

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ НАУЧНЫЙ ЦЕНТР
АГРОБИОТЕХНОЛОГИЙ РОССИЙСКОЙ АКАДЕМИИ НАУК

**В.К. КАЛИЧКИН, Р.А. КОРЯКИН,
П.К. КУЦЕНОГИЙ**

**ИНТЕЛЛЕКТУАЛЬНЫЕ
СИСТЕМЫ В СЕЛЬСКОМ ХОЗЯЙСТВЕ
(ТЕОРЕТИЧЕСКИЙ АНАЛИЗ ВОЗМОЖНОГО ПРИЛОЖЕНИЯ)**

Новосибирск 2020

УДК 004.9:631
ББК 4с51+32.97
К17

К17 **Каличкин В.К., Корякин Р.А., Куценогий П.К.** Интеллектуальные системы в сельском хозяйстве (теоретический анализ возможного приложения): монография / СФНЦА РАН. – Новосибирск: СФНЦА РАН, 2020. – 296 с.

Рецензенты:
доктор экономических наук *А.П. Задков,*
Д.А. Рассказов

Утверждено к печати экспертной комиссией СФНЦА РАН
(протокол № 4 от 14.10.2020).

ISBN 978-5-6044505-1-2

Проанализировано современное состояние интеллектуальных информационных систем от существующих концепций данных, информации и знаний до сложившейся практики их применения к сельскохозяйственным задачам. Сформулированы основные принципы и подходы цифровизации на примере процесса возделывания сельскохозяйственной культуры. На основе применения языка абстрактного моделирования UML разработаны концептуальные модели некоторых предметных областей сельскохозяйственной тематики. Представлена структура аграрной интеллектуальной системы, динамика ее функционирования и способы получения и хранения знаний о сельскохозяйственных объектах и процессах. Освещены вопросы машинного обучения, понятие образа события и перспективы применения численных моделей к прогнозированию урожайности сельскохозяйственных культур.

Книга предназначена для научных работников, IT-специалистов, аспирантов и студентов вузов сельскохозяйственного профиля, а также агрономов-технологов.

Рис. 36. Табл. 13. Библ. 332 назв.

УДК 004.9:631
ББК 4с51+32.97

ISBN 978-5-6044505-1-2

© Каличкин В.К., 2020
© Корякин Р.А., 2020
© Куценогий П.К., 2020
© СФНЦА РАН, 2020

MINISTRY OF SCIENCE AND HIGHER EDUCATION
OF THE RUSSIAN FEDERATION

FEDERAL STATE BUDGETARY SCIENCE INSTITUTION SIBERIAN
FEDERAL SCIENTIFIC CENTER OF AGROBIOTECHNOLOGIES
OF THE RUSSIAN ACADEMY OF SCIENCES

**V.K. KALICHKIN, R.A. KORYAKIN,
P.K. KUTSENOGIY**

**INTELLECTUAL SYSTEMS
IN AGRICULTURE**
(THEORETICAL ANALYSIS OF POSSIBLE APPLICATIONS)

Novosibirsk 2020

УДК 004.9:631
ББК 4с51+32.97
K17

K17 Kalichkin, V.K., Koryakin, R.A., Kutsenogiy, P.K. Intellectual systems in agriculture (theoretical analysis of possible applications): monography / SFSCA RAS. – Novosibirsk: SFSCA RAS, 2020. – 296 p.

Reviewers:
Doctor of Economic Sciences A.P. Zadkov,
D.A. Rasskazov

Approved for publication by the Expert Commission of SFSCA RAS
(Protocol No. 4 of 14.10.2020)

ISBN 978-5-6044505-1-2

The current state of intellectual IT-systems is analyzed: from the existing concepts of data, information and knowledge to the modern practice of their application to agriculture problems. The basic principles and approaches of digitalization are formulated on the example of agricultural crop cultivation process. By means of abstract modelling language UML, conceptual domain knowledge models for several agriculture areas have been developed. The structure of the agricultural intellectual system, the dynamics of its functioning, and methods of obtaining and storing knowledge about agricultural objects and processes are presented. The issues of machine learning, the fact image conception, and the prospects for applying numerical models to crop yield forecasting are highlighted.

The book is intended for researchers, IT specialists, agricultural university postgraduates and students, and for agronomy technologists.

Fig. 36. Table. 13. Bibl. 332 titles.

УДК 004.9:631
ББК 4с51+32.97

ISBN 978-5-6044505-1-2

© SFSCA RAS, 2020
© Kalichkin V.K., 2020
© Koryakin R.A., 2020
© Kutsenogiy P.K., 2020

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	7
Глава 1. Знания в искусственном интеллекте	15
1.1. Данные, информация, знания	15
1.2. Инженерия знаний	36
1.3. Модели представления знаний	49
1.4. Базы знаний	65
1.5. Управление знаниями	77
Глава 2. Цифровизация процесса возделывания сельскохозяйственной культуры	95
2.1. Понятие о цифровой зрелости интеллектуальной информационной системы	95
2.2. Использование элементов прикладного системного анализа	102
2.3. Шаблоны знаний	117
2.4. Накапливание данных и информации по сельскохозяйственным объектам	140
2.5. Ко второму этапу цифровой зрелости	149
Глава 3. Концептуальное моделирование	154
3.1. Абстрактный язык UML	155
3.2. Концептуальная модель ПО «агроэкологическая оценка земель»	159
3.3. Концептуальная модель ПО «выбор и размещение культуры»	171
3.4. Математические основания системы знаний	177
3.5. К созданию сельскохозяйственной ИИС	180
Глава 4. Элементы сельскохозяйственной интеллектуальной информационной системы	183
4.1. Аграрная интеллектуальная система	183
4.2. Структура и организация Хранилища Данных	193
4.3. Задачи организации ХД и использование накопленных знаний в агрономии	202
4.4. Репозиторий Знаний	210
Глава 5. Машинное обучение и интеллектуальный анализ данных	216
5.1. Постановка задачи применительно к предметной области, термины и определения	216
5.2. Достоверность и полнота исходных данных и релевантность прогноза. Понятие «образа» события	228
5.3. Влияние корректности данных на точность результата по опыту решения некорректных обратных задач. Влияние аппаратной функции на детализацию образа	232
5.4. Понятие «достоверности и независимости данных». Применимость образа в зависимости от специфики решаемой задачи	248
5.5. Перспективы применения описанных подходов к задаче прогноза урожайности сельскохозяйственных культур	255
5.6. Пример вычислений с применением описанного подхода	258
ЗАКЛЮЧЕНИЕ	266
БИБЛИОГРАФИЧЕСКИЙ СПИСОК	269

CONTENTS

INTRODUCTION	7
Chapter 1. Knowledge in Artificial Intelligence	15
1.1. Data, Information, Knowledge	15
1.2. Knowledge Engineering.	36
1.3. Knowledge Presentation Models	49
1.4. Knowledge Bases	65
1.5. Knowledge Management	77
Chapter 2. Agricultural Crop Cultivation Process Digitalization	95
2.1. Intellectual Information System Digital Maturity Index Conception .	95
2.2. System Analysis Methodology Applying.	102
2.3. Knowledge Templates	117
2.4. Agriculture Objects Data and Information Accumulation	140
2.5. To the 2 nd Stage of Digital Maturity	149
Chapter 3. Conceptual Modelling	154
3.1. Abstract language UML	155
3.2. Domain Knowledge Conceptual Model for Land Agroecological Properties	159
3.3. Domain Knowledge Conceptual Model for Crop Selection and Distribution	171
3.4. Mathematical Principles of Knowledge Systems	177
3.5. To Agricultural Intellectual Information System Creation.	180
Chapter 4. Agricultural Intellectual Information System Elements	183
4.1. Agrarian Intelligent System	183
4.2. Data Warehouse Structure and Organization	193
4.3. Data Warehouse Organization Problems and Knowledge Applica- tion in Agronomy	202
4.4. Knowledge Repository.	210
Chapter 5. Machine Learning and Data Mining	216
5.1. Formulation of the Problem within Domain: Terms and Definitions .	216
5.2. Reliability and Completeness of initial Data. Forecast Relevance. The Conception of “Image” of the Fact	228
5.3. Influence of Data Accuracy on the Result in case of Incorrect In- verse Problem Solution. Influence of measuring Device Sensitivity Function on possible Detailing of the Fact Image	232
5.4. «Data Reliability and Independence» Conception. The Applicabil- ity of the Fact Image Depending on the solving Problem	248
5.5. Possibilities to apply the described Approach for Crop Yield Forecast .	255
5.6. Test Calculations using Described Approach.	258
CONCLUSION.	258
BIBLIOGRAPHY.	269

ВВЕДЕНИЕ

В настоящее время проблемы информатизации сельского хозяйства значительно обострились в связи с необходимостью освоения и развития в России цифровой экономики (ЦЭ). В стратегии развития информационного общества в Российской Федерации дано следующее определение: «...цифровая экономика – хозяйственная деятельность, в которой ключевым фактором производства являются данные в цифровом виде, обработка больших объемов и использование результатов анализа которых по сравнению с традиционными формами хозяйствования позволяют существенно повысить эффективность различных видов производства, технологий, оборудования, хранения, продажи, доставки товаров и услуг».

В принятой правительством РФ 28 июля 2017 г. программе «Цифровая экономика Российской Федерации» основное внимание уделяется необходимости создания ряда индустриальных цифровых платформ под руководством профильных министерств или госкорпораций. Такой подход не лишён своих недостатков, так как остальным отраслям, в том числе и агропромышленному комплексу, придется продолжать стихийный процесс самоорганизации цифрового пространства, после чего переход на единую платформу ЦЭ, по мнению некоторых ученых, будет очень дорогим или интеграция окажется вообще невозможной.

В то же время до сих пор не удалось достичь того, чтобы общий менеджмент информации в сельском хозяйстве был бы организован в простой и совместимой форме. Принятая методология и организация статистического наблюдения в сельском хозяйстве, а также существующая в децентрализованном виде система сбора информации не позволяют в полной мере решать поставленные задачи его инновационного развития. Информационные системы Росстата (Статистический регистр Росстата и Единая межведомственная информационно-статистическая система) и Минсельхоза (например, Единая федеральная информационная система о землях сельскохозяйственного назначе-

ния) не удовлетворяют необходимым стандартным требованиям полноты, избыточности, непротиворечивости и актуальности, предъявляемым к базам данных; структура данных не ориентирована на экспорт в пакеты программ для проведения статистического и иного анализа; информация разбросана по разным базам данных, не стыкуется по структуре, различается в значениях, поскольку получена из разных источников сбора и обработки данных; конвертирование информации сопряжено с большой трудоёмкостью и высокой вероятностью ошибок.

В настоящее время появились принципиальные возможности по созданию практических инструментов прогнозирования сельскохозяйственной деятельности. Данная возможность связана с методами, базирующимися на принципах искусственного интеллекта. До появления этих методов было невозможно статистически и экономически значимо строить адекватные сельскохозяйственные модели. Объясняется это тем, что прогноз сельскохозяйственной деятельности – решение обратной задачи с заведомо неполным набором исходных данных (некорректная задача). При решении подобных задач нельзя искусственно навязывать прогнозам большую точность, чем позволяет заведомо не точно соизмеримая исходная информация. Одним из важнейших условий высокого качества системно-статистических прогнозов является формирование адекватных баз данных для соответствующих расчетов. С одной стороны, это касается достоверности используемого статистического материала, с другой – связано с его объёмом как в части географической представительности, так и в динамике.

Не только информация о производственных процессах и природных условиях ведения сельского хозяйства, но и сельскохозяйственные знания не систематизированы и слабо структурированы, не основываются на общей платформе представления и анализа данных и системном подходе, изложены в больших текстах (монографиях, статьях и рекомендациях). Исследования проводятся по узким направлениям, слабо развиты междисциплинарные исследования. Проблема осложняется нелинейным взаимодействием факторов и огромными размерностями задач.

Следует добавить, что сельское хозяйство ведётся в условиях неопределенности, которое является фундаментальным свойством

природы. Неопределённость – это факт, с которым все формы жизни вынуждены бороться, т.е. на всех уровнях биологических процессов существует неопределённость относительно возможных последствий событий и действий. Принятие решения в условиях неопределённости означает выбор такого варианта, когда одно или несколько действий имеют своим следствием множество частных исходов, но которые можно просчитать с некоторой долей вероятности. Президент РАН А.М. Сергеев (2017) о поведении сложных систем сказал: «Физики пришли к выводу, что наука не может предсказать развитие сложных систем на достаточно больших временах эволюции. Они в принципе непредсказуемы! Во всём мире ученые поняли, что не стоит больше искать способов их предсказания – надо изменить своё отношение к проблеме. А именно – найти способы ограничения динамического хаоса и компенсации действия непредсказуемых факторов...».

Один из ключевых факторов глобальной цифровизации экономики – развитие искусственного интеллекта. Под искусственным интеллектом понимается обычно способность автоматических или автоматизированных систем брать на себя некоторые функции интеллекта человека, например принимать оптимальные решения на основе анализа внешних воздействий и с учетом ранее полученного опыта. Можно выделить несколько направлений, в которых развиваются средства искусственного интеллекта. Среди них отметим экспертные системы, интеллектуальные игры, распознавание образов, робототехнику, общение с ЭВМ на естественном языке. Проблемы искусственного интеллекта тесно связаны с организацией знаний об окружающем мире в виде математических структур, таких как множества, графы, гиперграфы, фреймы, алгоритмы, которые отражают реальные связи и отношения между любыми объектами в природе (в частности, в предметной области).

С учетом сложившихся реалий настоящего времени – стремления к освоению «умного» сельского хозяйства, а также стратегии научно-технологического развития страны, в основе которой лежит цифровизация, мы посчитали необходимым в данной книге показать актуальность исследований по интеллектуальному анализу данных, а не просто применение цифровых технологий (ДЗЗ, параллельное вождение, автоматизированные внесение удобрений и обработка посевов

от вредных организмов и др.). Применение цифровых технологий в технологическом регламенте неплохо освещено не только в западной, но и отечественной литературе. Интеллектуальный анализ в сельском хозяйстве с использованием методов искусственного интеллекта для поддержки принятия решений специалистами аграрного производства в отечественной научной литературе освещён и разработан очень слабо. В этом отношении мы значительно отстали не только от зарубежных специалистов, но и от российских коллег, занимающихся данными проблемами в области биологии и медицины.

Мы убеждены, что ЦЭ, в том числе цифровое сельское хозяйство, без создания систем интеллектуального анализа данных и построения на этой основе систем поддержки принятия решений не найдёт эффективного практического применения.

Данная книга является попыткой теоретически обосновать возможность применения методов искусственного интеллекта к разработке предиктивных технологий сельскохозяйственного производства. Насколько это нам удалось, судить вам, уважаемые читатели.

INTRODUCTION

Nowadays, the problems of agriculture informatization have escalated due to Russian Federation digital economy exploration and development necessity. The following definition is formulated in Russian Federation information society development strategy: “Digital Economy is an economic activity with digitalized data as a key production factor, so that processing large volumes and applying the results of analysis of such data can significantly increase the efficiency of various types of production, technologies, equipment, storage, sale, delivery of goods and services, in comparison with traditional forms of management”.

The program “Digital economy of the Russian Federation” accepted by Russian Federation government on July 28, 2017 focuses on creating a number of industrial digital platforms directed by relevant ministries or state corporations. This approach has disadvantages since other industries, including the agro-industrial complex, have to continue the spontaneous process of digital space self-organization, after which, in opinion of some scientists, the transition to some unified digital economic platform will be either very expensive or impossible at all.

At the same time, it has not yet managed to get agriculture general information processing organized in a simple and compatible form. The statistical observation methodology and organization accepted in agriculture, as well as existing decentralized information gathering system, do not provide entire solutions of agriculture innovative development problems. The information systems of Rosstat (Rosstat Statistical Register and Unified Interdepartmental Information and Statistical System) and the Ministry of Agriculture (for example, Agricultural Land Unified Federal Information System) do not satisfy the necessary standard requirements for completeness, non-redundancy, consistency and relevance for databases; the data structure is not intended for export to statistical and other analysis software packages; information is scattered across different databases, does not fit together in structure, and differs in values because it is obtained from dif-

ferent data gathering and processing sources; information conversion is very time-consuming and error-prone.

Principal opportunities to develop agricultural activity forecasting practical tools have appeared nowadays. This is related to methods based on artificial intelligence principles. Prior to appearance of these methods, it was not possible to construct adequate agricultural models with statistical and economical significance. This is explained by the fact that any agricultural activity forecast is a solution to the inverse problem with a knowingly incomplete set of initial data (an incorrect problem). When solving such problems, the forecasts cannot be artificially imposed by an accuracy greater than allowed by knowingly not precisely comparable source information. One of the most important conditions for high-quality system-statistical forecasts is the formation of adequate databases for the corresponding calculations. On the one hand, this concerns the reliability of the statistical material used, on the other hand, it is related to its volume, both in terms of geographical representation and dynamics.

Information of production processes and agriculture activity nature conditions, as well as agricultural knowledge, are not systematized and poorly structured, are not based on any data representation, data analysis and a system approach common platform, and are presented in large texts (monographs, articles, and recommendations). Research is conducted within narrow areas, and interdisciplinary research is poorly developed. The problem is complicated by the nonlinear interaction of factors and huge dimensions of problems.

It should be added that agriculture is conducted under conditions of uncertainty, which is a fundamental property of nature. Uncertainty is a fact, with which all the life forms must fight, meaning there is uncertainty about the possible consequences of events and actions at all levels of biological processes. Decision making under conditions of uncertainty means choosing an option where one or more actions have many particular outcomes, which however can be calculated with a certain share of probability. President of the Russian Academy of Sciences A. M. Sergeev (2017) said the following about the behavior of complex systems: Physicists have concluded that science cannot predict the development of complex systems over sufficiently long evolutionary times. They are unpredictable in principle! All over the world, scientists have realized that it is no longer worth

looking for ways to predict them – one needs to change the attitude to the problem. Namely, to find ways to limit dynamic chaos and to compensate for the unpredictable factor effects ».

One of the economy global digitalization key factors is the artificial intelligence development. Artificial intelligence is usually understood as the automatic or automated system ability to take on certain human intelligence functions, for example, to make optimal decisions based on the external influence analysis and previous experience considering. There are several areas in which artificial intelligence tools are being developed. Among them, we note expert systems, mind games, image recognition, robotics, and natural language communication with computer. Artificial intelligence problems are closely related to the organization of knowledge about the surrounding world in the form of mathematical structures, such as sets, graphs, hypergraphs, frames, algorithms that reflect real connections and relationships between any objects in nature (particularly, in domain knowledge).

Taking into account the current present-day realities – the intention to explore “smart” agriculture and the country scientific and technological development strategy, based on digitalization, – we considered it necessary to show in this book the relevance of research on data mining, and not just digital technology applying (remote sensing, parallel driving, automated fertilization and treatment of crops from harmful organisms, etc.). Digital technology applying in technological regulations is well covered not only in Western, but also in Russian literature. However, data mining in agriculture with applying artificial intelligence methods to support agricultural production specialist decision making is very poorly covered and developed in the Russian scientific literature. In this respect, we are far behind not only foreign experts, but also our Russian colleagues dealing with these problems in biology and medicine.

We are convinced that the digital economy, including digital agriculture, will not find effective practical application without creating data mining systems and building decision support systems on this basis.

This book is an attempt to theoretically substantiate the possibility of applying artificial intelligence methods to the development of agricultural production predictive technologies. As far as we were able to do it, judge You, dear readers.

Наука – это особая разновидность Игры. Игра ведется по особым правилам, которые известны и понятны всем играющим, хотя они и не были еще ни разу классифицированы и кодифицированы.

...В процессе этой Игры создаются хитроумные и все усложняющиеся теоретические построения, которым сами игроки, кажется, не очень верят. Во всяком случае, законченное знание у них считается заблуждением, ибо только тот ученый считается по-настоящему талантливым, который сумел разрушить то, что было создано до него.

В.В. Налимов

ГЛАВА 1

ЗНАНИЯ В ИСКУССТВЕННОМ ИНТЕЛЛЕКТЕ

1.1. Данные, информация, знания

Данные, информация и знания являются ключевыми словами, а также фундаментальными понятиями в интеллектуальной обработке данных с помощью методов искусственного интеллекта (ИИ). Именно вокруг этих понятий не утихают научные дискуссии, начатые во второй половине XX в. Как отмечает один из российских экспертов в области информационных технологий Л. Черняк (2003), «...прежде всего потому, что со времени Клода Шеннона на инженерном уровне происходило смешение понятий, объединение представлений об информации и данных или сигналах, кодирующих эту информацию. С легкой руки первопроходцев информацией стали называть, по существу, наборы данных. До последнего времени, пока системы были относительно просты, отсутствие четкого разделения на данные и информацию не имело практического значения. Но с появлением сложных информационных систем, где функции распределены между человеком и машиной, где человек является активной составляющей, а также развитием таких дисциплин, как управление знаниями, требуются более точные определения базисных понятий: «данные», «информация» и «знания» [Черняк, 2003]. На эту же особенность информационно-интеллектуальных систем обращает внимание В. Палицин (2018). Г.Б. Бронфельд (2016) указывает, ссылаясь на И.Н. Бекмана (2009), что определений «информация» насчитывается более 500 [Бронфельд, 2016].

Одним из направлений определения информации является ее связь с энтропией. В классической термодинамике полный дифференциал энтропии замкнутой системы – величина больше нуля, что с физической точки зрения означает: в замкнутых системах энтропия должна увеличиваться. Отсюда вытекает определение об однозначном росте энтропии (не уменьшении) в результате любого термодинамического

процесса. С физической точки зрения, данный факт интерпретируется в том смысле, что в системе в итоге растёт степень «неупорядоченности» и любая система (включая Вселенную) движется в сторону неупорядоченного «белого» шума. Однако в живых системах степень упорядоченности однозначно растёт по мере развития и «обучения» данной системы. Сама жизнь противоречит данному постулату. Например, в тех формах жизни, с которыми мы знакомы, генетическая информация записана в виде упорядоченной (осмысленной) последовательности «букв» (нуклеотидов) в ДНК. Любая живая система по мере роста и размножения увеличивает количество копий ДНК, несящей все новую информацию, таким образом, система развивается в сторону, противоположную той, которую задает второе начало термодинамики.

В живых системах «обучение», связанное с накоплением информации, понимается в достаточно широком смысле, в биологии это может быть приспособленность к условиям окружающей среды. Человек, накапливая информацию, преобразует природу вокруг себя в сторону упорядоченности: строит дома, дороги, заводы и др. Мысль связать информацию с энтропией родилась на рубеже XIX–XX вв., хотя ее в основном связывают с именем Клода Шеннона. Связь энтропии с информацией описывается с помощью классического парадокса термодинамики, связанного с мысленным экспериментом, так называемым «демоном Максвелла». Суть парадокса заключается в том, что рассматриваются два сосуда с разными температурами, соединённые узкой трубкой с затворками, которыми управлял «демон». «Демон» мог измерять скорость отдельных летящих молекул и таким образом избирательно пропускать более быстрые в сосуд с высокой температурой, а более медленные – в сосуд с низкой. Из этого мысленного эксперимента вытекало кажущееся противоречие со вторым началом термодинамики, так как появлялась возможность создать тепловую машину без закачки дополнительной энергии в систему извне.

Парадокс может быть разрешён при помощи теории информации. Для измерения скорости молекулы «демон» должен был бы получить информацию о её скорости. Но всякое получение информации – материальный процесс, сопровождающийся возрастанием энтропии. Количественный анализ показал, что приращение энтропии при измере-

нии превосходит по абсолютной величине уменьшение энтропии, вызванное перераспределением молекул «демоном» [Бриллюэн, 2012].

Прежде всего, что такое «информация»? Шеннон отождествляет информацию с неопределенностью. Данные рассуждения позволяют ввести численные определения информации, тесно связанные с числом «независимых данных». Пусть есть некое количество N численных значений, которые являются результатом независимых действий (измерений). Величину $N + 1$ можно получить как неким независимым экспериментом, так и однозначно вывести из N предыдущих значений с помощью некоторого «знания». Во втором случае в новой системе будет $N + 1$ данных, но *не будет* новой информации по Шеннону. Шеннон полагал, что количественная мера информации должна быть непрерывной функцией от вероятности события p , а для независимых событий она должна быть аддитивной: количество информации, полученное в результате осуществления двух независимых событий, должно равняться количеству информации, полученному в результате осуществления совместного события. Таким образом, отличие «данных», «знаний» и «информации» приобретают математический смысл.

Определение, данное Шенноном для информации, во многих случаях подходит для машин, но, видимо, не соответствует человеческому пониманию этого слова. Хотя с практической точки зрения (физической или математической) может быть довольно ясно интерпретировано или проиллюстрировано. Например, если некоторая величина однозначно выводится (рассчитывается) из известных величин, то ее получение (измерение) не даст никакой новой информации. Это ответ в стиле «мы и так это знали, нам и так это было известно, мы лишь подтвердили то, что и без этого очевидно». Связь информации с неопределенностью прослеживается в современных технологиях машинного обучения. Далее в монографии будет рассмотрено предположение о том, что адекватное предсказание будущих событий возможно лишь на основе «образа» системы, допускающего значительные неопределенности граничных условий, в которые система помещается в результате численного эксперимента. Следовательно, «образ» должен быть стабильным, устойчивым в определенных пределах к изменчивости граничных условий. Если система является точной моделью,

результат расчета которой жестко завязан на численные значения граничных условий, то прогностическая ценность такой модели равна нулю. В машинном обучении такая система называется «переобученной». Дело в том, что ожидаемый результат будет достигнут только в случае 100%-го повторения условий, соответствующих данному результату. Во-первых, в реальном физическом мире и в условиях «живых» систем невозможно 100%-е повторение ранее наблюдавшихся условий. Во-вторых, точное воспроизводство условий в гипотетической реальности (мысленный или численный эксперимент) просто приведет к повторению уже известного результата и не добавит никакой новой информации для наблюдателя.

Несмотря на многочисленные попытки определения «данных», «информации» и «знаний», по-прежнему отсутствует четкая и полная картина того, что они собой представляют. Хотя многие определения актуальны, они далеко не полны. Ниже приведена таблица различных определений данных, информации и знаний от разных авторов, сгруппированных Э. Лью (2007). В таблицу также включены определения из Универсального словаря Вебстера (Merriam Webster's Collegiate Dictionary). По мнению Э. Лью, большинство определений имеют общую аномалию: они определяются посредством друг друга, т.е. данные определяются в терминах информации, информация – в терминах данных или знаний, знания – в терминах информации. Он считает, что «круговые» определения или аргументы являются логической ошибкой [Liew, 2007]. Тем не менее, мы считаем целесообразным привести это обобщение здесь, поскольку оно наглядно показывает пестроту мнений по этому вопросу, а также дает представление о сложности отношений между ними (табл. 1). Последняя строка отражает позицию авторов данной работы по отношению к данным, информации и знаниям.

Понимание сложной взаимосвязи между данными, информацией и знаниями лежит в источнике данных и информации. Источники обоих: 1) действия и 2) ситуации. Действия и ситуации формируют информацию (т.е. «релевантное значение» для кого-либо), которую или учитывают (измеряют), превращая в данные, или теряют. Информация по своей сути является сообщением, которое генерируется из действий и ситуаций. Однако информация хранится на носителях

Определения: данные, информация, знания (Liew, 2007)

Данные	Информация	Знания	Source (Источник)
1	2	3	4
Данные – это набор дискретных, объективных фактов о событиях ... как структурированные записи транзакций.	Информация – это поток сообщений.	Знание создается самим потоком информации, основанным на убеждениях и приверженности его владельца.	Nonaka I., Takeuchi H. The knowledge-creating company: How Japanese companies create the dynamics of innovation. – Oxford university press, 1995.
Данные должны быть организованы, чтобы стать информацией.	Информация должна быть помещена в контекст, чтобы стать знанием.		Garvin A.P. The art of being well informed: What you need to know to gain the winning edge in business. – Avery, 1996.
Данные являются элементами анализа.	Информация – это данные с контекстом.	Знание – это информация со смыслом.	Debra M.A. Innovation strategy for the knowledge economy: the Ken Awakening // Newton, MA: Butterworth Heinmann, 1997.
Данные – это набор дискретных, объективных фактов о событиях ... как структурированные записи транзакций.	Информация ... как сообщение ... в (различной) форме общения ... чтобы повлиять на суждение и поведение.	Знания – это сочетание опыта, ценностей, контекстуальной информации и экспертных знаний, которые представляют собой основу для оценки и включения нового опыта и информации...	Davenport T.H., Prusak L. Working knowledge: How organizations manage what they know. – Harvard Business Press, 1998.

Продолжение табл. 1

1	2	3	4
		Знание – это совокупность информации, техники и опыта, которые объединяются вокруг определенного предмета.	Horibe F. Managing knowledge workers: New skills and attitudes to unlock the intellectual capital in your organization. – John Wiley & Sons, 1999.
	Информация: обработанные данные... формализованные, фиксированные и объясненные; может быть легко упакованы в много-разовую форму. Информация – это данные, помещенные в контекст; это связано с другими частями данных. Информация о значении, и она формирует основу для знаний.	Знание: действенная информация... часто по-является в умах людей посредством опыта.	Tiwana A., Williams M. The Essential Guide to Knowledge Management: E-Business and Crm Applications. – Prentice Hall PTR, 2000.
		Знание ... включает в себя убеждения групп или отдельных людей и тесно связано с действиями.	V. Krogh G., Ichijo K, Nonaka I. Enabling knowledge creation: How to unlock the mystery of tacit knowledge and release the power of innovation. – Oxford University Press on Demand, 2000.
Информация ... данные, которые «находятся в процессе формирования», то есть данные, которые были сохранены, проанализированы и отображены, и пере-даны через разговорный язык, графические дисплеи или числовые таблицы.		Знание ... определяется как значимые связи, кото-рые люди устанавливают в своем уме между ин-формацией и ее приме-нением в действии в опреде-ленных условиях.	Dixon N.M. Common knowledge: How companies thrive by sharing what they know. – Harvard Business School Press, 2000.

Продолжение табл. 1

1	2	3	4
<p>Данные состоят из основной, нефильтрованной и, как правило, нефильтрванной информации.</p>	<p>Информация ... это гораздо более точные данные ... которые стали настолько полезными для анализа.</p>	<p>Знания принадлежат пользователю ... появляются только тогда, когда человеческий опыт и понимание применяются к данным и информации.</p>	<p>Kelley J. Knowledge Nirvana-Achieving The Competitive Advantage Through Enterprise Content Optimizing Team Collaboration. – Xulon Press, 2002.</p>
		<p>...смесь организованного опыта, ценностей, информации и идей, предлагающая основу для оценки нового опыта и информации.</p>	<p>Sydänmaanlakka P. An intelligent organization: Integrating performance, competence and knowledge management. – Capstone Publishing, 2002.</p>
<p>Данные: 1) фактическая информация, используемая в качестве основы для рассуждения, обсуждения или расчета; 2) информация, выводимая чувствительным устройством или органом, которая включает как полезную, так и не относящуюся к делу или избыточную информацию и должна быть обработана, чтобы быть значимой; 3) информация в цифровой форме, которая может быть передана в цифровой форме или обработана.</p>	<p>Информация: 1) передача или получение знаний или интеллекта; 2) знания, полученные в результате расследования или обучения; 3) факты, данные; 4) количественная мера содержания информации.</p>	<p>Знание: 1) познание; 2) факт или условие знания чего-либо со знакомством, приобретенным через опыт или общение; 3) диапазон информации или понимания; 4) сумма того, что известно: совокупность истины, информации и принципов, приобретенных человечеством.</p>	<p>Merriam Webster's Collegiate Dictionary 10th ed.</p>

1	2	3	4
<p>Данные: сохраняются символами и показаниями сигналов – 1) символы включают слова (текстовые и/или словесные), цифры, диаграммы и изображения (неподвижные и/или видео), которые являются строительными блоками общения; 2) сигналы включают сенсорные показания света, звука, запаха, вкуса и осязания.</p>	<p>Информация – это сообщение, которое содержит соответствующее значение или входные данные для принятия решения и/или действия. Информация поступает как из текущих (коммуникационные), так и из исторических (обработанные данные или «восстановленная картина») источников.</p>	<p>Знание – это 1) познание или признание (know-what), 2) способность действовать (know-how) и 3) понимание (know-why), которое находится или содержится в уме или в мозге.</p>	<p>Liew A. Understanding data, information, knowledge and their interrelationships // Journal of knowledge management practice. 2007. Vol. 8. № 2. P. 1–16.</p>
<p>Упорядоченные наборы данных, соответствующие стандарту входа для программ ЭВМ, предназначенных для интеллектуального анализа или их визуального отображения. Данные – это то, что можно обработать с помощью программно-стандартизированной функции.</p>	<p>Перевод входных и выходных данных программ ЭВМ, предназначенных для интеллектуального анализа и визуального отображения данных в термины предметной области. Презентационное описание работы программы ЭВМ, получившей одни данные на входе и выдавшей другие данные на выходе. Информация – это то, что объясняет и интерпретирует обработанную данные.</p>	<p>Состав и структура множества программно-стандартизированных функций, способных эффективно обрабатывать данные, связанные с предметной областью или её фрагментом. Точные алгоритмы получения ответов на вопросы предметной области с помощью функций обработки данных. Правила параллельного и последовательного применения функций обработки данных и программ ЭВМ для решения сложных задач.</p>	<p>Материалы внутренних семинаров Лаборатории предиктивных агротехнологий с использованием методов искусственного интеллекта СФНЦА РАН</p>

(база данных, книга, видео и др.) в форме данных или в человеческом разуме как знание (в простейшей форме ноу-хау или более высоких формах ноу-хау и знания). Пересечение данных и информации по отношению к информации и знаниям очевидно, т.е. они занимают разное пространство одновременно. Это также объясняет, почему многие воспринимают данные и информацию, а также знания как взаимозаменяемые. «Данные одного человека могут быть знанием другого человека и, наоборот, в зависимости от контекста» [Stewart, 1997]. Однако они не являются взаимозаменяемыми понятиями с точки зрения их принятых определений. В связи с этим возникает вопрос, что же такое книга: знания, информация или данные? Книга – это все перечисленное, но в различном контексте: знание с точки зрения автора, информация для потенциального читателя, а также данные, которые содержатся на носителе, называемом «книга» [Liew, 2007].

Следует отметить, что отличия знаний от данных с развитием средств информатики сглаживаются, так как в последние 20 лет разработчики систем управления базами данных делают их все более похожими на знания. Примером может служить применение семантических сетей (формализм для представления знаний) для проектирования баз данных, появление объектно-ориентированных баз данных, хранимых процедур (это делает в какой-то мере данные активными).

Одним из первых проблему соотношения «данные – информация – знания» поднял Р. Акофф (1989), системный теоретик и профессор Вашингтонского университета. По его мнению, содержание человеческого разума можно разделить на пять категорий [Ackoff, 1989]:

- *данные*: символы;
- *информация*: данные, которые обрабатываются, чтобы быть полезными; дает ответы на вопросы «кто», «что», «где» и «когда»;
- *знание*: применение данных и информации; отвечает на вопросы «как»;
- *понимание*: оценка «почему»;
- *мудрость*: оцениваемое понимание.

Р. Акофф указывает, что первые четыре категории относятся к прошлому; они имеют дело с тем, что было или что известно. Лишь пятая категория (мудрость) относится к будущему, потому что включает в себя видение и разработку проблемы. С мудростью люди могут создавать будущее, а не просто понимать настоящее и прошлое. Однако

достичь мудрости нелегко: люди должны последовательно проходить через другие категории.

Д. Роули (2007), изучив популярные информационные системы и учебники по управлению знаниями, выделила ряд неоднозначных и противоречивых определений, подмножество которых показано в табл. 2. Она также перефразировала Р. Акоффа следующим образом. Мудрость – это способность повышать эффективность. Знание – это ноу-хау, и именно это делает возможным преобразование информации в инструкции. Информация дает ответы на вопросы «кто», «что», «где» и «когда». Данные определяются как символы, которые представляют свойства объектов, событий и их окружающей обстановки. Они являются продуктами наблюдения [Rowley, 2007].

Со временем этот подход в работах западных авторов оформился в виде «DIKW-модели» [Rowley, 2007; Bernstein, 2009; Frické, 2009; Baskarada, Koronios, 2013; Aven, 2013]. «DIKW-модель», производное от англ. data, information, knowledge, wisdom (данные, информация, знания, мудрость). В «DIKW-моделях» отражаются иерархические отношения подчинения понятий данных, информации, знания и мудрости, где каждый уровень добавляет определенные свойства к предыдущему уровню. Обычно изображается в виде пирамиды, в ее основании D (data – данные) находится уровень данных. Следующий уровень I (information) – уровень информации добавляет контекст. Следующий уровень K (knowledge – знание) – уровень знания добавляет «как» (механизм использования). Следующий уровень W (wisdom – мудрость) – уровень знания добавляет «когда» (условия использования) [Седакин, 2009]. Э. Лью (2013) добавляет в эту классическую модель еще и интеллект (DIKIW: data, information, knowledge, intelligence, wisdom) [Liew, 2013].

С. Янсен с соавт. (2017), обсуждая развитие следующего поколения (NextGen) сельскохозяйственных моделей и поддерживающего их модельного сообщества с позиции развития информационных и компьютерных технологий (ИКТ), представляют DIKW-модель в виде цепочки знаний (knowledge chain) [Janssen et al., 2017]. Цепочка знаний состоит из набора связанных шагов, с помощью которых данные обрабатываются в информацию, знания и, наконец, мудрость, используемые в процессе принятия решений (рис. 1).

Определение данных, информации, знаний и мудрости (Rowley, 2007; Baskarada S., Koronios A., 2013)

Мудрость	<p>Мудрость – это накопленные знания, которые позволяют понять, как применять концепции из одного домена в новые ситуации или проблемы (Jessup and Valacich, 2003).</p> <p>Мудрость – это высший уровень абстракции с дальновидностью и способностью видеть дальше горизонта (Awad and Ghaziri, 2004, p. 40).</p> <p>Мудрость – это способность действовать критически или практически в любой конкретной ситуации. Она основана на этическом суждении, касающегося системы убеждений человека (Jashapara, 2005, p. 17–18).</p>
Знание	<p>Знание – это комбинация данных и информации, к которой добавляется экспертное мнение, навыки и опыт, чтобы получить ценный актив, который может быть использован для принятия решений (Chaffey and Wood, 2005, p. 223).</p> <p>Знание – это данные и/или информация, которые были организованы и обработаны для передачи понимания, опыта, накопленных знаний и опыта их применения к текущей ситуации, проблемы или деятельности (Turban et al., 2005, p. 38).</p> <p>Знание основывается на информации, извлекаемой из данных... В то время как данные являются собственностью вещи, знание – это свойство людей, которое предрасполагает их действовать определенным образом (Boddy et al., 2005, p. 9).</p>
Информация	<p>Информация – это данные, которые повышают ценность понимания предмета (Chaffey and Wood, 2005, p. 233).</p> <p>Информация – это данные, которые были сформированы в форме, которая имеет смысл и полезна для человека существа (Laudon и Laudon, 2006, p. 13).</p> <p>Информация – это совокупность данных, которая облегчает принятие решений (Awad and Ghaziri, 2004, p. 36).</p>
Данные	<p>Данные не имеют смысла или ценности, потому что они не имеют контекста и интерпретации (Jessup и Valacich, 2003, Bocij et al., 2003, Groff and Jones, 2003).</p> <p>Данные являются дискретными, объективными фактами или наблюдениями, которые являются неорганизованными и необработанными, и не передают какого-либо конкретного значения (Awad and Ghaziri, 2004, Chaffey and Wood, 2005, Pearlson and Saunders, 2004, Bocij et al., 2003).</p> <p>Элементы данных являются элементарным и записанным описанием вещей, событий, действий и транзакции (Laudon and Laudon, 2006, Turban et al., 2005, Boddy et al., 2005).</p>

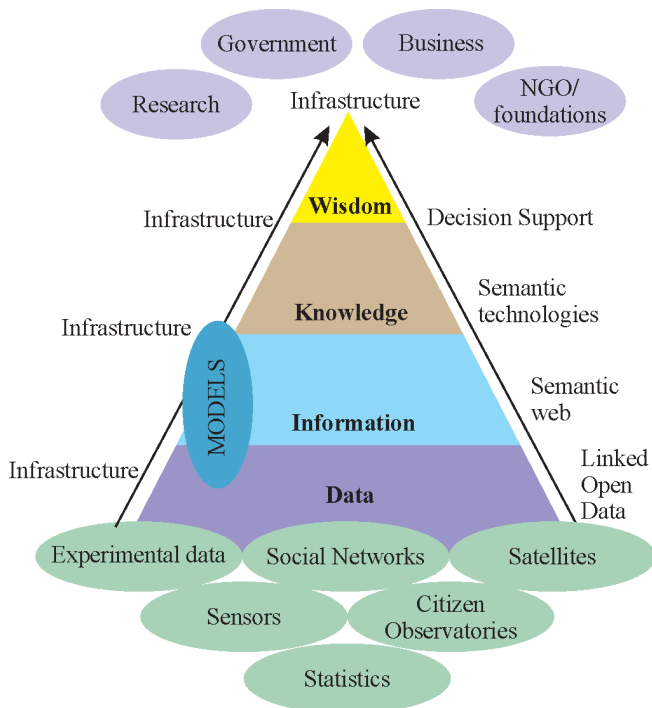


Рис. 1. Пирамида знаний, связывающая данные с информацией, знаниями и мудростью, в которой данные являются сырьем для разработки приложений, направленных на принятие решений через мудрость (Lokers et al., 2016; Janssen et al., 2017)

Авторы постулируют, что данные представляют собой сырье, которое в сочетании с описанием и качественными атрибутами приводит к информации. Информация может быть связана с другими источниками информации и помещена в причинно-следственные цепи для получения знаний. В конечном счете, знание служит вкладом в решения, основанные на мудрости, которые не могут быть оцифрованы и которые существуют в сознании лица, принимающего решения [Janssen et al., 2017].

Здесь возникает польза математического определения информации в современных методах ИИ. Данные сами по себе являются источником информации в подходе «обучение без учителя», но лишь в

том объеме, в каком содержат эту информацию. Не весь и не любой массив данных может являться однозначным источником *новой* информации, а лишь в той части, в которой данные являются истинно «новыми» и «независимыми». Парадокс в том, что истинно «новые» данные содержат информацию вне зависимости от «наблюдателя». Польза «наблюдателя» – лишь в извлечении этой информации в удобной форме, т.е. в ее трансформации.

Вторая перспектива развития моделей для сельского хозяйства сосредоточена на прикладных цепочках (application chains). Сельскохозяйственные модели должны быть задействованы в инфраструктуре, состоящей как из программного обеспечения (например, в слоях доступа к данным, обработки, анализа и визуализации), так и из оборудования (т.е. серверов, вычислительных мощностей и хранения), как показано на рис. 2.

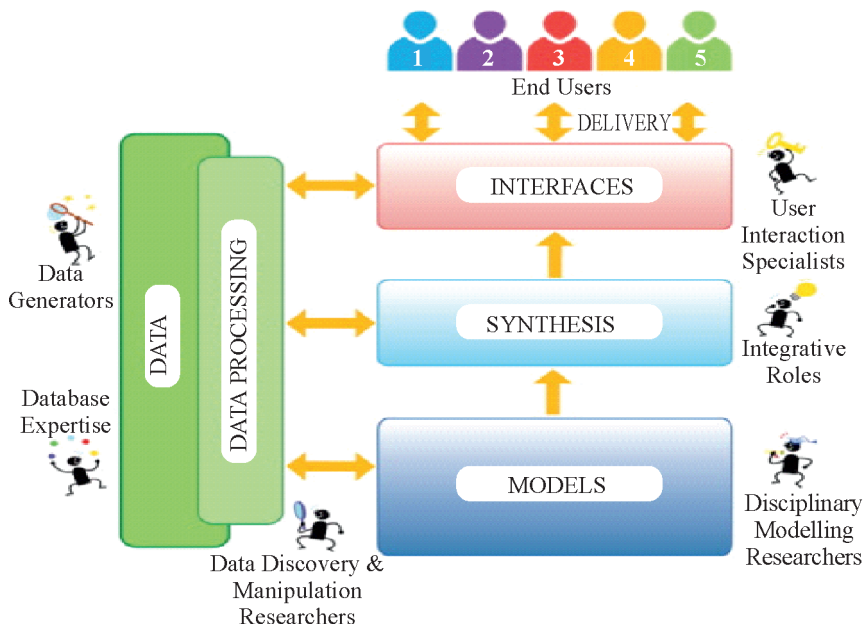


Рис. 2. Цепочки приложений, описывающие поток данных и информации через слои моделирования, синтеза и взаимодействия с конечными пользователями с ролью различных участников информационной цепи [Janssen et al., 2017]

На основе данных, содержащихся в инфраструктуре, приложения, ориентированные на конечных пользователей, служат для получения информации и знаний, например, прогноза получения дохода или оценки ущерба, причиненного болезнями сельскохозяйственных культур. Цепочки приложений могут быть простыми или сложными, могут включать, в частности, доступ к данным, извлечение, преобразование (например, суммирование или интерполяция) и операции интеграции; одну или несколько моделей; интеграцию выходных данных из разных моделей; этапы преобразования, анализа и визуализации выходных данных модели. При разработке цепочек приложений должны учитываться не только конечные пользователи, но и весь спектр пользователей инфраструктуры ИКТ NextGen, включая сборщиков первичных данных, специалистов по базам данных, инженеров по знаниям, разработчиков программного обеспечения и конечных пользователей знаний и информации [Janssen et al., 2017].

В последние годы DIKW-модель стала популярна и в среде отечественных информационных специалистов [Седакин, Цветков, 2007; Цветков, 2013; Ожерельева, 2015; Бронфельд, 2016; Пахомов, 2017]. Она дает подход к определению информационных понятий, включая такие как носитель информации, данные, информация, знания. Кроме очевидной наглядности, модель позволяет раскрыть отношения между всеми известными понятиями, включая информацию и знания, и установить обусловленность этих отношений. Также в модели показано то, что возникновение данных отношений не абсолютно, а относительно, связано с выполнением некоторых условий.

Авторам настоящей книги представляется рабочим следующий вариант DIKW-модели (рис. 3).

Кружками выделены источники данных, на верхнем уровне – цель использования модели. Угловыми светлыми стрелками обозначен переход на новый цикл целевого использования модели. В схеме видно, что у каждого из элементов DIKW по два входящих источника, которые иллюстрируют интуитивный и механистический способы движения от источников данных к эффективному видению и прогнозированию. Например, интеллект нарабатывается из длительного наблюдения за ситуациями или на основе уже готовых знаний. Первый случай характеризует опытного специалиста (например, агронома),

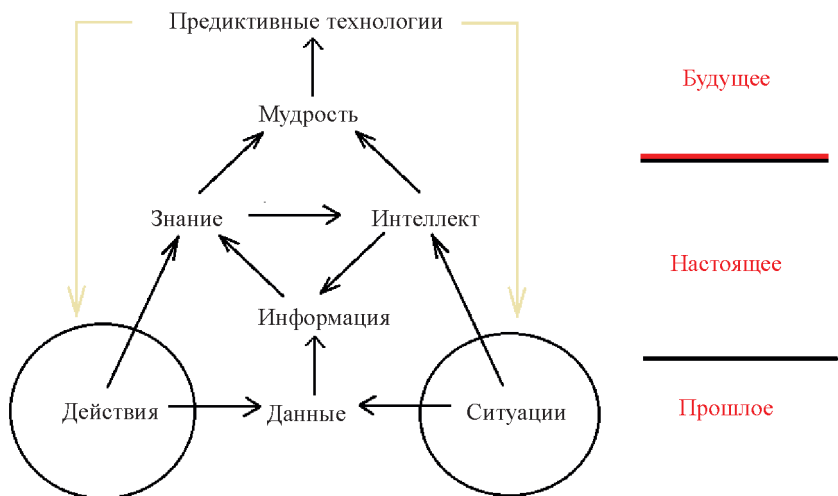


Рис. 3. Схема формирования и функционирования DIKW-модели

наблюдающего за одними и теми же явлениями в течение длительных периодов времени, второй – симуляцию этого специалиста с помощью ИИ или коллектива, слаженно работающего по готовым инструкциям.

Понимание авторами данной работы терминов «Данные», «Информация» и «Знания» уже приведено в табл. 1. Это понимание основано на возможности применения к данным, информации и знаниям программно-стандартизированных функций их обработки. ИИ можно характеризовать в качестве автоматизированного центра принятия решений об использовании той или иной функции обработки данных из доступного множества таких функций. Такое решение может быть принято по каким-то надёжно фиксируемым или считываемым признакам ситуации или на основании структурированного опыта применения функций обработки данных к различным предметным областям (ПО) или их фрагментам.

Мудрость в понимании авторов – совокупный опыт принятия решений (об использовании функции или программы ЭВМ для обработки данных в ПО), увязанный с конечными последствиями этих решений: удалось ли решить вопрос или задачу, возникли ли при этом проблемы на практике, общая оценка помощи машинной обработки данных и др.

Предиктивные технологии – множество прогнозов и рекомендаций, направленных на воспроизведение ситуаций и действий, уже знакомых интеллекту и содержащихся в массиве накопленных знаний, в случае которых выработка эффективной реакции не несёт ничего нового и в рамках накопленного опыта стандартна. При этом прогнозы имитируют ещё не наступившие ситуации, а рекомендации – ещё не выполненные действия, направленные на распределение проявления этих ситуаций, обычно тех из них, которые наиболее выгодны с точки зрения ПО.

Отметим, что в сельском хозяйстве выдача прогнозов и рекомендаций – дискретный процесс. Например, на протяжении цикла возделывания сельскохозяйственной культуры существует множество моментов времени, когда результаты применения предиктивных технологий будут актуальными для практики. В каждый из этих моментов времени складываются более или менее стандартные ситуации и планируется более или менее стандартные действия, но совокупность случайных факторов (климат, погода, состояние сельскохозяйственной техники) существенно меняет и ситуации, и планируемые действия настолько, что малейшая ошибка весьма значительно отражается на урожайности.

С. Баскарада и Э. Корониос (2013) провели семиотическое теоретическое и эмпирическое исследования иерархии данных, информации, знания и мудрости и их качественного измерения [Baskarada, Koronios, 2013]. В частности, проанализировав литературные источники и семиотику Пирса, измененную американским философом Чарльзом У. Моррисом, они предложили так называемую семиотическую лестницу (табл. 3). Последняя строка в таблице добавлена авторами работы.

Известно, что Ч. Моррис (1938) расширил семиотику Пирса, сосредоточив внимание на отношениях между знаками и другими знаками, знаками и объектами, а также знаками и интерпретаторами [Morris, 1938]. В результате он предложил три дополнительных области исследования. Синтаксика (или синтаксис) – это исследование формальных или структурных отношений между знаками, т.е. логика и грамматика знаковых систем (имеет дело с физической формой, а не с содержанием/значением знаков). Семантика – это изучение отношений знаков к объектам, которые они обозначают. Семантика имеет дело с правилами такого состояния, при котором знаки применяют-

Семиотическая лестница (Baskarada, Koronios, 2013)

Уровень	Аспекты представительства
Social (общественный)	Убеждения, ожидания, обязательства, контракты, социальные законы, культура ...
Pragmatic (прагматический)	Намерения, общение, разговоры, переговоры, речевые акты ...
Semantic (семантический)	Значения, предложения, обоснованность, истина, значение, обозначения ...
Syntactic (синтаксический)	Формальная структура, язык, логика, данные, записи, дедукция, программное обеспечение, файлы ...
Empirical (эмпирический)	Структура, разнообразие, шум, энтропия, пропускная способность канала, коды, эффективность ...
Physical (физический)	Сигналы, физические различия, аппаратные средства, скорости, экономика, законы природы, ...
Automatic (машинный)	Автоматизированное управление, алгоритмы, программы ЭВМ, средства настройки, администрирования и программирования ...

ся к объектам. Прагматика – это изучение отношения знаков для интерпретаторов, прагматика связывает вопрос о знаках с вопросом о намерениях. Синтаксику, семантику и прагматику трудно, а вернее, невозможно рассмотреть в изоляции друг от друга.

В семантическую лестницу добавлены также еще три уровня – физический, эмпирический и социальный. Физический уровень использует методы физики и техники для моделирования сигналов, их источников, пунктов назначения и маршрутов, которые они передают. Эмпирический уровень имеет дело с кодированием и частотой сигнала, социальный – со значениями и действиями (Stamper et al., 2000).

Машинный уровень введен нами в семантическую лестницу в связи с тем, что семантика ИИ все-таки отличается от приведенных выше уровней тем, что связана с машинными языками программирования, автоматизированным управлением, алгоритмами, программами ЭВМ, средствами настройки, администрирования и др.

С. Баскарада и Э. Корониос (2013) предложили также определения для оценки качества данных, информации, знаний и мудрости (табл. 4).

Определения качества данных, информации, знаний и мудрости (Baskarada, Koronios, 2013)

Термин	Определение	Пример
Качество данных	Качество данных – это разрыв между физическими и техническими характеристиками знаков (чем меньше разрыв, тем выше качество)	Низкое качество данных: плохой почерк, подарпаннный CD/DVD, поврежденный компьютерный файл и др.
Качество информации	Качество информации – это информация соответствующая назначению	Если, при прочих равных условиях, следовать рецепту А и более эффективно производить вкусную еду, чем по рецепту В, тогда рецепт А представляет собой информацию более высокого качества.
Качество знаний	Качество знаний – это знание, пригодное для выполнения цели	Если, при прочих равных условиях, когнитивно-поведенческая терапия (КПТ) более эффективна для уменьшения фобий, чем психоаналитическая терапия, тогда КПТ составляет более качественные психотерапевтические знания.
Качество мудрости	Качество мудрости – это пригодность мудрости по назначению	Если, при прочих равных условиях, увеличение общественных расходов на образование приводит к более эффективному развитию общества, чем увеличение государственных расходов на правоохранительные органы, тогда растущие общественные расходы на образование составляют более качественную мудрость.

Совокупность качественных свойств знаний, которую в отечественной литературе впервые описал Д.А. Поспелов [Искусственный интеллект..., 1990], имеет в ИИ специфический смысл. Под знаниями понимается форма представления информации в ЭВМ, которой присущи такие особенности, как: а) внутренняя интерпретируемость (когда каждая информационная единица должна иметь уникальное имя, по которому система находит ее, а также отвечает на запросы, в

которых это имя упомянуто); б) структурированность (включенность одних информационных единиц в состав других); в) связность (возможность задания временных, каузальных пространственных или иного рода отношений); г) семантическая метрика (возможность задания отношений, характеризующих ситуационную близость); д) активность (выполнение программ инициируется текущим состоянием информационной базы). Именно эти характеристики, подчеркивал Д.А. Поспелов, отличают знания в ИИ от данных и определяют ту грань, за которой данные превращаются в знания, а базы данных перерастают в базы знаний.

В инженерии знаний различают также следующие основные понятия о знаниях, заимствованные из семиотики (науки о знаковых системах):

- экстенциональные знания – поверхностные или конкретные знания;
- интенциональные знания – глубинные или абстрактные знания (знания о закономерностях);

- синтаксис – структура знаковой системы (данных или знаний);
- семантика – смысл знаковой системы (знаний), т.е. эквивалентное ее представление в другой парадигме представления знаний (внутренней);

- прагматика – цели, связанные со знаковой системой (например, цели или назначение предложения на естественном языке – команда, вопрос, пояснение и т.п.).

Г.В. Рыбина (2010), ссылаясь на работы Т.А. Гавриловой и В.Ф. Хорошевского (2000), а также А.П. Частикова с соавт. (2003), для сравнения данных и знаний предлагает использовать способ, связанный с выделением основных форм (уровней) существования знаний и данных (табл. 5).

По мнению Л.А. Гладкова с соавт. (2009), понятие «знания» остается расплывчатым и неформальным. Под знаниями понимают нечеткие и четкие множества фактов и сведений, описывающих информацию относительно определенной предметной области, позволяющих решать определенный класс задач. Согласно О. Уотермену, 1989; А.Г. Теслинову, 1989; В.В. Васильковой, 1999; Л.А. Гладкову и др., 2009, выделяют следующие типы знаний:

- физические, знания об окружающей среде и внутренних свойствах объектов;

Основные формы существования знаний и данных (Рыбина, 2010)

Знания (Зн)	Данные (Д)
Зн ₁ – знания в памяти человека	Д ₁ – результат наблюдений над объектами или данными в памяти человека
Зн ₂ – материализованные знания (энциклопедии, учебники, справочники, словари и др.)	Д ₂ – фиксация данных на материальных носителях (таблицы, графики и др.)
Зн ₃ – поле знаний (структурированное, полуформализованное описание Зн ₁ и Зн ₂), модель представления знаний	Д ₃ – модель данных (некоторая схема описания, связывающая несколько объектов)
Зн ₄ – знания на языках представления знаний (формализация Зн ₃)	Д ₄ – данные на языке описания данных
Зн ₅ – база знаний в ЭВМ (на машинных носителях информации)	Д ₅ – база данных на машинных носителях информации
Выделяют три уровня: Зн ₁ (знания) > Зн ₃ (поле знаний) > Зн ₅ (база знаний – БЗ)	Выделяют три уровня: Д ₁ (внешний) > Д ₃ (логический) > Д ₅ (физический)

- предметные, знания о данной предметной области;
- синтаксические, знания о правилах построения рассматриваемых структур;
- семантические, знания о конкретном смысле и значении элементов в структуре;
- декларативные, знания о фактах и данных;
- процедурные, знания об алгоритмах, методах, эвристиках;
- метазнания, обобщенные знания;
- «Инь и Янь», знания о преобразованиях и переменных на основе принципов комплементарности (противоположности);
- расплывчатые (синергетические), знания об интеллектуальном опыте природы.

К этому перечню, по нашему мнению, следует добавить знания об условиях, времени и эффекте применения всех предыдущих знаний.

Основные характеристики знаний – это интерпретируемость, концептуальность, структурированность, именованность, иерархичность, активность, рефлексивность, связность, степень согласованности и готовность к практическому применению [Гладков и др., 2009].

По-видимому, понятие «знания» перестаёт быть расплывчатым только тогда, когда их приобретение, обработка и применение направлены на реализацию конкретной цели в рамках конкретной ПО. Цель данной книги – ответить на вопрос: как сложившиеся на настоящий момент представления учёных о структуре и свойствах данных, информации, знаний и мудрости могут помочь в разработке эффективных предиктивных технологий в сельском хозяйстве. Более узко: как могут быть детализированы каждый блок и каждая связь в представленной выше DIKW-модели (см. рис. 3), чтобы она превратилась из абстрактного указания в практическое руководство для создания базы комплектов данных и функций, обрабатывающих эти комплекты, с привязкой к реальным моментам в сельскохозяйственных процессах, когда необходимо использовать ИИ.

Многочисленные характеристики знания, его качества и структуры дают возможность изучать выбранную проблему с помощью DIKW-модели многослойно, в различных приближениях, с учетом интересов получателей результатов применения предиктивных технологий. Например, перечисленные 10 типов знаний могут быть выделены или из анализа действий и ситуаций, или путём преобразования интуитивного интеллектуального понимания в информацию с последующим её встраиванием в некоторую модель представления знаний. Если существуют эти 10 типов знаний, то существуют и специализированные способы их приобретения, накопления и представления. Ещё более специализированно они структурируются в конкретной ПО. Конечная стадия специализации – это инженерное владение знаниями: как, сколько и что именно нужно предоставить на вход некоторой стандартизированной функции обработки данных с целью получения настолько же детально известного комплекта данных и применить его для принятия решения или для входа другой функции.

1.2. Инженерия знаний

Теория ИИ характеризуется как «...наука о знаниях, о том, как их добывать, представлять в искусственных системах, перерабатывать внутри системы и использовать для решения задач» [Поспелов, 1986], а история ИИ, за исключением ранних этапов, – как история исследований методов использования (извлечения, накопления, структуризации, представления) знаний, называемых также инженерией знаний (knowledge engineering).

Инженерия знаний – область ИИ, связанная с разработкой экспертных систем и баз знаний. Определена Э. Фейгенбаумом и П. МакКордак в 1983 г. как: «...раздел (дисциплина) инженерии, направленный на внедрение знаний в компьютерные системы для решения комплексных задач, обычно требующих богатого человеческого опыта» (https://lrc-lib.ru/wikipedia/Инженерия_знаний).

В широком смысле инженерия знаний занимается интеграцией знаний с компьютерными системами для того, чтобы решать сложные проблемы, обычно требующие высокого уровня человеческой экспертизы, а также участия методов анализа данных с помощью ИИ. В основном инженерия знаний осуществляется с помощью взаимодействия инженера по знаниям (когнитолога) и эксперта в определенной предметной области в форме непосредственного живого общения [Гаврилова, Хорошевский, 2000]. Однако это не единственная форма извлечения знаний, хотя и довольно распространенная. В некоторых работах упоминается около 15 неавтоматизированных и более 20 автоматизированных методов приобретения и формирования знаний [Boose, 1989; Welbank, 1990].

По мнению Л.С. Болотовой (2012), «Тематика инженерии знаний, несмотря на кажущееся большое многообразие методов, остаётся наименее раскрытой. Среди них практически отсутствуют подходы, обеспечивающие единое и целостное представление как о самом процессе получения экспертных знаний, так и о способах представления его в рамках единого концептуального решения. Представляется, что причина такой ситуации в следующем:

– сложность получения знаний от лиц, владеющих ими, т.е. экспертов, поскольку они далеко не всегда в состоянии «выложить» его на лист и даже просто сформулировать необходимым образом;

– сами ЛПР (лицо, принимающее решение – *добавлено нами*) в большинстве своём не владеют методиками структуризации и перевода знаний из категории «скрытых» в категорию явных;

– инженеров по знаниям, владеющих методами и техниками выявления и проектирования нужных, в том числе новых знаний, воспитать в наших условиях очень сложно» [Болотова, 2012].

Следует отметить, что трактовка термина «знание» в эпистемологии и в ИИ различается. Так, Л.А. Микешина (2005) подчеркивает принципиальную разницу понятий «информация» и «знание», указывая, что упрощенно различие между ними можно свести к формуле: «информация – это знание минус человек... Информационная модель знания... является лишь намеком на представленное знание, по которому человек способен творчески воссоздать само знание» [Микешина, 2005]. Философы считают, что информация общедоступна, а знание всегда принадлежит индивиду и не может существовать отдельно от человека-субъекта. Информация объективно существует и свободно циркулирует в обществе, может быть зафиксирована на бумажном или электронном носителе и не связана с каким-либо определенным субъектом. Она доступна широкому кругу лиц, а знания являются достоянием их создателя. Информация может тиражироваться, издержки на производство ее новых копий невысоки, создание же новых знаний требует высокого интеллектуального уровня развития личности, а также существенных усилий и затрат. Информация – это сырье для мыслительной деятельности человека, порождающего знание. Воспринимая и перерабатывая информацию, человек увеличивает собственное знание [Алексеева, 1993; Красильников, 2012; Корниенко и др., 2013; Лысак, 2017].

Из опыта работы Лаборатории предиктивных агротехнологий с использованием методов искусственного интеллекта СФНЦА РАН (далее Лаборатория) в 2019, 2020 гг. и определения термина «знания», приведённого в последней строке табл. 1, сформировался механизм сохранения и накопления знаний, основанных на специальных таблично-списочных формах – шаблонах знаний.

Шаблон знаний – это форма-анкета для заполнения научным сотрудником или экспертом, отвечающая трём базовым свойствам:

– снабжена инструкцией заполнения, понятной человеку, ориентирующемуся в ПО;

- заполненные шаблоны знаний в совокупности должны исчерпывающе описывать ПО и все происходящие в ней процессы;
- заполненные шаблоны знаний в совокупности должны быть достаточны для экспертной программной реализации имитационной модели ПО.

В результате DIKIW-модель, приведённая на рис. 3, преобразуется в вид, изображенный на рис. 4.

Квадратом с закруглёнными краями выделен блок, обычно являющийся основой для большинства статей в области прогнозирования протекания процессов в сельском хозяйстве. О шести шаблонах знаний, разработанных в Лаборатории для сбора и накопления знаний в области растениеводства и земледелия, подробно рассматривается во второй главе.

К. Уиг (1995) считает, что «...знание состоит из истин и убеждений, перспектив и концепций, суждений и ожидания, методологии и ноу-хау» [Wiig, 1995], И. Нонака и Х. Такечи (1995) – «...знание является фактором производства» [Nonaka, Takeuchi, 1995].

Подход к знанию, применяемый в ИИ, И.Ю. Алексеева (1993) называет технологическим. Она предполагает конструирование и исследование

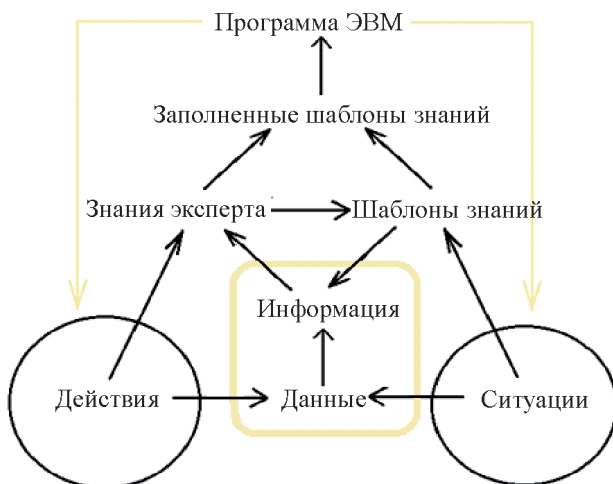


Рис. 4. Измененная схема формирования и функционирования DIKIW-модели

дование позитивных объектов, посредством которых знание фиксируется и передается. В качестве таковых выступают знаковые системы – модели и программы. С позиций технологического подхода к знанию особого внимания заслуживают вопросы экономичности представления знаний, эффективности используемых моделей в решении задач [Алексеева, 1993].

Извлечение знаний имело свое формальное начало в середине 1980-х годов в контексте инженерии знаний для экспертных систем [Экспертные системы..., 1987; Нейлор, 1991; Джарратано, Райли, 2007]. Экспертные системы – это компьютерные программы, которые воплощают предметно-специфические знания и выполнение которых (например, принятие решений, решение проблем, проектирование) осуществляется на уровнях, типичных для человека-эксперта (но не обязательно точно так же, как эксперты-люди). Инженерия знаний здесь в широком смысле определяется как процесс построения систем, основанных на знаниях [Cooke, 1999]. В структуре экспертных систем выделяют два основных компонента: базу знаний и систему логического вывода. В связи с этим возникает вопрос: присуще ли знание только человеку или его носителем могут быть интеллектуальные системы? Проведенный анализ показывает (см. табл. 1), что в большинстве работ понятия «информация» и «знания» определяются через информацию, а порой и просто отождествляются с ней. Например, Д.А. Пospelов (1990) под знаниями понимает форму представления информации в ЭВМ, которой присущи внутренняя интерпретируемость, структурированность, связность, семантическая метрика и активность [Искусственный интеллект..., 1990]. В.Н. Ручкин и В.А. Фулин (2009) считают, что «...знанием является информация, которая отражает объективные свойства и связи некоторых объектов, явлений, процессов, сущностей и отношений между ними как в субъективном, так и в научном (объективном) выражении» [Ручкин, Фулин, 2009, Болотова, 2012]. И.Г. Сидоркина (2015) под знаниями понимает «...представленная на естественном языке в соответствии с предметной областью совокупность данных, фактов и правил вывода, включающая информацию о свойствах объектов, закономерностях процессов и явлений, а также правилах использования этой информации в целях констатации фактов или принятия решений и реализующая свойства означивания, активности, интерпретируемости» [Сидоркина, 2015]. В канве

этого определения И.Г. Сидоркиной нами создавались шаблоны знаний в целях использования информации в интеллектуальных системах, представляющих собой множество функций обработки данных и множество комплектов данных, являющихся входами и выходами этих функций.

Как отмечает Г.Б. Бронфильд (2012), для описания знаний в ИИ давно сформировалась представление о двухуровневой структуре знаний, наиболее часто используемое в России [Попов, 1987; Гаврилова, Хорошевский, 2000; Вагин, 2002]. В ее основе находятся данные (понятия), т.е. отдельные факты, характеризующие объекты, процессы и явления предметной области, а также их свойства. В связи с этим знания – это хорошо структурированные данные, или данные о данных, или метаданные (можно идентифицировать как информацию). Есть традиционные представления знаний, использующих эту основу, в частности продукционные модели, семантические сети и фреймы. Однако постепенно стал развиваться более широкий спектр структур представления знаний и информации (например, DIKIW-модель). В частности, Г.Б. Бронфильдом (2012) предложено представление знаний и мудрости в виде путей на графе информации. По мнению автора, такой подход открывает новые возможности аналитического изучения знаний и информации, реализации новых подходов в создании интеллектуальных систем и делает более четкими и наглядными взаимосвязи между информацией и знанием [Бронфельд, 2012].

В ИИ обычно подразделяют знание на декларативное («знать, что»), к которому относятся констатация факта, комментирование события, и процедурное («знать, как») – порядок выполнения процедуры, определяемый располагаемыми данными в четко заданном формате. Трудности приобретения знаний в ИИ – это трудности изучения структуры экспертного знания и механизмов его функционирования. Кроме того, имеются сложности, связанные с вербализацией неявного знания (знания субъекта), «знания, как» [Алексеева, 1993].

В инженерии знаний процесс работы со знаниями обычно делится на четыре подпроцесса: извлечение и приобретение знаний, представление знаний и манипулирование знаниями [Искусственный интеллект..., 1990, 1990а; Гаврилова, Хорошевский, 2000; Башмаков, Башмаков, 2005; Палагин и др., 2012].

1. *Извлечение знаний* из различных источников, в том числе из естественно-языковых объектов, включает два основных раздела: анализ исходной информации, формализацию качественных знаний и интеграцию знаний. Первый раздел связан с созданием методов, позволяющих переходить от знаний, выраженных, например, в естественно-языковой форме, к их аналогам, пригодным для ввода в память информационной системы. Второй раздел связан с интеграцией знаний, получаемых от различных источников, в некоторую взаимосвязанную и непротиворечивую систему знаний о ПО.

Здесь уже нужно упомянуть технологию Big Data, потому что гораздо эффективнее не анализировать исходную информацию, а просто её складывать, причём складывать в соответствии с абстрактными объектами шаблонов знаний в виде вариантов их реализации. Тогда одновременно выполняется анализ, формализация и интеграция знаний. Из Big Data следует извлекать уже заведомо качественные данные – те, которые нужны для конкретной задачи.

Например, в области сельского хозяйства динамический контекст, сложность и местная специфика, а также многие задачи, которые сельскому хозяйству необходимо выполнять, требуют более гибких способов управления генерацией, интеграцией и обменом знаниями. Все заинтересованные стороны, включая собственно сельскохозяйственных производителей, должны быть признаны в качестве равных соавторов генерирования знаний, а все виды знаний, как формальных, так и неформальных, должны объединяться в инновационных процессах. Сети знаний, которые способствуют обмену знаниями, совместному обучению и выработке новых, более интегрированных решений, имеют решающее значение для обеспечения устойчивости сельского хозяйства [Klerkx, Jansen, 2010; Hermans et al., 2015; Ingram, 2018; Šūmane et al., 2018].

2. *Приобретение знаний*. Знаний, содержащихся в источниках информации, отчуждённых от специалиста, как правило, недостаточно. Значительную часть профессионального опыта эти специалисты не могут выразить словесно (профессиональное умение или интуиция). В связи с этим для того, чтобы приобрести такие знания, нужны специальные приёмы и методы. Они используются в инструментальных системах по приобретению знаний, создание которых – одна

из задач инженерии знаний. Полученные от экспертов знания нужно оценить с точки зрения их соответствия ранее накопленным знаниям и формализовать для ввода в память системы. Кроме того, знания, полученные от различных экспертов, необходимо согласовать между собой, так как нередки случаи, когда эти знания оказывались внешне несовместимыми и даже противоречивыми. Рассматриваемый под-процесс включает такие разделы как организация работы с экспертами, оценка и формализация знаний и согласование знаний.

В работе Б. Бьюкенена с соавт. (1983) приобретение знаний трактуется более широко. В частности, они предлагают выполнять анализ приобретения знаний в терминах модели процесса проектирования экспертной системы (рис. 5) [Buchanan et al, 1983].

Как видно из рис. 5, проектирование экспертной системы начинается с анализа ПО, а весь процесс состоит из пяти этапов.

Идентификация. Анализируется класс проблем, которые предполагается решать с помощью проектируемой системы, включая данные, которыми нужно оперировать, и критерии оценки качества решений. Определяются ресурсы, доступные при разработке проекта, – источники экспертных знаний, трудоемкость, ограничения по времени, стоимости и вычислительным ресурсам. Например, в задаче поддержки агротехнологий – это набор моментов вмешательства в ведение хозяйством процесса возделывания культуры и набор вопросов, которые в эти моменты задаёт агроном некоторой информационной экспертной системе. Количество моментов и количество вопросов могут отличаться, количество вопросов может быть существенно больше.

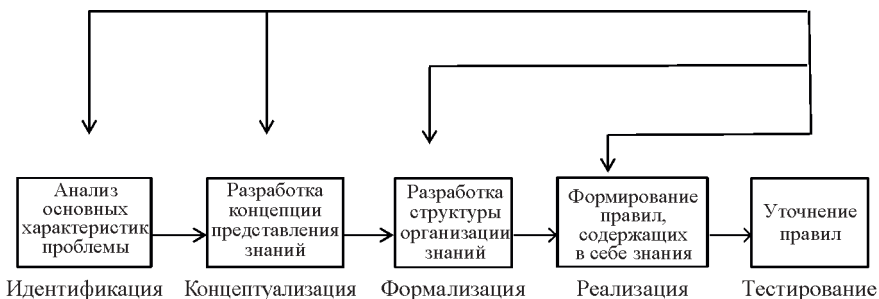


Рис. 5. Схема процесса приобретения знаний (Buchanan et al, 1983; цит. по Jackson, 1998)

Концептуализация. Формулируются базовые концепции и отношения между ними. Сюда же входят и характеристика различных видов используемых данных, анализ информационных потоков и лежащих в их основе структур в предметной области в терминах причинно-следственных связей, отношений частное/целое, постоянное/временное и т.п.

Формализация. Предпринимается попытка представить структуру пространства состояний и характер методов поиска в нем. Выполняется оценка полноты и степени достоверности (неопределенности) информации и других ограничений, накладываемых на логическую интерпретацию данных, таких как зависимость от времени, надежность и полнота различных источников информации.

Реализация. Преобразование формализованных знаний в работающую программу, причем, на первый план выходит спецификация методов организации управления процессом и уточнение деталей организации информационных потоков. Правила преобразуются в форму, пригодную для выполнения программой в выбранном режиме управления. Принимаются решения об используемых структурах данных и разбиении программы на ряд более или менее независимых модулей.

Тестирование. Проверка работы созданного варианта системы на большом числе репрезентативных задач. В процессе тестирования анализируются возможные источники ошибок в поведении системы. Чаще всего таким источником является имеющийся в системе набор правил. Оказывается, что в нем не хватает каких-то правил, другие не совсем корректны, а между некоторыми обнаруживается противоречие.

3. *Представление знаний.* Этот процесс предполагает разработку формальной научной теории, включающей построение модели знаний, системы представления знаний и базы знаний (БЗ). В инженерии знаний системы представления знаний включают совокупность процедур, необходимых для записи знаний, извлечение их из памяти и поддержки хранения знаний в рабочем состоянии. Системы представления знаний оформляются как БЗ, являющихся естественным развитием баз данных (БД). Представление знаний – это соглашение о том, как и в какой формальной теории описывать реальный мир. Обычно этот процесс осуществляется следующим образом. На естественном языке вводятся основные понятия и отношения между ними, но при этом используются ранее определённые понятия и отношения, смысл которых уже известен. Далее устанавливается соответствие между характеристиками (чаще всего количественными) понятий и подходя-

шей математической моделью. Основная цель представления знаний – строить математические модели реального мира и его частей.

Системой представления знаний называют совокупность средств, позволяющих:

- описывать знания о предметной области с помощью языка представления знаний, шаблонов знаний + UML;
- организовать хранение знаний в системе (накопление, анализ, структурное обобщение и организация знаний) Big Data или просто наборы данных, являющихся входными для конкретных моделей;
- выводить новые знания из имеющихся и объединять их;
- находить требуемые знания;
- обновлять знания через заполнение и корректировку шаблонов знаний, через уточнение и детализацию модели ПО;
- осуществлять интерфейс между системой и пользователем.

4. *Манипулирование знаниями.* К этому процессу относятся такие разделы как пополнение знаний, классификация знаний, обобщение знаний, вывод на знаниях (резольционные методы, квазиаксиоматические системы и системы правдоподобного вывода), рассуждения с помощью знаний, объяснения на знаниях, решения прикладных задач ПО.

В механизмах автоматизированного приобретения знаний выделяют три поколения. Это автоматизированные системы приобретения знаний от экспертов с использованием множества наработанных технологий и методологий [Бова и др., 2010]. Первое поколение – автоматизированные системы извлечения знаний без участия экспертов из существующих знаний и из текстов, массивов данных (Data mining, Knowledge Discovery in Databases) [Курейчик и др., 2008].

Технологии второго поколения начиная с 90-х годов XX в. представлены применением методологии KADS и характеризуются также дополнительными особенностями: глубокими знаниями и машинным обучением [Lavrac, Mozetic, 1989; Breuker, Wielinga, 1989].

В основе KADS лежит идея о том, что экспертная система является не контейнером, наполненным представленными экспертом знаниями, а «операционной моделью», которая демонстрирует некоторое нужное поведение в реальном мире. Приобретение знаний, таким образом, включает в себя не только извлечение специфических знаний о ПО, но и интерпретацию извлеченных данных применительно к некоторой концептуальной оболочке и формализацию их таким способом, чтобы программа могла действительно использовать их в процессе

работы [Wielinga et al., 1992; Schreiber et al., 1993; Jackson, 1998].

В основу оболочки KADS положены следующие базовые принципы:

- использование множества моделей, позволяющее преодолеть сложность процессов инженерии знаний;
- четырехуровневая структура для моделирования требуемой «экспертности» – набора качеств, лежащих в основе высокого уровня работы специалистов;
- повторное использование родовых компонентов модели в качестве шаблонов, поддерживающих нисходящую стратегию приобретения знаний;
- процесс дифференциации простых моделей в сложные;
- важность преобразования моделей «экспертности» с сохранением структуры в процессе разработки и внедрения.

Главным мотивом создания оболочки KADS было преодоление сложности знаний. В настоящее время у инженеров по знаниям имеется возможность использовать при построении экспертных систем множество самых разнообразных методов и технологий. Однако при этом остаются три основных вопроса [Jackson, 1998]:

- определение проблемы, которую необходимо решить с помощью экспертной системы;
- определение функций, которые возлагаются на экспертную систему применительно к этой проблеме;
- определение задач, которые необходимо решить для выполнения возложенной функции.

В подходе, который реализован при создании KADS, стадия «концептуализации» разбивается на две части: модель кооперации, или коммуникации, и модель «экспертности». Первая отвечает за декомпозицию процесса решения проблемы, формирование набора простейших задач и распределение их между исполнителями, в качестве которых могут выступать люди и машины. Вторая модель представляет процесс, который обычно называется извлечением знаний, т.е. анализ разных видов знаний, которые эксперт использует в ходе решения проблемы.

Кроме указанных, в состав оболочки KADS входит и модель проектирования, включающая технологии вычислений и механизмы представления знаний, которые могут быть использованы для реализации спецификаций, сформулированных предыдущими моделями. Представителем данной методологии являются оболочка CommonKADS (<https://commonkads.org/>).

Третье поколение – это более сложные, гибкие, использующие графические возможности современных рабочих станций, достижения программных средств CALS- и CASE-технологий. В зарубежной литературе активно обсуждается проблема машинной обработки больших данных (big data), которую иногда называют «четвертой парадигмой науки».

В 2009 г. была опубликована книга «The Fourth Paradigm: Data-Intensive Scientific Discovery» под редакцией Tony Hey, Stewart Tansley и Kristin Tolle, которая содержит серию эссе ученых и программистов, описывающих, как различные научные области трансформируются в результате экспоненциального увеличения научных данных [Hey et al., 2009]. По мнению Tony Hey, одного из редакторов книги, парадигмы науки во времени менялись следующим образом: 1) тысяча лет назад – Experimental Science – описание природных явлений; 2) последние несколько сотен лет – Theoretical Science – законы Ньютона, уравнения Максвелла и др.; 3) последние несколько десятилетий – Computational Science – моделирование сложных явлений; 4) в настоящее время – Data-Intensive Science – ученые перегружены данными из различных источников (данные, полученные с помощью приборов и инструментов, в результате моделирования, генерируемые сенсорными сетями и др.). В связи с этим созданы наборы инструментов и технологий eScienceis, способные поддерживать корпоративные базы и хранилища данных, развивать сотрудничество среди мирового научного сообщества для получения и анализа данных, их визуализации, научной коммуникации и распространения.

Для облегчения научных исследований (открытий) этими же авторами предложено решить задачу создания естественного пользовательского интерфейса (natural user interface – NUI), который дал бы возможность разработчикам программ и исследователям создавать интуитивно понятные, мультимодальные, ориентированные на пользователя, научные приложения. Научная система NUI (scientific NUI system) должна начать понимать намерения ученых, чтобы анализировать и фильтровать данные для получения соответствующего набора информации, действительно позволять осуществлять научные открытия. Кроме того, система должна быть способна представлять эту информацию в интуитивно понятном и персонализированном виде, а также в режиме «без потерь», который позволяет интерактивно запрашивать результаты [Tolle et al., 2011].

В статье К. Лиакоса с соавт. (2018) осуществлен обзор исследований, посвященных применению машинного обучения и обработки больших данных в сельскохозяйственном производстве. Исследования осуществляют по следующим направлениям:

- управление возделыванием сельскохозяйственных культур, включая приложения по прогнозированию урожайности, выявлению болезней, качеству урожая сельскохозяйственных культур и распознаванию видов сорняков и культурных растений;

- управление животноводством, включая знания о здоровье и состоянии животных;

- управление водными ресурсами;

- управление плодородием почв.

Применяя машинное обучение к сенсорным и иным данным, системы управления сельскохозяйственным предприятием (фермой) превращаются в программы с ИИ в режиме реального времени и предоставляют сельхозтоваропроизводителям (фермерам) обширные рекомендации и идеи для поддержки принятия решений и действий [Liakos et al., 2018]. В каждой из этих программ есть точки обращения к системе, набор задач, решаемых в каждой точке, а также случаи, когда при поступлении некоторых критических данных (или становящихся критическими в совокупности с уже существующими в системе) система сама выходит на связь с человеком, давая прогнозы и рекомендации.

Продолжая тему использования методов ИИ в сельском хозяйстве, нельзя не упомянуть серию статей в специализированном выпуске журнала «Agricultural Systems» за 2017 г. В журнале обсуждаются идеи создания нового поколения сельскохозяйственных систем, моделей и знаний под общим названием NextGen. В этих статьях подчеркивается, что необходимость создания нового поколения моделей сельскохозяйственных систем мотивируется несколькими соображениями. Во-первых, складывается консенсус в отношении того, что необходимо устойчивое и более продуктивное сельское хозяйство, способное решать местные, региональные и глобальные проблемы продовольственной безопасности XXI в. Данный консенсус предполагает, что необходимы новые и усовершенствованные инструменты, которые могут быть использованы для оценки устойчивости существующих и перспективных сельскохозяйственных систем. Эти отдельные, но взаимосвязанные проблемы в свою очередь создают

спрос на достижения в области аналитических возможностей имеющихся данных [Antle et al., 2017]. Во-вторых, существует большая и растущая база знаний о процессах, лежащих в основе сельскохозяйственных систем, на которых можно строить модели нового поколения [Jones et al., 2017]. В-третьих, быстрые достижения в области сбора данных и управления ими, моделирования, вычислительной мощности и информационных технологий дают возможность использовать эти знания новыми и более продуктивными способами для достижения эффективных и устойчивых сельскохозяйственных систем [Janssen et al., 2017].

Необходимость создания NextGen обосновывается еще и тем, что системы моделирования и поддержки принятия решений в настоящее время не успевают за технологиями получения данных в сельском хозяйстве (дистанционное зондирование, IoT-технологии, точное земледелие и др.). Многие фреймворки (frameworks), используемые в этих системах, датируются 1970–1990 гг., до появления современных передовых технологий сбора данных, вычислений, хранения, доступа, обработки, языков программного обеспечения и стандартов кодирования [Janssen et al., 2017].

Кроме того, одной из основных проблем применения моделей сельскохозяйственных систем в настоящее время признается то, что они разработаны, как говорится, «в академических кругах» для нужд исследователей, поэтому слабо ориентированы на конечного пользователя и мало используются в производстве. Ключевым новшеством нового поколения моделей должна быть более тесная связь с продуктами знаний, которые должны принять форму удобных для пользователей аналитических инструментов и мобильных приложений. Это позволит использовать модели и их результаты гораздо большим количеством заинтересованных сторон, чем это возможно сейчас [Antle et al., 2017a].

Инструментальными средствами для обработки знаний являются:

- алгоритмические языки высокого уровня (СИ+, Паскаль, ЛИСП, PROLOG, SMALLTALK, РЕФАЛ и др.);
- языки представления знаний (OPS 5, KRL, FRL и др.);
- программные среды (KE, ARTS, GURU, G2 и др.);
- специальные оболочки для создания экспертных систем (BUILD, EMYCIN, EXSYS Prof., ЭКСПЕРТ и др.).

1.3. Модели представления знаний

Модели представления знаний (МПЗ) начали разрабатывать с 70-х годов XX в. для описания семантики документов, баз данных и прикладных расчетно-логических задач. Различные подходы отличались собственной спецификой, более всего заметной на синтаксическом уровне формального инструментария описания смысла. Их применяли в вопросно-ответных [Белоногов, Кузнецов, 1983] и естественно-языковых системах [Искусственный интеллект..., 1990], а также в интеллектуальных решателях задач и системах управления [Минский, 1979; Пospelов, 1981; Нильсон, 1985]. Это направление получило название «системы представления знаний» (knowledge based systems) [Слэйгл, 1973; Гаврилова, Хорошевский, 2000] или «модели представления знаний» (knowledge representation models) [Бибило, Романов, 2011; Бова, 2014].

Одной из проблем при создании экспертных систем является проблема представления знаний для их переработки в машине. В настоящее время существуют различные способы представления знаний. По мнению А.А. Корниенко с соавт. (2013), системы представления знаний имеют особенность: именно эти системы моделируют деятельность человека, которая осуществляется в неформальном виде. МПЗ ориентированы на информацию, получаемую от экспертов. Эта информация часто носит противоречивый характер, но специфика ее применения требует, чтобы она была представлена в однозначном, формализованном виде. Последнего можно добиться посредством обращения к идеям многозначной логики, теории нечетких множеств или использованию математических моделей [Корниенко и др., 2013].

К наиболее известным и часто применяемым МПЗ относятся следующие:

- логическая, формально-логическая модель;
- модель представления знаний в виде семантической сети;
- продукционная модель;
- фреймовая модель.

Логическая, формально-логическая модель – знания, необходимые для решения, и сама решаемая задача описываются определенными утверждениями на логическом языке (в основном предикатов первого порядка). Знания составляют множество аксиом, а решаемая задача

представляет собой теорему, требующая доказательства. Процесс доказательства теоремы составляет логическую модель представления знаний.

В основе логических моделей представления знаний лежит понятие формальной системы. Формальная система (ФС) задается четверкой [Бондарев, Аде, 2002; Журавлёв и др., 2010]:

$$M = \{T, P, A, R\},$$

где T – множество базовых элементов (алфавит); P – множество синтаксических правил (формулы); A – множество аксиом; R – множество правил вывода.

Суть элементов, образующих ФС, заключается в следующем. Множество T состоит из конечного или бесконечного числа элементов различной природы. Элементы множества T – алфавит ФС, на основе которого строятся все остальные составные части ФС. На множество T никаких ограничений не накладывается. Важно только, чтобы для T существовала процедура проверки принадлежности некоторого элемента множеству T .

Множество синтаксических правил P позволяет строить из элементов T синтаксически правильные совокупности базовых элементов. На множество синтаксических правил также не накладывается особых ограничений. Требуется только, чтобы существовала конструктивная процедура, которая позволяла бы за конечное число шагов дать однозначный ответ на вопрос, является ли данная совокупность элементов из T синтаксически правильной. Такие совокупности называют правильно построенными формулами (ППФ).

Среди всех ППФ выделяют некоторое подмножество аксиом A . При этом должна существовать процедура, позволяющая для любой ППФ решить вопрос, является ли она аксиомой данной ФС.

Множество R – это конечное множество отношений между ППФ, называемых правилами вывода. Правило вывода – это отношение на множестве формул.

Применяя правила вывода к множеству аксиом A , можно получать новые ППФ, к которым можно опять применить правила из множества R . Это позволяет осуществлять вывод новых ППФ. Множество R называют также множеством семантических правил. Множество

ППФ, получаемое после применения правил к аксиомам, называется множеством семантически правильных совокупностей.

Логические модели представления знаний характеризуются хорошим математическим обоснованием, а также теоретической обоснованностью выводов, осуществляемых в системе. Цель их использования – дать базис для построения других методов представления знаний или для реализации языков программирования ИИ, таких как Prolog [Гаврилов, 2002; Братко, 2007]. Особое место среди них занимают методы для представления нечетких знаний: нечеткая логика, псевдофизические логики, лингвистические переменные. Наиболее развитые из этих моделей – псевдофизические логики [Поспелов, 1981, 1989; Иноземцев, 2014], в которых аккумулируются концепции нечетких множеств, лингвистической переменной и логики предикатов первого порядка.

Преимущества логических моделей, использующих язык логики предикатов, связаны с дедуктивными возможностями исчисления предикатов, теоретической обоснованностью выводов, осуществляемых в системе. Однако логические модели в сложных ПО или в соответствующих фрагментах знания могут оказаться слишком громоздкими и недостаточно наглядными, поэтому используются в основном в строго формализованных задачах [Поспелов, 1981; Братко, 2007].

Модель представления знаний в виде семантической сети. Сетевая модель имеет аналогом структуру долговременной памяти человека и пригодна для представления знаний любых типов. Она начала развиваться в виде семантических сетей с целью моделирования естественного языка. Одной из основополагающих теоретических работ по семантическим сетям считается труд американского психолога Р. Куиллиана о «семантической памяти» («Semantic memory. Semantic information processing» [Люггер, 2004]. В этой модели сетевая структура использовалась для представления семантических (смысловых) отношений между концептами (словами, обозначающими некоторые понятия ПО). Обычно используемая для моделирования знаний семантическая сеть состоит из двух множеств: узлов (понятий) и соединяющих узлы ребер (дуги), которые отражают отношения между ними. В качестве узлов (понятий) в модели используются объекты ПО, их свойства и значения, а также события, процессы, явления. Ре-

бра (дуги) могут иметь или не иметь тип, направление и количественную оценку отношений между узлами (понятиями). Главное преимущество сетевой модели – способность выражения достаточно тонких смысловых оттенков знаний [Слэйгл, 1973; Лозовский, 1984, 1990; Нильсон, 1985; Кузнецов, 1986; Загорулько, 1996].

Семантическую сеть C в общем случае можно определить как

$$C = (x_1, x_2, \dots, x_n, R_1, R_2, \dots, R_m),$$

где x_1, x_2, \dots, x_n – некоторые фиксированные множества объектов (содержимое понятий); R_1, R_2, \dots, R_m – система отношений на элементах этих множеств.

В семантической сети, содержащейся в разработанном нами Репозитории Знаний, представлены шесть множеств объектов: метакатегории, категории, страницы, сельскохозяйственные объекты, группы пользователей, пользователи.

В качестве понятий в семантической сети обычно выступают абстрактные или конкретные объекты (пашня, трактор, агротехнология, Петров). «Пашня» – в категориях, включённых в метакатегорию «Земля», «агротехнологии» – метакатегория, трактор – нижестоящая категория внутри метакатегории «агротехнологии».

Виды отношений: вложенность одной категории в другую, принадлежность страницы категории, горизонтальная связь между страницами (перекрёстная ссылка), пересечение метакатегорий и др.

В качестве отношений используются также следующие [Гаврилова, Хорошевский, 2000]:

– таксономические («класс – подкласс – экземпляр», «множество – подмножество – элемент» и др.). Данный тип отношения называют также отношением АКО (A Kind Of – является разновидностью), IS A (является, это есть) или гипонимии (гипероним – общая сущность; гипоним – частная сущность);

– структурные («часть – целое»). Данный тип отношения называют также отношением Part of (является частью), Has part (состоит из, включает в себя), агрегации (лат. *aggregatio* – присоединение), композиции (лат. *compositio* – составление, связывание, сложение, соединение) или меронимии (холоним – сущность, включающая в себя другие; мероним – сущность, являющаяся частью другой);

- родовые («предок» – «потомок»);
- производственные («начальник» – «подчиненный»);
- функциональные (определяемые обычно глаголами «производит», «влияет» и т.п.);
- количественные (больше, меньше, равно и др.);
- пространственные (далеко от, близко от, за, под, над и др.);
- временные (раньше, позже, в течение и др.);
- атрибутивные (иметь свойство, иметь значение);
- логические (И, ИЛИ, НЕ);
- казуальные (причинно-следственные).

Все эти виды отношений встраиваемы в сеть, например, можно сделать связи между сельскохозяйственными объектами с обозначением расстояния между ними. Интересный практический вопрос: к чему относить метеостанции? Ясно, что к метакатегории «Внешний источник данных». А к метакатегории «агротехнологии» или «цифровизация» или к их потомкам метеостанции имеют отношение? Вообще, каков порядок определения места любого нового понятия в иерархической структуре семантической сети? Этот порядок должен быть разработан при участии эксперта. Когда-нибудь само отнесение нового понятия должно стать автоматическим или полуавтоматическим.

По нашему мнению, порядок должен быть примерно такой:

- определяем одну и только одну метакатегорию, к которой данное понятие относится;
- смотрим во вложенных категориях этой метакатегории, где находится место новому понятию в иерархическом дереве этих категорий;
- по ходу описания нового понятия, т.е. при создании посвящённой ему страницы в Репозитории Знаний, связываем с другими понятиями, неважно из каких категорий или метакатегорий;
- отвечаем на вопрос, может ли данное новое понятие относиться к одной или нескольким категориям вне иерархического дерева выбранной на первом шаге метакатегории; если ответ «да», то привязываем понятие к этим категориям;
- далее нужно проследить связи нового понятия до одного или нескольких сельскохозяйственных объектов.

Например, если это метеостанция, то нужно провести цепочку связей до тех сельскохозяйственных объектов, при анализе задач на кото-

рых будут использоваться данные именно этой метеостанции (в этом, кстати, заключается формирование прецедента по шаблонам знаний).

Отношения также классифицируются по степени участия (арности) понятий в отношениях:

- унарное (рекурсивное) связывает понятие само с собой;
- бинарное связывает два понятия;
- N-арное связывает более двух понятий.

В зависимости от типов связей, используемых в модели, различают классифицирующие сети, функциональные сети, сценарии. В классифицирующих сетях используются отношения структуризации. Такие сети позволяют в БЗ вводить разные иерархические отношения между информационными единицами. Функциональные сети характеризуются наличием функциональных отношений. Их также называют вычислительными моделями, поскольку они позволяют описывать процедуры «вычислений» одних информационных единиц через другие. В сценариях используются каузальные отношения, а также отношения типов «средство – результат», «орудие – действие» и др.

Эти отношения находятся в разрабатываемой нами интеллектуальной системе «защиты» в шаблонах знаний и соответственно выяснять данные отношения в нашей семантической сети можно будет через метакатегории шаблонов знаний. Преимущество такого подхода в том, что это легко реализовать программно, на практике. Что касается теории, то можно (скорее всего, с помощью интуиционистской логики) доказать теорему: семантическая сеть (имеющая отношение к сельскому хозяйству) с произвольным количеством различных отношений может быть встроена в иерархическую (таксономическую) с фиксированным количеством корневых элементов (подразумевается, что через грамотно выстроенные шаблоны знаний можно задавать любые отношения). Или ещё короче: берём любое отношение в сельскохозяйственной семантической сети и утверждаем, что это отношение может быть преобразовано в иерархическое отношение дерева категорий шаблонов знаний.

Если в сетевой модели допускаются связи различного типа, то ее называют семантической сетью.

Известны несколько топологий сетей. Деревья – самый распространенный вид иерархии – граф с одной вершиной. Правила тополо-

гии для данного графа: вершина графа представляет собой один общий тип, каждый другой узел имеет лишь одного родителя. Ациклический граф – это граф без циклов, имеющий ветви, которые расходятся и опять сходятся вместе, при этом узел графа может иметь несколько узлов-родителей. Решетка – в отличие от деревьев узлы в решетке могут иметь несколько узлов родителей. Ограничения: любая пара типов X и Y , как минимум, должна иметь общий гипертип (родителя) и подтип (потомка). Вследствие этого решетка выглядит, как дерево, имеющее одну главную вершину, которая является гипертипом всех категорий, и другую вершину – подтип всех узлов.

Различным модификациям семантических сетей присущ ряд сходств, а именно: узлы семантических сетей представляют собой концепты (понятия) предметов, событий, состояний; различные узлы одного концепта относятся к различным его значениям; дуги семантических сетей определяют отношения между узлами-концептами (пометки над дугами указывают на тип отношения); отношения между концептами представляют собой временные, пространственные, логические отношения, также лингвистические падежи, такие как агент, объект, реципиент, отношения между предложениями и другое; концепты организованы по уровням иерархии в соответствии со степенью обобщенности.

В настоящее время наиболее известными семантическими сетями, например в области обработки естественного языка, являются сети WordNet и PyTez, построенные на основе формализации человеческого восприятия окружающего мира [Лукашевич, 2011]. В семантической сети WordNet понятия сформированы на основе отношения синонимии, поэтому называются синсетамы (от англ. set of synonyms – «множество синонимов», сокр. synset) [Fellbaum, 1998]. Эти сети используются в качестве «золотого стандарта» для оценки систем автоматической обработки естественного языка и различных языковых ресурсов. Они служат в качестве общепринятого эталона, с которым выполняется сопоставление нового метода или языкового ресурса на основании некоторой заданной количественной меры качества.

Другим примером семантической сети служит Semantic Web, предложенная в 2001 г. В ней документы доступны в форме, удобной для чтения как человеком, так и машиной в виде автоматических агентов [Berners-Lee et al., 2001]. В настоящее время данный подход реализо-

ван лишь отчасти в виде форматов и публикуемых машиночитаемых метаданных документов. Это привело к возникновению технологии поиска сущностей или семантического поиска, когда на странице результатов поиска представляются фактографическая информация об организации, событии или личности, которой был посвящен поисковый запрос, см. Google Knowledge Graph [Singhal, 2012].

Продукционная модель как способ представления знаний предложена Э. Постом в 1943 г. Основу этой модели составляет множество так называемых продукционных правил, которые применяют в тех областях, где значительная часть знаний основана на опыте специалистов, эмпирических закономерностях и ассоциациях [Уотермен, 1989; Представление..., 1989; Поспелов, 1990]. Модель базируется на логике «из ситуации следует действие», представляет систему правил вида «Если..., то...» и обеспечивает способ представления выводов, рекомендаций, указаний и стратегий. Этот тип МПЗ получил свое название в связи с тем, что он основан на конструкциях, позволяющих использовать сочетания уже известных фактов (образующих ситуации или условия) для получения новых знаний или ситуаций. Правила, на основании которых такие преобразования в конкретной ПО имеют смысл, называются «производящими правилами» или «продукциями».

Продукция представляет собой обусловленную опытом и (или) знаниями абстрактных истин причинно-следственную конструкцию, состоящую из двух частей. Первая часть – это некоторая ситуация в ПО, описываемая с помощью абстрактных или конкретных знаний. В продукции она играет роль утверждения или условия, снабженного некоторыми оценками истинности его компонентов. Эту часть продукции принято называть антецедентом. Антецедент в продукции играет роль посылки, из которой на основании истинных знаний (фактов), имеющихся в памяти системы, следует вторая часть продукции – вывод, обусловленный истинностью посылки. Вторая часть продукции называется консеквентом. Консеквент как следствие посылки может представлять собой новую ситуацию или действие, приводящее к изменению прагматики, парадигматики или синтагматики имеющихся знаний о сущностях ПО.

В общем виде под продукцией понимается следующее выражение [Васильев, Чернов, 2010]:

«(*i*); *Q*; *P*; *C*; $A \Rightarrow B$; *N*»,

где (*i*) – имя (метка) продукции; *Q* – сфера применения, вычлняющая из предметной области некоторую ее часть, в которой знание, заключенное в продукции, имеет смысл; *P* – предусловие, содержащее информацию об истинности данной продукции и ее приоритетности, используемую в стратегиях управления выводом при выборе данной продукции для исполнения; *C* – условие применимости ядра продукции. Обычно *P* представляет собой логическое выражение. Когда *P* принимает значение «истина», ядро продукции активизируется. Если *P* ложно, то ядро продукции не может быть использовано; $A \Rightarrow B$ – ядро продукции (интерпретация ядра может быть различной, например: «Если *A* истинно, то *B* истинно», «Если *A* имеется в базе знаний, то *B* надо внести в базу знаний», «Если *A* текущая ситуация, то надо делать *B*» и т.п.); *N* описывает постусловия продукции. Они актуализируются только в том случае, если ядро продукции реализовалось. Постусловия продукции описывают действия и процедуры, которые необходимо выполнить после реализации *B*. Например, после внесения удобрений необходимо в базе удобрений уменьшить их количество на складе.

Продукционная модель используется в деятельности Лаборатории, которая представлена как задачи в виде направленного дерева с комментариями. Проект-менеджер – ответственный человек за реализацию проекта – должен держать в голове все знания о проекте в виде продукционной модели. Иногда проект-менеджер применяет какие-то программные пакеты, папки, блокноты, но это только дополнительные средства для того, чтобы удержать модель в памяти. В сельском хозяйстве по этой модели много сделано, например технологические карты, по которым можно смело возделывать культуры в тепличных условиях без воздействия случайных факторов, что повсеместно и происходит. Осуществляется продукционная модель, искажаемая регулярными случайными факторами.

Если в памяти системы хранится некоторый набор продукции, то они образуют систему продукции. Система продукции – это совокупность множества продукции, выполнение которых задается с помощью стратегии управления выводом. Стратегия определяется машиной вывода:

$$M = \langle M_1, M_2, M_3 \rangle,$$

где M_1 формирует набор продукции, для которых выполнены условия их применимости; M_2 производит выбор из этого набора той, которая будет выполняться на данном шаге процесса, M_3 вносит необходимые изменения в продукции, входящие в систему, на основании информации, которая содержалась в постуловии продукции, выполненной на данном шаге процесса.

В производственных системах обычно выделяют следующие компоненты:

- неструктурированная или структурированная БД;
- некоторое число производственных правил или продукции;
- интерпретатор, который последовательно определяет, какие продукции могут быть активированы в зависимости от условий, в них содержащихся; выбирает одно из применимых в данной ситуации правил продукции; выполняет действие из выбранной процедуры.

Правилом называется упорядоченная тройка множеств $\Pi = \{C, A, D\}$, где C – условие правила; A – множество добавляемых правилом фактов; D – множество удаляемых правилом фактов. Фактами являются атомарные формулы исчисления предикатов первого порядка без свободных переменных. Что касается правил, то в них атомарные формулы со свободными переменными из множеств C, A и D превращаются в факты в процессе применения правил, т.е. в результате выполнения соответствующих подстановок (m_1, m_2, \dots, m_n) на места свободных переменных (x_1, x_2, \dots, x_n) и проверки для каждой формулы $P(x_1, x_2, \dots, x_n)$ из C выполнения условия $(m_1, m_2, \dots, m_n) \in I(P)$, т.е. выполнимости в текущем состоянии рабочей памяти [Осипов, 2011].

Производственные модели получили широкое распространение благодаря таким достоинствам, как простота формулировки отдельных правил, пополнения и модификации, а также механизма логического вывода. В качестве недостатка производственного подхода отмечают низкую эффективность обработки информации при необходимости решения сложных задач.

Для разработки производственных моделей используют алгоритмические языки PROLOG, CLIPS, OPSS, EXSYS, GURU, Кара, а также инструментальные системы KEE, ARTS, PIES и др.

Фреймовая модель. В 1974 г. М. Минский (1979) в книге «Фреймы для представления знаний» изложил теорию фреймов – дал понятие базовой структуры представления знаний фреймами и управления выводом, описал специфику универсального фреймового языка. Фреймы интерпретированы им как сравнительно большие единицы представления знаний, раскрыта их иерархическая структура, учитывающая степень абстракции; проанализированы и возможности представления комбинации декларативных и процедурных знаний [Минский, 1979]. М. Минский (1979) суть теории фреймов интерпретировал следующим образом: попадая в новую ситуацию, радикально меняя свое отношение к текущим обстоятельствам, человек вызывает из памяти структуру-фрейм. Фрейм являет собой единицу представления знания, заполненную в прошлом. Однако ее детали можно изменить, ориентируясь на происходящую ситуацию. Фрейм представляет собой сеть, состоящую из нескольких вершин и отношений: фиксированная информация об истинном состоянии объекта, а также терминальные слоты (терминалы), которые следует заполнить конкретными знаниями и данными. Каждый фрейм может быть дополнен информацией о способах его применения, последствиях этого применения, действиях, которые надо выполнить в определенных ситуациях.

Фрейм объединяет в себе декларативную информацию о свойствах и структуре объекта, процедурную – операции и действия, реализующие реакции фрейма на внешние воздействия, и целевую составляющую, определяющую назначение фрейма и применимость его в той или иной ситуации. В системе всегда существует активный фрейм, который в данный момент контролирует взаимодействие с внешней средой и отвечает за восприятие информации. Новый фрейм активизируется с наступлением новой ситуации и передачей ему управления от текущего. Каждый фрейм имеет уникальное имя и состоит из слотов, которые делятся на информационные, их значения описывают объект или ситуацию реального мира, отображаемую фреймом, и служебные.

Каждый слот в качестве значения может содержать не только числовую или символьную константу, но и функцию, правило, с помощью которого можно найти значение слота, приказ на выполнение

программы, ссылку на другие фреймы и слоты того же фрейма и др. Благодаря этому организуется сеть фреймов, в явном виде вводится иерархическая структура, основанная на различной степени абстракции характеристик объектов и ситуаций.

Фреймы подразделяются на два типа [Макаренко, 2009]:

- фреймы-прототипы (или классы), хранящиеся в базе знаний;
- фреймы-примеры (или экземпляры), которые создаются для отображения реальных фактических ситуаций на основе поступающих знаний.

Фреймы-прототипы используются для порождения фреймов-примеров.

В качестве примера можно рассмотреть фрейм для понятия «внесение»:

Фрейм «Внесение»	Слоты фрейма: (субъект, X_1) (объект, X_2) (место, X_3) (время, X_4) (условие, X_5)
------------------	---

В этом фрейме указаны имена слотов (субъект, объект и т.д.), но вместо их значений стоят переменные (X_1 , X_2 и т.д.). Такой фрейм называется фреймом-прототипом, или протофреймом.

Протофреймы хранят знания о самом понятии. Например, понятие «внести» связано с наличием слотов с указанными именами. Внесение осуществляет X_1 в месте X_3 во время X_4 , если выполнено условие X_5 . Берет X_1 нечто, обозначенное как X_2 .

Подставляя вместо всех переменных конкретные значения, получим конкретный факт-описание:

Фрейм «Внесение»	Слоты фрейма: (субъект, <i>механизатор</i>) (объект, <i>удобрение</i>) (место, <i>поле севооборота № 3</i>) (время, X_4) (условие, <i>план внесения удобрений, сроки и дозы</i>)
------------------	--

Фреймы, в которых обозначены все основные слоты (они неким образом помечаются в описании фрейма), называются фреймами-экземплярами, или экзофреймами. В UML это называется экземплярами

класса. В случае адаптации UML под сельскохозяйственные задачи конкретные экземпляры класса участвуют в диаграммах прецедентов. Диаграммы прецедентов строят на базе диаграмм объектов, объект – это совокупность последовательностей, последовательности составляются из методов, методы строятся по связям, классам и атрибутам. Поскольку в состав фрейма могут входить слоты с именами действий, фреймы годятся для представления как декларативных, так и процедурных знаний. В качестве значения слота может выступать новый фрейм, что позволяет на множестве фреймов осуществлять иерархическую классификацию. Диаграмма классов – это очень удобное свойство фреймов, так как человеческие знания, как правило, упорядочены по общности.

Важнейшим свойством теории фреймов является заимствованное из теории семантических сетей наследование свойств. И во фреймах, и в семантических сетях наследование происходит по АКО-связям (A-Kind-Of = это). Слот АКО указывает на фрейм более высокого уровня иерархии, откуда неявно наследуются, т.е. переносятся, значения аналогичных слотов [Гриф, Гениатулина, 2009]. Например, из фрейма «Сельскохозяйственная культура» явно наследуются определенные характеристики, относящиеся к определенной культуре (пшеница, рожь, овес, кукуруза и др.), но не всем этим характеристикам могут быть присвоены значения. В данном случае можно также сказать, что приведено описание класса UML с набором атрибутов (или протофрейма с набором слотов).

Преимущества фреймовых моделей заключаются в их экономичности, позволяющей сократить время автоматизированного поиска информации, и удобстве для описания определенных областей знания (соответствующих фрагментов реальности, изучаемых в данных областях), когда выделяются основные (с точки зрения задач, для которых создается интеллектуальная система) объекты ПО и (или) система понятий, в которых будут анализироваться конкретные ситуации, а также описываются свойства объектов (понятий) и отношения между ними. Это свидетельствует о преимуществах использования шаблонов знаний. И хотя в целом для этих типов моделей существуют значительные проблемы с организацией вывода, фреймовые системы многими оцениваются как перспективные благодаря возможностям

подведения под них достаточно строгих логических и математических оснований [Корниенко и др., 2013].

В нашем случае математическое основание можно подвести, доказав, что система шаблонов знаний (или система выбранных абстрактных объектов UML) является конкретной алгебраической системой в виде многомерного многообразия с конкретными операциями.

Фреймовая модель хорошо согласуется с возможностями объектно-ориентированных языков программирования как в части описания объектов, так и средств описания процедур обработки таких описаний. Однако существуют и специальные средства для программирования фреймовых моделей. Это языки FRL (Frame Representation Language), KRL (Knowledge Representation Language) и фреймовая «оболочка» Карра.

Интегрированная модель представления знаний. По мнению Ю.А. Загорулько (2013, 2015), использование каждой МПЗ по отдельности (семантические сети, фреймы и продукции) имеет ограниченное значение для адекватного отражения ПО и создания БЗ. В связи с этим инструментальные средства разработки интеллектуальных систем (ИС) должны включать различные взаимодополняющие друг друга методы и средства представления и обработки знаний. При этом возможны разные способы интеграции таких средств, начиная с интеграции на уровне подсистем и компонентов ИС и заканчивая интеграцией на уровне единого формализма.

В связи с этим Ю.А. Загорулько считает, что необходимо создавать такие МПЗ, которые обладали бы набором методов и средств, достаточными для представления в ИС всех необходимых знаний. Такие модели должны обеспечивать представление как декларативных, так и процедурных знаний, что является необходимым требованием, так как без представления этих двух типов знаний невозможна разработка полноценной ИС. Кроме того, они должны обеспечивать представление и обработку как точных, так и неточных данных, что сделает их пригодными для решения задач в условиях неопределенности знаний и данных [Загорулько, 2013, 2015].

Автором предложена интегрированная модель представления знаний (ИМПЗ-модель), которая в его нотации представлена в следующем виде [Загорулько, 2013]:

$$M = \langle ONT, SN, FN, PS, IP, NF \rangle,$$

ONT – прикладная онтология, описывающая основные понятия и отношения (классы объектов и отношений) предметной и проблемной областей приложения; *SN* – объектно-ориентированная семантическая сеть (ОО-сеть), классы объектов и отношений которой, как и их свойства, определяются онтологией *ONT*; *FN* – функциональная сеть, включающая локальные вычислительные модели, заданные на значениях атрибутов объектов семантической сети *SN*; *PS* – система правил-продукций, описанных в терминах классов и отношений онтологии *ONT*; *IP* – средства императивного программирования; *INF* – общий механизм вывода/обработки знаний и данных, представленных в ИМПЗ-модели, объединяющий механизмы вывода/обработки каждой из интегрируемых моделей.

Существуют также и другие теоретические концепции представления знаний. К ним относятся:

– концептуальные сети – попытка создать диалоговую программу для общения с машиной на естественном языке (ЕЯ) – разработчик Р. Шенк с учениками [Шенк, 1980]. Программа реализована на языке LISP в виде трех основных функций: 1) семантическая трансляция – преобразование предложения на естественном языке в соответствующую концептуальную модель; 2) концептуальная память – манипулирование концептуальными структурами, соответствующее «человеческим» мыслительным операциям; 3) концептуальная генерация – преобразование концептуальной структуры в текст на естественном языке. Это достаточно простой алгоритм, и вот как видится его реализация: по текстовой формулировке задачи выделить все встречающиеся в тексте слова, отделить из них и проигнорировать служебные, далее зайти в БЗ и представить дерево (граф) из встречающихся слов со всеми связями. Такая функция фактически уже есть в Википедии. После получения этого дерева необходимо довести его до абстрактного объекта UML «объект», далее этот объект превратить в прецедент, т.е. привязать к конкретным условиям (конкретному сельскохозяйственному объекту). Однако это всё возможно, когда будет реализовано приложение, и это приложение будет сопровождать постоянная группа программистов или хотя бы один программист;

– аппарат недоопределенных типов данных и разработанный на его основе метод недоопределенных вычислительных моделей дает возможность оперировать неточно заданными значениями и частично описанными объектами [Нариньяни, 1986; Нариньяни и др., 1998];

– в рамках парадигмы программирования в ограничениях знания можно задавать в виде системы ограничений на значения параметров моделируемых объектов [Ушаков, Телерман, 2000; Apt, 2003];

– дескрипционные логики и построенные на их основе языки и системы представления знаний [Baader et al., 2003].

– ситуационное управление – метод управления сложными техническими и организационными системами, основанный на идеях теории ИИ: представление знаний об объекте управления и способах управления им на уровне логико-лингвистических моделей, использование обучения и обобщения в качестве основных процедур при построении управления по текущим ситуациям, использование дедуктивных систем для построения многошаговых решений [Поспелов, 1986];

– универсальный семантический код (УСК) – его называют также языком описания ситуаций. Простой ядерной конструкцией в УСК является тройка вида SAO. В этой тройке S соответствует субъекту, совершающему акцию А, а О – объекту, на который направлена данная акция. В УСК вводится замкнутая система операций, позволяющая из простых ядерных конструкций строить более сложные цепочки [Мартынов, 2001];

– растущие пирамидальные сети (РПС) – ациклические ориентированные графы с вершинами истоками (рецепторами) и вершинами-концепторами. Концепторы могут описывать понятия на основе разнообразных свойств (значений признаков), задаваемых рецепторами как источниками данных, и формируются в процессе работы сети. В начале работы в РПС есть только рецепторы. При появлении конкретных признаков рецепторы переходят в «состояние возбуждения». Этот процесс распространяется по сети вверх. Концептор активизируется, если все его признаки «возбуждены» [Гладун, 2004]. Может достаточно легко реализоваться на базе одной из моделей семантической сети. Осуществляется проход по сети и на основе входных данных задачи ставится признак возбуждения, если процент возбуждения

превышает заданный, то концептор активируется. Концепторы – метакатегории, рецепторы – сельскохозяйственные объекты и связанные с ними пользователи;

– признаковые структуры (ПС) – формализм, объединяющий фреймовый и логический подходы в рамках иерархии вложенных кортежей. Для ПС в отличие от логических термов снимаются следующие ограничения: 1) структуры, размеченные символически, не используют явного указания позиций аргументов; 2) для структур не требуется указывать фиксированное число составляющих их аргументов; 3) устраняются различия между идентификаторами функций и аргументов; 4) переменные и взаимные ссылки аргументов обрабатываются отдельно [Knight, 1989];

– онтологии для описания семантики – концептуальные модели ПО, использующие одну или несколько таксономий и включающие описание понятий, отношений, свойств понятий и ограничений на значения свойств. Онтологии являются одним из современных средств описания семантики ПО (в том числе для описания информационных ресурсов), в котором использованы основные результаты известных ранее моделей знаний. Получили широкую популярность у разработчиков баз знаний (БЗ), с помощью которых можно достаточно просто представить согласованную систему понятий моделируемой области [Guarino, 1998; Гаврилова и др., 2006; Abdul-Ghafour et al., 2007; Каширин, Каширин, 2010; Загорулько, Загорулько, 2011; Скобелев, 2012; Бова и др., 2015].

1.4. Базы знаний

Центральным звеном инженерии знаний является создание БЗ, которое в общем случае заключается в извлечении экспертных знаний и их представление в оптимальной модели. Для извлечения знаний используются коммуникативные и когнитивные методы, с помощью которых инженер по знаниям помогает эксперту формализовать необходимые сведения. Наиболее важный вопрос при создании БЗ – выбор МПЗ о свойствах сущностей ПО и отношениях (связях) между ними. Эта модель должна определять не только структуры информации различных уровней, но и обеспечить их максимальную адекватность структуре внутренних операций компьютера и структуре язы-

ков программирования, используемых для реализации модели. При этом нельзя оставлять без внимания и такое важное условие, как максимальное соответствие МПЗ характеру задач, для решения которых создается система.

БЗ – совокупность программных средств, обеспечивающих поиск, хранение, преобразование и запись в памяти ЭВМ сложно структурированных информационных единиц (знаний) [Толковый словарь..., 1992].

Считается, что БЗ – это особый тип моделей ПО, характеризующихся высокой степенью субъективности [Макаренко, 2009]. Если проводить аналогию с классическим математическим моделированием, то классические модели с той или иной степенью точности отражают свойства и отношения непосредственно самой ПО (модели первого типа), тогда как БЗ – это чаще всего отражение того, как ПО формируется в голове эксперта (модели второго типа) [Аршинский, 2016]. На это двойное отражение дополнительно накладываются особенности восприятия данного отражения инженером по знаниям. Следует вывод, что БЗ – это достаточно субъективная конструкция.

Помимо знаний и представлений эксперта, полученных непосредственно от него самого или из написанных им трудов, в настоящее время в качестве источника знаний выступают «большие данные» (Big data). БЗ, построенные на этой основе, ближе к моделям первого типа, однако и здесь сведения, предлагаемые системами анализа данных, обрабатываются специалистами и сверяются с их субъективным представлением о ПО [Аршинский, 2016]. Субъективность БЗ порождает ряд особенностей. Это в первую очередь известные НЕ-факторы А.С. Нариньяни [Нариньяни, 2008]: неопределенность, недостоверность, нечеткость, неполнота и др. Установлено, что в ходе общения инженера по знаниям с экспертами в среднем теряется до 76% информации. Кроме того, ошибки человека-эксперта при внесении информации в БЗ в среднем равны 15%. Влияние этих факторов порождает фундаментальную проблему качества БЗ [Нитежук, 2018].

Надо иметь в виду, что при создании БЗ могут возникнуть сложности, к которым относятся противоречивость, неполнота, избыточность, наличие нигде не используемых терминов. Это порождает про-

блемы использования БЗ [Долинина, 2010; Поспелова, 2013, 2014]:

- невозможность получить достоверные заключения;
- наличие неопределенных понятий и терминов, невозможность вывода в определенных ситуациях;
- наличие неиспользуемых фактов, правил, терминов;
- наличие понятий, совпадающих по смыслу с введенными ранее.

В связи с этим важным этапом разработки БЗ является процесс верификации – доказательства того, что она не содержит перечисленных ошибок. Важнейшей частью верификации при этом является проверка БЗ на непротиворечивость. Данный тип ошибки способен привести к проблемам логического вывода, а значит – к невозможности использования БЗ в качестве модели ПО [Нитежук, 2018].

Основными технологическими этапами создания БЗ являются [Гаврилова, Хорошевский, 2000; Дубровин, 2008; Бычков и др., 2016; Юрин, 2016]:

- идентификация проблемной области (ПрО);
- концептуализация ПО;
- формализация БЗ.

Идентификация ПрО. Цель этого этапа – определить, для решения какой задачи (или задач) создается БЗ. Идентификация задачи заключается в составлении ее неформального (вербального) описания, в котором указываются:

- общие характеристики задачи (содержательная постановка задачи, исходные данные);
- подзадачи, на которые может быть разбита задача;
- знания, необходимые для решения задачи;
- предполагаемое содержание и форма представления результата решения задачи.

В процессе идентификации задач инженер по знаниям и эксперт работают в тесном контакте. Исходное неформальное описание задачи, данное экспертом, используется инженером по знаниям для уточнения терминов и ключевых понятий. Эксперт корректирует описание задачи, объясняет, как решать ее и какие рассуждения лежат в основе того или иного решения. После нескольких циклов, уточняющих описание, эксперт и инженер по знаниям составляют окончательное неформальное описание задачи.

Концептуализация ПО. На данном этапе проводится содержательный анализ ПрО, выявляются используемые понятия и их взаимосвязи, определяются методы решения задач. Этот этап завершается созданием ядра модели ПО, включающего основные концепты и отношения между ними. Существует два подхода к процессу построения концептуальной модели ПО.

Признаковый, или атрибутивный, подход предполагает представление полученной от экспертов информации в виде элементарных триад «идентификатор объекта – название признака (атрибута) объекта – значение признака (атрибута)». Если идентификатор объекта (его наименование) является уникальным для данной ПО, то никакой другой информации (кроме элементарных триад, в которых признаками объектов могут быть и их связи с любыми другими объектами) для описания ПО не понадобится. Если это условие не соблюдается (наименование неуникально), то потребуется дополнительная информация для того, чтобы отличать, например объекты, имеющие одинаковые наименования. Такая дополнительная информация может иметь характер обучающей, если она представляет собой прецеденты (примеры) правильных экспертных заключений на предмет сходства или различия одноименных объектов ПО. Этот подход развивается в рамках направления, получившего название «машинное обучение» (machine learning).

Структурный (когнитивный) подход основан на выделении сущностей ПО по семантике или характеру отношений между ними. Этот подход предполагает выделение следующих когнитивных элементов знаний: 1) понятий, 2) взаимосвязей понятий, 3) метапонятий, 4) семантических отношений.

Выделяемые понятия ПО должны образовывать систему, которой свойственны:

- отсутствие избыточности;
- полнота описания сущностей (процессов, фактов, явлений и т.д.) ПО;
- достоверность (соответствие выделенных единиц смысловой информации их реальным наименованиям);
- непротиворечивость (отсутствие омонимии).

Важной фазой концептуального анализа ПО является установление семантических отношений между понятиями и метапонятиями. Установить семантические отношения – значит определить специфи-

ку взаимосвязи, выявленной в результате применения того или иного метода анализа. Для этого необходимо каждую выявленную взаимосвязь осмыслить и отнести к одному из типов отношений, отражаемых в модели ПО.

Существует около 200 базовых типов таких отношений: «часть – целое», «род – вид», «причина – следствие», пространственные, временные и др. (см. раздел 1.3). Для каждой ПО, помимо базовых отношений, могут существовать и уникальные для некоторых ее сущностей.

Формализация БЗ. На этом этапе все ключевые понятия ПО и отношения между ними выражаются на некотором формальном языке, который или выбирается из числа уже существующих, или создается заново. Определяются состав средств и способы представления декларативных и процедурных знаний. Итогом этапа является формальное описание ПО (декларируемых в БЗ фактов), а также формальное описание постановки задачи и процедур, с помощью которых эта задача будет решаться системой.

Приведем в качестве примера БЗ небольшой фрагмент из экспертной системы MYCIN. Эта компьютерная программа, основанная на правилах, разработана в 1970-х годах в Стэнфордском университете для диагностики и рекомендации терапии при серьезных инфекциях. Подробнее об этой системе можно посмотреть по адресу: <http://samoychiteli.ru/document23918.html>, или в книге «Computer-based medical consultations: MYCIN. – Elsevier, 2012».

БЗ этой системы организована в виде множества правил в форме: если условие 1 и... и условие 2 удовлетворяются, то прийти к заключению 1 и... и к заключению *n*.

Эти правила преобразованы в операторы языка LISP.

Перевод на обычный язык типичного правила MYCIN выглядит следующим образом:

ЕСЛИ

- 1) микроорганизм обладает грамотрицательной окраской, и
- 2) микроорганизм имеет форму палочки, и
- 3) микроорганизм аэробный,

ТО

есть основания предполагать (0.8), что этот микроорганизм относится к классу Enterobacteriaceae.

Эти правила названы оргправилами (ORGRULES) и в них сконцентрированы знания о таких микроорганизмах, как *Streptococcus*, *Pseudomonas* и Enterobacteriaceae. Правило говорит о том, что если микроорганизм имеет форму палочки, пятнистую окраску и активно развивается в среде, насыщенной кислородом, то с большой вероятностью его можно отнести к классу Enterobacteriaceae. Число 0.8 называется уровнем соответствия (tally) правила, т.е. мерой правдоподобия заключения, сделанного на основании сформулированных условий.

Каждое правило такого вида можно рассматривать как представление в машинной форме некоторого элемента знаний эксперта. Возможность применить правило определяется тем, удовлетворяются ли в конкретной ситуации условия, сформулированные в первой его части. Сформулированные условия также носят нечеткий характер и могут удовлетворяться с разной степенью истинности. Поэтому в результате импортирования правил из БЗ применительно к конкретной ситуации формируется более общее правило, включающее и оценки уровня истинности соблюдения условий. В этом случае степень уверенности, связанная с каждым заключением, является функцией от оценок истинности соблюдения условий и уровня соответствия, отражающего степень уверенности эксперта при формулировке первичных правил.

Фактически правило является парой «предпосылка – действие», такое правило называют «продукцией». Предпосылка – это совокупность условий, уверенность в достоверности предпосылки зависит от того, насколько достоверна оценка условий. Условия – это предположения о наличии некоторых свойств, которые принимают значения «истина либо ложь» с определенной степенью достоверности. Примером может служить условие в приведенном выше правиле: «организм имеет форму палочки».

Действие – это или заключение, или рекомендация о том, какое действие предпринять. Примером заключения может служить вывод, что данный организм относится к определенному классу.

Помимо правил, в БЗ MYCIN также хранятся факты и определения. Для их хранения используются разные структурные формы:

- простые списки, например списки всех микроорганизмов, известных системе;

– таблицы знаний с записями об определенных клинических показателях и значениях, которые эти показания имеют при разных условиях; примером может служить информация о форме микроорганизмов, известных системе;

– система классификации клинических параметров соответственно контексту, в котором эти параметры рассматриваются, например, являются ли они свойством (атрибутом) пациентов или микроорганизмов.

Значительная часть знаний хранится не в виде правил, а в виде свойств, ассоциированных с 65 клиническими параметрами, известными программе MYCIN. Например, форма – это атрибут микроорганизма, который может принимать самые разнообразные значения, например «палочка» или «коккон». Система также присваивает значения параметрам и для собственных нужд – или для упрощения мониторинга взаимодействия с пользователем, или для индексации при определении порядка применения правил.

Особо следует остановиться на создании нечеткой базы знаний (НБЗ), поскольку при решении задач в сельском хозяйстве разработчики БЗ сталкиваются с проблемой неопределенности, неточности и др. НБЗ (их также называют БНЗ – база нечетких знаний и БНП – база нечетких правил) называется совокупность продукционных правил, которые описывают взаимосвязь между входами и выходами с использованием нечетких термов «низкий», «средний», «высокий» и т.п. При этом термы формализуют нечеткими множествами [Заде, 1976]. Идентификацию, в результате которой получают модель зависимости в форме НБЗ, обычно называют нечеткой. При использовании НБЗ этап структурной идентификации заключается в определении входных и выходных переменных моделей, иерархической организации входных переменных, формировании терм-множеств лингвистических переменных и описании зависимости нечеткими продукционными правилами [Ротштейн, 1999; Штовба, 2007; Штовба и др., 2016].

Первые две процедуры являются общими для любого метода идентификации, последние две – специфическими для нечеткой идентификации. В результате структурной идентификации получается грубая модель, которая в общих чертах описывает исследуемую зависи-

мость. На этапе параметрической идентификации настраивают модель – изменяют функции принадлежности нечетких термов, а также веса правил в НБЗ Мамдани и коэффициенты в заключениях правил БЗ Сугено [Штовба, 2007; Терновой, Штогрина, 2015; Штовба и др., 2016]. Качество НБЗ оценивают по критериям точности, компактности и интерпретабельности [Cordon, 2011; Штовба и др., 2012].

С.Д. Штовба с соавт. (2016) предлагают технологию создания НБЗ, состоящую из четырех этапов: 1) генерации списка адекватных правил; 2) селекции правил; 3) редукции antecedентов отобранных правил; 4) параметрической настройки НБЗ (настройка функций принадлежности и весов правил). Первые три этапа соответствуют структурной идентификации, четвертый – параметрической. Критерии точности и интерпретабельности задействованы на первом и четвертом этапах, критерии точности и компактности – на втором и третьем. Также предлагаются новые модели сохранения интерпретабельности НБЗ во время параметрической идентификации. Как подчеркивают авторы, компьютерные эксперименты на 7 задачах из UCI Machine Learning Repository Калифорнийского университета в Ирвине показали, что предлагаемая информационная технология синтезирует НБЗ с показателями точности и компактности на уровне конкурентных технологий. При этом предлагаемая технология обеспечивает интерпретабельность НБЗ [Штовба и др., 2016].

Под интерпретабельностью авторы понимают возможность содержательного объяснения структуры и параметров модели. Нечеткая модель является интерпретабельной, если выполняются следующие условия [Штовба, 2007а]:

- БЗ не является противоречивой или избыточной, т.е. не содержит правил с одинаковыми antecedентами;
- БЗ согласована с количеством термов, т.е. каждый терм фигурирует хотя бы в одном нечетком правиле;
- для произвольного входного вектора на выходе получаем непустое нечеткое множество;
- изолированно функция принадлежности каждого терма содержательно интерпретируется, т.е. соответствующее нечеткое множество является нормальным и выпуклым;
- каждое терм-множество содержательно интерпретируется.

В русскоязычной научной литературе работ по проблеме создания БЗ в сельском хозяйстве насчитываются единицы. Одной из первых работ в этом направлении были исследования А.Ф. Петрушина (2005) в кандидатской диссертации «Комплекс программ формирования и обработки баз данных и знаний в агрономии». В результате выполненных исследований получены следующие результаты [Якушев, Петрушин, 2000; Якушев и др., 2003; Петрушин, 2005]:

- разработана алгоритмическая основа создания программного комплекса формирования проблемно-ориентированных БД и БЗ по комплексированию и синтезированию агротехнологий и обоснована его реализация на языке Delphi 6;

- обоснована целесообразность разработки и создания специализированной системы управления разнородными атрибутивными данными, обеспечивающая также эффективную среду формализации, хранения и извлечения знаний;

- разработана и создана подсистема формирования и управления пространственно-атрибутивными данными с навигационным геоинформационным обеспечением;

- предложен интеллектуальный редактор управления агротехнологическими знаниями, которые описываются специально разработанными шаблонами, обеспечивающими электронное представление, формализацию и эффективную обработку. Интеллектуальный редактор функционирует в двух режимах. В режиме «эксперт БЗ» происходит формирование базовых агротехнологий и технологических адаптеров в базу знаний для хранения, в режиме «пользователь БЗ» производится настройка на ту или иную схему обработки знаний;

- программный комплекс обеспечивает комплексирование и синтезирование различных вариантов агротехнологий возделывания сельскохозяйственных культур. Количество этих вариантов зависит от содержания хранимых в БЗ правил выбора и дифференциации агроприемов, их ресурсного обеспечения, почвенно-климатических особенностей сельскохозяйственных полей, вида культур (сортов) и уровня программируемого урожая. Если в шаблоне описания технологических операций определен критерий её выбора или определены правила выбора других компонент агроприема, то автоматически синтезируется отвечающая этим требованиям выходная информация.

При этом генерация агротехнологий может быть решена в пакетном режиме для всех полей хозяйства или для заданного конкретного поля (соответствующий порядок определяется интеллектуальным редактором в режиме «пользователь БЗ»).

В АФИ продолжают исследования по проблеме создания систем поддержки принятия решений (СППР) в точном земледелии. Формализация в электронном виде накопленных в агрономии знаний – создание БЗ – признается важнейшим вопросом [Якушев, 2016; Якушев и др., 2018]. Первым этапом при создании БЗ является извлечение знаний из различных источников и их структурирование. Этот процесс включает в себя изучение различных источников информации, в том числе сюда входит эмпирический опыт экспертов в ПО. Вторым этапом является описание «поля знаний» – абстрагированное описание основных сущностей и взаимосвязей между понятиями ПО, выявленных из системы знаний эксперта в виде графа или онтологии верхнего уровня, таблицы или текста. Для приведения знаний к единообразной форме представления принято использовать стандартные онтологии верхнего уровня (онтология единиц измерений, геоонтология, онтология основных формальных понятий и т.п.).

Заключительным этапом обработки знаний является создание БЗ, которая может состоять из онтологий, продукционных правил, а также математических моделей, определенным образом встроенных в СППР. В АФИ разработано специализированное программное обеспечение (приведённое выше), позволяющее описывать агротехнические операции, используя базовые агротехнологии и технологические адаптеры, а также эмпирический опыт экспертов. Описание сущностей, входящих в агротехническую операцию, а также их свойств происходит в специальном программном модуле «глоссарий агротехнологий» [Якушев и др., 2018].

Н.В. Бендик и О.В. Борхошкин (2019), сотрудники Иркутского ГАУ им. А.А. Ежовского, создали БЗ для сельскохозяйственных товаропроизводителей Иркутской области. В разрабатываемой БЗ представление информации осуществляется при помощи правил продукционной МПЗ, т.е. определяются посылки «если» и заключения «то». БЗ содержит интеллектуальный редактор, интерфейс пользователя, графический интерфейс, базы текстов и графических изображений.

Предложена структура БЗ, а также описаны объекты базы. К основным разделам растениеводства отнесены землепользование, сельскохозяйственные культуры, повышение продуктивности земли. К объектам животноводства относятся животные и птицы, виды животных, кормление, содержание, болезни, планирование продуктивности животных, продукция животноводства и ее обработка. Помимо этого, выделены разделы, которые авторы будут детализировать в дальнейшем: экономика сельского хозяйства, техника и оборудование, строительство в сельском хозяйстве. Описана область применения разрабатываемой БЗ, а также дальнейшие этапы проектирования [Бендик, Борхошкин, 2019].

Функционирует несколько информационных ресурсов в свободном доступе в Интернет, которые можно использовать для нужд сельского хозяйства.

БЗ «Агрос» – крупнейшая в АПК документографическая база знаний. Объем базы составляет более 1 200 000 записей. Видовой состав документов: статьи из сериальных изданий, статьи из разовых сборников, материалы конференций, книги, авторефераты диссертаций, нормативно-технические документы, неопубликованные переводы, депонированные рукописи. В БЗ содержатся знания по следующей тематике:

- сельское хозяйство (растениеводство; биология сельскохозяйственных растений; защита растений от вредителей, болезней и сорняков; почвоведение; земледелие; сельскохозяйственная мелиорация; агрохимия; животноводство; биология сельскохозяйственных животных; ветеринария; механизация и электрификация; экономика и организация сельского хозяйства и проблемы АПК; охота и охотничье хозяйство; охрана окружающей среды в условиях сельскохозяйственного производства);

- лесное хозяйство;

- рыбное хозяйство;

- строительство в сельском хозяйстве;

- пищевая промышленность;

- домоводство (включая ремесла и промыслы, народную медицину, кулинарные рецепты).

Описания документов приводятся на языке документа или на английском языке. Все зарубежные документы сопровождаются обогащенным переводом основного заглавия на русском языке. Часть документов дополнена рефератами.

БЗ «Мелиорация и орошаемое земледелие» призвана способствовать укреплению в определенной степени регионального сотрудничества путем оказания информационной поддержки внедрению инструментов Интегрированного управления водными ресурсами (ИУВР). В состав базы включены объекты, с которыми приходится сталкиваться водникам в их повседневной практической деятельности по управлению водными ресурсами.

БЗ «Экология» включает информацию о книгах, авторефератах диссертаций, статьях из сборников научных трудов и периодических изданий по следующим темам:

- загрязнение и охрана атмосферы, вод, суши, почв и недр;
- влияние промышленности, сельского и лесного хозяйства на окружающую среду;
- производство экологически чистых продуктов, альтернативные методы ведения хозяйства, органическое земледелие, экологически безопасные технологии и др.

БЗ Союза органического земледелия, находящаяся по адресу: <https://soz.bio/baza-znaniy/>, содержит рубрики «Селекция», «Питание», «Защита», «Техника», «Вебинары», «Законы», «Стандарты». Способствует распространению знаний об экологических технологиях возделывания сельскохозяйственных культур.

Платформа Direct. Farm – информационный ресурс наподобие Википедии с вкраплением полезных сервисов. Задача платформы – предоставить фермеру возможность оперативного получения необходимых сведений и знаний, поделиться собственным опытом, оптимизировать работу при помощи современных цифровых технологий (<https://ftimes.ru/312825-onlajn-servis-znayushhij-vsyo-o-selskom-hozyajstve.html>)

Глобальный онлайн доступ к исследованиям по сельскому хозяйству можно осуществить через AGORA (Access to Global Online Research in Agriculture – AGORA), программу, которая бесплатно или

за очень незначительную цену предоставляет доступ к большинству научных журналов в сфере сельского хозяйства, биологических, экологических и социальных наук государственным институтам в развивающихся странах. AGORA запущена в октябре 2003 г., предоставляет доступ к 15 500 журналам от научных мировых институтов-лидеров. Целью AGORA, которая осуществляется под эгидой ФАО ООН, является улучшение качества и эффективности исследований в области сельского хозяйства и образования в странах с низким доходом и в свою очередь улучшение продовольственной безопасности (<http://www.fao.org/agora/background/ru/>).

В рамках AGORA осуществляется Программа Research4Life – государственно-частное партнерство ВОЗ, ФАО, ЮНЕП, ВОИС, Корнельского и Йельского университетов, Майкрософт и более 185 научных издательств из Международной ассоциации научных, технических и медицинских издательств. Целью партнерства является сокращение разрыва в научных знаниях между промышленно развитыми и развивающимися странами. С 2002 г. благодаря четырем программам Research4Life более 8500 организаций из более 115 развивающихся стран и территорий бесплатно или за низкую плату получили доступ к более 20 000 рецензируемых международных журналов, книг и баз данных, предоставленных ведущими мировыми научными издателями.

Существует несколько специализированных редакторов БЗ и оболочек экспертных систем (Expert System Designer, Expert System Creator, ARITY Expert Development Package, CxPERT, Exsys Developer и др.), которые позволяют реализовать формализованное описание понятий ПО и структур БЗ на определенном языке представления знаний.

1.5. Управление знаниями

Управление знаниями (УЗ) – область теоретической и практической деятельности, начавшая формироваться в начале 90-х годов XX в. и получившая наибольшее распространение в формате развития экономики знаний. Термин «knowledge management – управление знаниями» ввел К. Уиг (1995) [Wiig, 1995]. Термин обозначает систематическое формирование, обновление и применение знаний в целях повышения эффективности работы предприятия (организации, фир-

мы). Данное направление возникло в связи с проблемами обработки информации в крупных корпорациях, которые стали критическими для их развития. Узким местом совершенствования производства была обработка знаний, накопленных специалистами компании, так как именно знания обеспечивают конкурентное преимущество производственных технологий и продуктов, эффективность и выживаемость организации.

УЗ трактуется в научной литературе как совокупность процессов, которые управляют созданием, распространением, обработкой и использованием информации внутри предприятия [Гаврилова, Хорошевский, 2000; Мильнер, 2003; Тузовский и др., 2005; Тугускина и др., 2019; Baskerville, Dulipovici, 2006; Asrar-ul-Haq et al., 2016]. По нашему мнению, проблему УЗ следует рассматривать шире, чем только как средство эффективного менеджмента конкретного предприятия.

На Западе проблемы УЗ разрабатываются довольно активно, поскольку это направление считается важной областью исследований информационных систем и практической деятельности. Данная область опирается на теоретические основы информационной экономики, стратегического управления, организационной культуры и структуры, организационного поведения, искусственного интеллекта (ИИ), управления качеством и измерения эффективности деятельности организации. Эти теории используются в качестве основы для новых концепций, которые дают обоснование для УЗ, определяют процесс УЗ и позволяют оценивать результаты данного процесса. Появляются новые концепции, в том числе экономика знаний, альянс знаний, культура знаний, организация знаний, инфраструктура знаний и объективность (equity) знаний [Lee, Choi, 2003; Baskerville, Dulipovici, 2006; Alavi et al., 2005; Singh, 2011; Tzortzaki, Mihiotis, 2014].

Следует иметь в виду, что в УЗ осуществляется избирательный подход к освоению знаний в отличие, например от библиотковедения, которое пытается организовать все знания и информацию. Акцент на эффективное принятие решений квалифицирует УЗ как дисциплину действий, ориентированную на передачу знаний туда, где они могут быть применены. УЗ трактуется также как цикл получения, хранения, оценки, распространения и применения знаний. В

конечном счете, есть две основные миссии УЗ в качестве актива для улучшения процессов в организации:

- использование того, что организация «знает», чтобы она могла лучше использовать свои активы знаний;

- подключение генераторов знаний, обладателей и пользователей для облегчения потока знаний через организацию [Jennex et al., 2014].

М. Дженнекс с соавт. (2009) предполагает, что успех КМ можно рассматривать как способность организации эффективно использовать активы знаний для влияния на производительность и повышения общей эффективности и результативности [Jennex et al., 2009]. Т. Давенпорт с соавт. (1998) определили четыре цели для проектов, основанных на знаниях: создание хранилищ знаний, улучшение доступа к знаниям, улучшение среды знаний и управление знаниями как активом. Проекты УЗ успешны, когда есть рост ресурсов, вложенных в проект, рост содержания знаний, вероятность того, что проект выживет без поддержки конкретного человека, и некоторые доказательства финансовой отдачи. Факторы, которые приводят к успеху проектов, связанных с УЗ, включают гибкие структуры знаний, дружественную к знаниям культуру, ясные цель и язык, а также многочисленные каналы передачи знаний [Davenport et al., 1998].

При разработке систем УЗ выделяют следующие этапы [Гаврилова, Хорошевский, 2000]:

- *накопление информации*. Как правило, это стихийный и бессистемный процесс. Информация при этом имеет произвольную форму (таблицы, графики, формулы, словесные описания, рисунки и др.), различные источники формирования (системы мониторинга, эксперты, специалисты) и носители информации (бумажные, фото, видео, электронные). Задача заключается в том, чтобы привести в этот процесс формализованные процедуры, сделать его упорядоченным, управляемым и эффективным;

- *извлечение знаний*. Это один из наиболее сложных и трудоемких этапов, требующий высокой эрудиции и креативности от разработчика. От успешности этого этапа зависит дальнейшая жизнеспособность (адекватность модельных представлений, эффективность функционирования) системы. В качестве инструментов извлечения

знаний, как правило, используют SWOT, PEST-анализы, когнитивный и морфологический анализы, методы математического моделирования и др. В настоящее время недостаточно или вообще не используются в менеджменте сельскохозяйственных предприятий инструменты ИИ. Эти инструменты позволяют автоматизировать процесс извлечения знаний (обеспечить полноту исследования, необходимую скорость, высокое качество, устойчивость процесса);

– *структурирование знаний*. На этом этапе должны быть выделены основные понятия исследуемой предметной области, выработана структура представления информации, обладающая максимальной наглядностью, простотой изменения и расширения. В качестве наглядного инструментария на этом этапе выступает морфологический анализ. Морфологический анализ – это метод систематизации всех теоретически возможных решений, основанный на анализе структуры объекта;

– *формализация и программная реализация системы управления знаниями*. Данный этап предполагает представление структурированной информации в форматах машинной обработки данных, т.е. на языках описания данных и знаний, организацию автоматизированной обработки и поиск информации. Характерным примером является OLAP-куб. Роль и место механизмов и методов ИИ на этом этапе основополагающие. Существующие системы УЗ относятся к пассивным – работают по запросу ЛПР-руководителя (менеджера).

М. Оливейра с соавт. (2014) проанализировали большое количество мировой литературы и классифицировали УЗ (knowledge management – КМ) по практикам (practice) и технологиям (technologies). Разница заключается в использовании технологий: практики могут использовать технологии, но данные технологии могут осуществляться и без технической поддержки. Например, неформальный разговор может быть лицом к лицу, но также может быть и по телефону. Наиболее распространенные практики УЗ приведены в табл. 6 [Oliveira et al., 2014].

Распространенные практики УЗ [Oliveira et al., 2014]

KM practice	Authors
1	2
Практические сообщества – люди, объединенные общими интересами и опытом для обмена знаниями	[Aurum <i>et al.</i> 2008; Benbya & Belbaly 2005; Brown & Duguid 2001; Chua & Goh 2008; Kankanhalli <i>et al.</i> 2003; O’ Sullivan 2007; Plessis 2008; Saito <i>et al.</i> 2007; Wenger <i>et al.</i> 2002]
Неформальный разговор – разговор между сотрудниками	[Hansen <i>et al.</i> 1999; Jasimuddin 2007; Kankanhalli <i>et al.</i> 2003; Thomas <i>et al.</i> 2001]
Встречи и телефонные звонки – разговор между сотрудниками лицом к лицу или на расстоянии с использованием некоторой технологии	[Hansen <i>et al.</i> 1999; Jarrar 2002; Jasimuddin 2007; Kankanhalli <i>et al.</i> 2003; Thomas <i>et al.</i> 2001]
Мозговой штурм – неформальные встречи, предназначенные для генерирования идей с целью решения проблем или определения возможностей	[Hansen <i>et al.</i> 1999; Kankanhalli <i>et al.</i> 2003]
Лучшие практики – действия или методы, принятые фирмой, чтобы найти лучший способ сделать что-то	[Benbya & Benbya 2005; Bollinger & Smith 2001; Jarrar 2002; Saito <i>et al.</i> 2007]
Креативные помещения – пространство в фирме, позволяющее сотрудникам быть изобретательными и инновационными	[Curado & Bontis 2010]
Извлеченные уроки – регулярные встречи для обсуждения успехов и неудач в отношении процесса и продукта, обеспечение обучения и выявление уроков, которые могут быть полезны в других ситуациях	[Benbya & Benbya 2005; Bollinger & Smith 2001; Davenport <i>et al.</i> 1998; Jarrar 2002; Saito <i>et al.</i> 2007]
Наставничество – человек с большим опытом работы в фирме способствует личному развитию людей с меньшим опытом работы в фирме	[Bollinger & Smith 2001; Hansen <i>et al.</i> 1999; Henriques & Curado 2009; Lawrence 2008]
Организационный бюллетень/газета – частный бюллетень для внутреннего распространения	[Curado & Bontis 2010]

1	2
Мобильность персонала – перемещение людей между офисами, командами и деятельностью	[Hansen <i>et al.</i> 1999; Kankanhalli <i>et al.</i> 2003]
Рассказывание правдивых или вымышленных историй	[Kaye & Jacobson 1999]
Работа в команде – деятельность, проводимая в команде	[Bollinger & Smith 2001; Coakes <i>et al.</i> 2008; Saito <i>et al.</i> 2007]
Индивидуальная или групповая тренировка либо лицом к лицу, либо на расстоянии	[Aurum <i>et al.</i> 2008; Hansen <i>et al.</i> 1999; Jarrar 2002; Kankanhalli <i>et al.</i> 2003; Saito <i>et al.</i> 2007]

Технологии УЗ, наиболее распространенные для применения, определенные в литературе, перечислены в табл. 7 [Oliveira *et al.*, 2014].

Таблица 7

Распространенные технологии УЗ [Oliveira *et al.*, 2014]

KM technologies	Authors
1	2
Блог – электронный дневник, опубликованный в сети	[Sullivan 2007]
Электронный дискуссионный форум – позволяет людям публиковать сообщения и комментировать другие сообщения	[Bollinger & Smith 2001; Hansen <i>et al.</i> 1999; Kankanhalli <i>et al.</i> 2003; O’Sullivan 2007]
Е-mail – асинхронный обмен сообщениями в электронном виде	[Bollinger & Smith 2001; Jasimuddin 2007; Kankanhalli <i>et al.</i> 2003; Saito <i>et al.</i> 2007]
Экспертные системы – программное обеспечение, которое пытается дать ответ на проблему	[Aurum <i>et al.</i> 2008; Benbya & Benbya 2005]
Мгновенный обмен сообщениями – синхронный обмен сообщениями от людей, подключенных к Интернету	[Jasimuddin 2007; Saito <i>et al.</i> 2007]

1	2
Инtranет – частная сеть, использующая интернет-протокол	[Benbya & Benbya 2005; Bollinger & Smith 2001; Chua & Goh 2008; Jarrar 2002; Jasimuddin 2007; O’-Sullivan 2007; Saito <i>et al.</i> 2007; Skok & Kalmanovitch 2005]
Репозиторий – информационная система для хранения и распространения знаний	[Aurum <i>et al.</i> 2008; Benbya & Benbya 2005; Chua & Goh 2008; Chua 2004; Davenport <i>et al.</i> 1998; Hansen <i>et al.</i> 1999; Kankanhalli <i>et al.</i> 2003; Lai <i>et al.</i> 2009; Probst <i>et al.</i> 2000; Saito <i>et al.</i> 2007; Wang <i>et al.</i> 2008]
Симуляционные программы – программное обеспечение, которое воспроизводит реальные ситуации	[Curado & Bontis 2010]
Телефон – телекоммуникационная технология, которая позволяет людям общаться, находясь в разных местах	[Jasimuddin 2007; Kankanhalli <i>et al.</i> 2003]
Видеоконференция – телекоммуникационные технологии (аудио и видео), которые позволяют встречаться с людьми, находящимися в разных местах	[Bollinger & Smith 2001; Hansen <i>et al.</i> 1999; Kankanhalli <i>et al.</i> 2003]
Сообщения голосовой почты – человек получает предварительно записанное сообщение, а затем имеет возможность оставить сообщение взамен	[Hansen <i>et al.</i> 1999]
Wiki – средство хранения знаний, созданных совместно	[Bollinger & Smith 2001; Grace 2009; O’Sullivan 2007; Saito <i>et al.</i> 2007]
Желтые страницы/справочники экспертов – определение людей со знаниями по конкретной теме	[Bollinger & Smith 2001; Davenport <i>et al.</i> 1998; Hansen <i>et al.</i> 1999; Jarrar 2002; Kankanhalli <i>et al.</i> 2003]

УЗ как академическая дисциплина активно развивается во всем мире. Тем не менее, по мнению португальских исследователей К. Курадо и Н. Бонтиса (2011), общепризнанной структуры или модели УЗ до сих пор не существует [Curado, Bontis, 2011; Senseuse, Cahyaningsih, 2018]. Наиболее применяемыми из известных моделей УЗ с близкими эпистемологическими и онтологическими взглядами являются следующие: Nonaka’s SECI (Japan), March’s Ex-Ex (USA) и Boisot’s I-space (Europe).

Модель Nonaka's SECI (Japan). Известные японские специалисты в области управления знаниями И. Нонака и Х. Такечи (1995) в работе «Компания – создатель знания», посвященной исследованию инновационных механизмов японских компаний, объясняют их феноменальные успехи в конце XX в. осмыслением и использованием концепции создания, интеграции и использования организационного знания [Nonaka, Takeuchi, 1995].

Они предложили модель, включающую четыре основных типа трансформации знания (рис. 6).

Модель включает следующие основные элементы [Маринко, 2004]:

- две формы знания – неявное (неформализованное) и явное (формализованное);
- динамику взаимодействия (передача);
- три уровня социальной агрегации (индивид, группа, контекст);
- четыре процесса «создания знания» (социализация, экстернализация, комбинация и интернализация).

Согласно этой модели, создание организационного знания происходит в результате непрерывного динамического взаимодействия не-

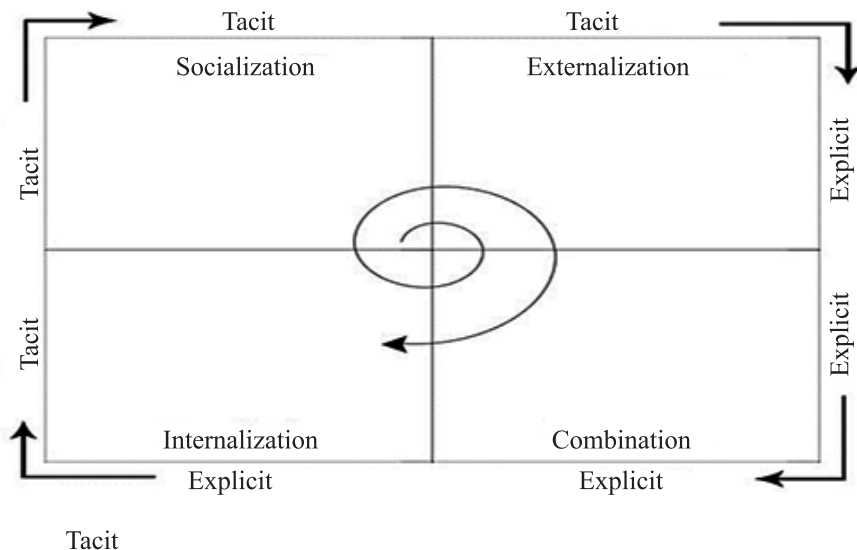


Рис. 6. Модель Nonaka's SECI

явного (Tacit) и явного (Explicit) знания, при котором используются разные типы трансформации знаний, а их выбор определяется разными пусковыми механизмами:

- социализация (Socialization) начинается с организации поля взаимодействия, способствующего непосредственному распространению опыта и интеллектуальных моделей, т.е. неявного знания сотрудников; это метод обмена неявными знаниями посредством наблюдения, подражания, практики и участия в официальных и неформальных сообществах и группах;

- экстернализация (Externalization) инициируется диалогом или так называемым коллективным размышлением с использованием метафор и аналогий; это техника выражения неявных знаний в явных понятиях. Как только неявное знание глубоко усвоено, этот процесс становится ключом к обмену знаниями и их созданию;

- комбинация (Combination) «включается» образованием связей между только что созданным и уже имеющимся знанием, но принадлежащим другим структурам организации (при этом формируется и используется системная модель, включающая различные точки зрения на ситуацию); это метод интеграции концепций в систему знаний. Некоторые примеры или случаи могут представлять собой обобщение в форме отчета, анализа тенденций, краткого исполнительного резюме или новой базы данных для организации контента;

- интернализация (Internalization) стимулируется обучением на практике с использованием формализованных инструкций и описаний; это техника воплощения явных знаний в неявные знания.

При этом разные типы трансформаций формируют знание специфического содержания:

- социализация создает дружественное знание (общие ментальные модели и навыки);

- экстернализация дает концептуальное знание (совместное формирование концепций и базовых моделей);

- комбинация способствует системному знанию (созданию системных моделей, прототипов и новых технологий);

- интернализация создает операционное знание об управлении проектами, процессами, использовании продуктов и технологий с помощью обучения.

Носителями знания являются люди с их индивидуальным знанием (явным и неявным). Организация может рассматриваться как среда коллективной деятельности, связанная с интенсивными коммуникациями и обменом информацией, для достижения определенных целей. Следовательно, знание организации – это знание ее сотрудников, распространенное на уровень всей компании, ставшее достоянием всех ключевых специалистов и обогащенное их индивидуальным знанием. В этом организационном знании можно выделить более общие фундаментальные знания о бизнесе компании и более специализированные, известные определенным сотрудникам. Принципиальное значение имеют трансформации явного и неявного знания отдельных индивидумов и распространение этих изменяющихся, трансформируемых знаний внутри (а иногда и за пределами) организации. Поскольку при каждом новом цикле происходит обогащение и изменение знания, то процесс можно представить в виде разворачивающейся спирали. При этом постепенно происходит включение в процесс создания организационного знания все большего числа людей в организации. Таким образом, знание распространяется по организации, выходя на групповой, межгрупповой, организационный и даже межорганизационный уровень

Кроме процессов, обозначенных в четырех квадрантах матрицы SECI, представленной на рис. 7, выделяют также факторы, определяющие скорость трансформации знания при переходах от одного квадранта к другому. К этим факторам относят стимулы к распространению знания, культуру, лидерские качества и организационные мероприятия.

В 1991 г. американский исследователь Д. Марш предложил модель под названием «разработка – эксплуатация» («Exploration – Exploitation» – «Ex-Ex») [March, 1991; Curado С., Bontis, 2011; Абрамчук и др., 2011; Пястолов, 2011]. Модель «Ex-Ex» J. March основана на идее взаимосвязи между этими процессами (см. рис. 7).

Разработка (развитие новых знаний) обеспечивает создание новых продуктов и услуг, т.е. разработка нового знания является результатом организационной гибкости, управления рисками, экспериментаторства, инноваций и импровизаций. Эксплуатация заключается в балансировке существующих знаний, доработке продуктов и процессов на

		Выход		
		Индивид	Группа	Организация
Вход	Индивид	Знания индивидуального уровня	Индивидуально-групповое опережающее обучение	Разработки с проектированием будущего
	Группа	Групповое-индивидуальное закрепляющее обучение	Знания группового уровня	Групповое организационное опережающее обучение
	Организация	Эксплуатация с обратной связью	Организационно-групповое закрепляющее обучение	Знания организационного уровня

Рис. 7. Модель March's Ex-Ex [Абрамчук и др., 2011]

основе существующего знания. Динамическая природа модели проявляется посредством взаимосвязи между организационной ассимиляцией нового знания, полученного на индивидуальном уровне (например, опережающее обучение – *feed-forward learning*), и освоением индивидом предсуществующего организационного знания (например, закрепляющее обучение – *feed-back learning*). Эта взаимосвязь – напряжение (*tension*) – появляется, потому что оба эти процесса должны быть скоординированы на трех уровнях: индивидуальном, групповом и организационном.

Поддержание баланса двух процессов требует лидерских качеств менеджеров организации, управляющих трансформациями. Изменение внешних условий приводит к смещению приоритетов: в конкурентном окружении успех обеспечивается эксплуатацией имеющихся компетенций, а в меняющихся условиях выживанию способствует разработка новых компетенций. Нередко фирма попадает в ловушку обучения, когда одна из стратегий начинает преобладать. Чаще этой стратегией оказывается эксплуатация, так как отдача в этом случае ближе во времени и в пространстве.

В соответствии с концепцией М. Буазо (модель *Voisot's I-space*), социальное обучение протекает в информационном пространстве

(I-space) через шесть этапов: сканирование, кодификацию, абстрагирование, диффузию (распространение), абсорбцию, воздействие [Voisot, 1995; Voisot et al., 2004]. Схематически эта модель представлена в форме куба, сторонами которого являются кодификация, абстрагирование, диффузия (рис. 8).

Под сканированием подразумевается процесс сбора исходных данных, но с учетом индивидуальной интерпретации под влиянием накопленного опыта и знаний. Сканирование может быть очень быстрым, если данные хорошо кодифицированы, и медленным, если данные размещены случайным образом в специфическом контексте.

Кодификация представляет собой индивидуальную реакцию на получение сканированных данных. Изначальная информация может быть раздробленной, но классификация, интерпретация и устранение лишнего позволяют индивиду кодифицировать знание. Путем абстрагирования индивид формирует общие правила применения кодифицированных подходов к более широкому спектру ситуаций. Так же как и кодификация, абстрагирование опирается на опыт индивида в подобных ситуациях, но если знание было однажды кодифицировано и абстрагировано, его диффузия может быть очень быстрой.

Распространение хорошо кодифицированного и абстрагированного знания среди населения не представляет проблем, но некодифи-

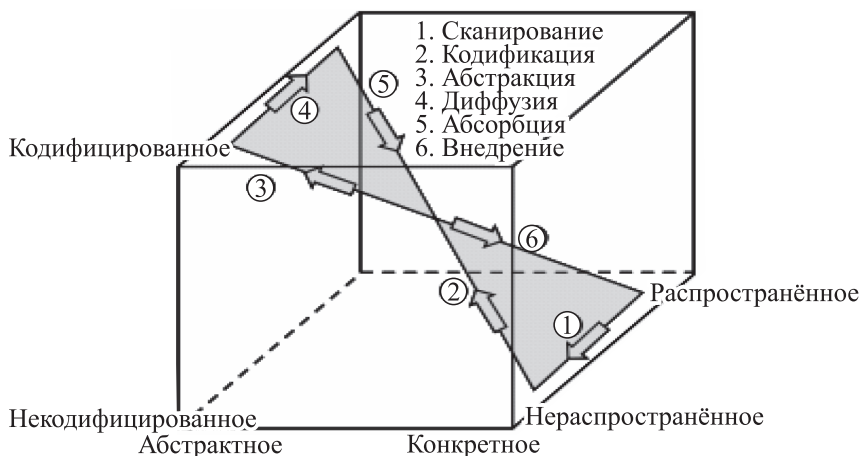


Рис. 8. Модель Voisot's I-space [Абрамчук и др., 2011]

цированное знание в специфическом контексте для ускорения своего распространения требует, чтобы источник передал содержание этого контекста получателю информации. Абсорбция имеет место тогда, когда, получив переданное знание, индивид модифицирует свое поведение в формате «обучения в процессе деятельности». В ходе абсорбции знания индивиды накапливают практический опыт. Применение новых кодифицированных подходов обеспечивает со временем обобщение знания.

Воздействие происходит тогда, когда новое абсорбированное знание применяется в конкретных ситуациях, когда индивиды модифицируют поведенческие паттерны, совершают новые действия и воздействуют на окружение при помощи полученного знания [Curado С., Bontis, 2011; Пястолов, 2011].

Модель Boisot's I-space рассматривает компании как живые организмы, когда процесс роста и развития активов знаний внутри компаний постоянно меняется. Это означает, что компании должны применять динамическую стратегию управления знаниями, которая учитывает динамический характер организационного цикла обучения.

Т.А. Нестик (2013) провел анализ используемых моделей УЗ в российских организациях. Он отметил, что выделяются две стратегии УЗ в современных компаниях: ориентированная на кодификацию знаний и ориентированная на персонализацию знаний [Нестик, 2013].

Стратегия кодификации опирается на широкое использование IT-инструментов поиска данных и поддержку электронных БЗ, позволяющих накапливать интеллектуальный капитал организации. Такой подход практически не оставляет места для инноваций и креативности сотрудников, однако он дает возможность опираться на проверенные опытом, надежные способы решения возникающих проблем. Кодификация предполагает существенные инвестиции в IT-инфраструктуру, организацию доступа к информации в зависимости от должностного положения и функций сотрудников, а также стимулирование персонала к пополнению и использованию баз знаний. Часто данный подход используется при реализации масштабных IT-проектов в крупных компаниях, например при внедрении систем автоматизированного управления (ERP-системы, CRM-системы), а также при переходе к электронному документообороту.

Стратегия персонализации знаний, напротив, ориентирована не столько на технологии, сколько на развитие креативности сотрудников и облегчение обмена знаниями. Она хорошо зарекомендовала себя в организациях с матричной и проектной структурой, где для решения возникающих проблем используются очные встречи, электронная переписка и видеоконференции. Основное внимание при реализации данной стратегии уделяется развитию социальной сети проектных команд и отдельных сотрудников, которая позволяет совместными усилиями находить решение проблем. Сотрудники вознаграждаются не столько за кодификацию своих знаний в корпоративных базах, сколько за обмен знаниями друг с другом, поддержание конструктивного диалога поверх функциональных границ.

Т.А. Нестик (2013) указывает, что после анкетирования экспертов 20 московских организаций из сферы энергетики, производства и оптовых продаж с численностью персонала от 200 до 20 000 сотрудников, можно констатировать, что УЗ еще не стало общепринятым направлением деятельности в российских компаниях. Однако в целом за последние 3 года отмечается тенденция к отказу от стратегии кодификации знаний в пользу персонализации или к попыткам совмещения элементов обеих стратегий [Нестик, 2013].

Т.А. Нестик (2013) также пишет, что в настоящее время УЗ в российских компаниях основывается на пяти основных моделях, каждая из которых имеет свои социально-психологические особенности.

Первая модель – УЗ в рамках проекта. Данный подход предполагает интенсивный анализ опыта, накапливаемого в ходе проекта, и его тиражирование в других аналогичных проектах. Преимуществом данной модели является относительная простота применяемых технологий, которые не требуют сложной ИТ-инфраструктуры и кодификации: в основном это регулярные совещания проектных команд по анализу опыта и различные общекорпоративные мероприятия, позволяющие подразделениям и проектным группам обмениваться передовыми практиками.

Вторая модель УЗ основана на использовании виртуальных групп экспертов. Пример такого подхода – «сетевые группы» в компании Лукойл.

Третьей моделью УЗ является создание центров компетенций – структурных подразделений, отвечающих за сбор и распространение передового опыта в определенной сфере деятельности.

Четвертая модель УЗ, широко представленная в российских компаниях, основана на развитии внутренних коммуникаций и использовании методов организационного развития. Данная модель наиболее близка к стратегии персонализации знаний и ориентирована на стимулирование инновационной деятельности сотрудников.

Пятая модель, постепенно получающая распространение в России, опирается на открытые инновации, т.е. ориентирована на поиск знаний и экспертов за пределами компании. Условия жизни в сетевом обществе поставили компании перед необходимостью перейти от создания внутри- и межорганизационных сетей обмена знаниями к так называемым открытым инновациям. Эффект такого подхода состоит в том, что он за короткое время повышает осведомленность сотрудников о стратегических целях компании, приверженность персонала и его вовлеченность в реализацию стратегических инициатив [Нестик, 2013].

В публикации Т.А. Гавриловой с соавт. (2017) проведен анализ опроса 104 слушателей и выпускников программы EMBA (Executive Master of Business Administration) Высшей школы менеджмента в 2010–2015 гг. Слушателям было предложено произвести оценку развития системы УЗ в собственных компаниях. Анализ результатов анкетирования позволил авторам сделать следующие выводы [Гаврилова и др., 2017]:

- вне зависимости от общего уровня зрелости УЗ в компании, в среднем у организаций, представленных респондентами, похожие сильные и слабые стороны;

- стратегический этап строительства и поддержки знания, выражаемый в поддержке УЗ со стороны руководства, оснащении информационными системами, предоставляющими доступ к актуальным знаниям, а также в поддержке продуктивных начинаний, является наиболее развитым в российских организациях;

- проблемным является стратегический этап совершенствования информации в связи с отсутствием сотрудников-координаторов по обмену знаниями, а также с невысоким уровнем вовлеченности сотрудников в процессы обмена и обновления знаний [Гаврилова и др., 2017].

М.В. Власов и А.Ю. Шишкина провели сравнительный анализ российского опыта построения моделей УЗ (табл. 8) [Власов, Шишкина, 2015].

Характеристика российских моделей генерации знаний

Название модели, автор, год	Краткое описание модели	Особенности (преимущества/недостатки) модели
1	2	3
Модель организации внутрифирменного управления знаниями, Мильнер, 2003	Включает производство знаний, распределение знаний, использование знаний	На пересечении указанных процессов возникает новая функция – функция управления передачей знаний
Корпоративная модель генерации знаний, Аксенова, 2006	Основана на принципах самообучающейся организации	Акцент сделан не на накопление новых знаний, а на их эффективное использование. Применима в научных и образовательных учреждениях
Функциональная модель, Бунина, 2008	Опирается на развитие и мобилизацию знаний организационными методами. Характеризуется расширением взаимодействия явного и неявного знания по мере перехода с одного на другой онтологический уровень	Не рассматривает вопросы стимулирования производства новых знаний
Интегрированная модель, Смирнова, 2008	Взаимосвязь трех составляющих: технологической, организационной и человеческой, с элементами моделей управления знаниями в организациях: людьми, процессами, технологиями	Модель сложно адаптируется под специфику деятельности организации
Циклическая модель, Суслов, 2012	В основе модели – трансформация знаний путем внутриорганизационного обучения	Недостаточно проработаны вопросы повышения эффективности использования полученных в организации знаний
Трехуровневая модель, Гумерова, Шаймиева, 2013	Содержит показатели для трех уровней: макроуровень – рынок знаний промышленности РФ, мезоуровень – промышленно-развитых округов, микроуровень – промышленных предприятий	Возникает вопрос о выделении еще одного уровня – деятельности конкретного индивида

1	2	3
Интегрированная модель представления и обработки знаний, Черняховская, Федорова, Владимирова, 2014	Состоит из трех видов моделей: – объектная модель управления знаниями; – онтологическая модель поддержки принятия решений; – модель формирования правил и прецедентов в онтологии поддержки принятия решений	Модель предусматривает создание институциональной базы системы управления знаниями. Однако не рассмотрены такие процессы генерации знаний как передача и накопление знаний

Управление знаниями – актуальная проблема. Вот что по этому поводу писал Б.З. Мильнер (2003) еще в начале XXI в.: «Можно считать вполне установленным, что все более возрастающая доля стоимости продуктов и услуг принадлежит «незримой собственности» – знаниям, нематериальным активам. Поворот организации и управления к этой огромной, многообразной и малоизученной сфере, в сущности говоря, является событием знаковым и рубежным. Это тем более важно, что разработка теории управления знаниями находится еще на начальной стадии, а «все более возрастающая доля» управляется и регулируется слабо и бессистемно. Научиться управлять знаниями – это войти в особый мир понятий, категорий, вариантов, структур, методов, процедур, технологий, оценок. Это – новая область приложения организационных механизмов, управленческих приемов и экономических стимулов, порождающая реальные конкурентные преимущества компаний» [Мильнер, 2003].

В заключении этой главы отметим, что всякое познание представляет собой моделирование. Модель – это искусственно созданная система, в которой отражено сходство структуры и функции с системой-оригиналом. Модель всегда упрощает и искажает оригинал. Однако упрощение принципиально необходимо в связи с большой сложностью аграрной системы и накопленного знания о ее функционировании.

Аграрные науки используют главным образом словесные описания или словесные модели с добавлением отдельных схем, цифр или уравнений. Сельскохозяйственные модели и сельскохозяйственное знание не являются точными, они не позволяют количественно рассчитать изменения системы-оригинала, например точно прогнозировать урожай-

ность какой-либо культуры. Неточный (недоопределенный) характер сельскохозяйственных знаний предполагает применение принципиально новых (нетрадиционных) методов познания (пытающихся ускорить работу человека-аналитика путем переложения части его функций на возможности компьютера) для выявления взаимосвязей между различными объектами и процессами, т.е. для проведения исследований, ранее традиционно осуществляемых только человеком. Эти исследования в мире в последние годы получили дальнейшее развитие.

Предпосылками для применения методов ИИ являются в основном описательный характер накопленных сельскохозяйственных знаний, а также невозможность конструктивистского подхода – точного (количественно) предсказания течения процессов и принятия конкретного решения на основании известных общих закономерностей.

Даже исследовательские базы данных отдельных ученых, ведущих наблюдения по очень узкой тематике, могут содержать сведения о нескольких десятках различных показателей, число записей может составлять несколько сотен. Эмпирический анализ даже такого малого объема информации весьма сложен и трудоемок. Об объеме получаемых данных в настоящее время может свидетельствовать тот факт (по данным аналитического отдела американской компании GTE), что только одни научные учреждения за один день записывают, фиксируют информации примерно на 1 терабайт. При этом следует учесть, что наука не самый большой источник данных.

Предпосылками использования новых подходов к анализу данных, информации и знаний является также развитие аппаратного и программного обеспечения. Ключевой возможностью применения информационных технологий, например Data mining (добыча данных, интеллектуальный анализ данных), за последние несколько лет стало снижение цены на устройства хранения информации. Уменьшение цены на процессоры с одновременным увеличением их быстродействия и распространение клиент-серверной архитектуры также способствуют развитию технологий, связанных с обработкой больших массивов информации.

Таким образом, использование методов ИИ в разработке и применении предиктивных технологий сельскохозяйственной деятельности имеют широкие перспективы и могут способствовать развитию отечественного сельского хозяйства.

ГЛАВА 2

ЦИФРОВИЗАЦИЯ ПРОЦЕССА ВОЗДЕЛЫВАНИЯ СЕЛЬСКОХОЗЯЙСТВЕННОЙ КУЛЬТУРЫ

Сельское хозяйство характеризуется множеством различных факторов, описывающих экономические, пространственные, экологические, социально-демографические, технические и другие явления, происходящие в агросистеме. Так же, как в любой неоднородной среде, наиболее важные свойства агросистемы обнаруживаются во взаимодействии явлений. Причем, наиболее существенными для понимания и управления системой часто оказываются те, которые не наблюдаются ни в одном из перечисленных явлений, а проявляются в результате их взаимодействия. Все это в целом делает агросистемы весьма специфическим и трудным объектом для исследования. Механизмы происходящих процессов, внутренние их стимулы, характер взаимодействий между элементами столь сложны и разнообразны по своей природе, нестабильны и случайны, что возникают значительные препятствия даже на начальном этапе исследования – определении состава данных, которые характеризуют состояние агросистемы, декомпозиции ее на блоки и выявлении причинно-следственных связей между ними [Липкович, 2008; Москвин, 2011].

2.1. Понятие о цифровой зрелости интеллектуальной информационной системы

В отечественной практике сложилось так, что моделированием агросистем занимались в основном экономисты при разработке программ развития сельскохозяйственного предприятия и его управления. Основным инструментом реализации такого типа задач были экономико-математические модели [Леньков, 1997; Смагин, 2007; Нечаев и др., 2012]. Экономико-математическая модель сельскохозяйственного предприятия представляет собой оптимизационную задачу, концентрированное выражение наиболее существенных взаимосвязей и закономерностей

поведения управляемой системы в математической форме [Гатаулин и др., 1990]. В настоящее время существует множество апробированных экономико-математических моделей. Это модели по оптимизации отраслевой структуры производства и посевных площадей [Родионов, Макарова, 2016]; формирования севооборотов [Шеров, 2005], рационов кормления сельскохозяйственных животных и расчета потребности в кормах [Скоркин и др., 2013; Харитоновна, Волкова, 2015]; состава и структуры машинно-тракторного парка [Кухарев и др., 2012]; распределения и использования минеральных удобрений [Иванов, Никонова, 2016] и др. Особенностью такого моделирования является то, что оно позволяет из области допустимых решений выбрать одно лучшее по заранее заданному критерию. Дополнительным стимулом к широкому использованию моделей данного типа были относительная простота математического аппарата и наличие стандартных программных средств реализации данных моделей на персональных ЭВМ.

Однако в настоящее время все больше исследователей склоняются к тому, что оптимизационные модели имеют ограниченное значение. Эти модели, как правило, ориентированы на поиск параметров системы, характеризующих ее состояние в конкретный момент времени, тогда как процесс достижения данных параметров остается вне пределов их компетенции (за исключением динамических оптимизационных моделей, использующихся для решения некоторых задач развития, но отличающихся довольно высоким уровнем научной абстракции). Наиболее эффективным методом исследования процессов функционирования агросистем признается имитационное моделирование [Stermann, 2001; Улезько и др., 2012; Носонов, 2017; Алетдинова и др., 2019; Осенний, Франциско, 2019]. Кроме того, в отечественной сельскохозяйственной науке и практике происходит становление качественно новых подходов и методов управления использованием земельных ресурсов, базирующихся на информационных технологиях. Это связано с появлением концепции интеллектуального (умного) сельского хозяйства (smart farming) и необходимостью освоения для управления производством цифровых технологий и интеллектуального анализа данных [Барсегян и др., 2004; Анищенко, 2019; Ганичева и др., 2019; Walter et al., 2017; Wolfert et al., 2017].

Общие принципы цифровых технологий и интеллектуального анализа данных могут быть сведены к следующим двум правилам и четырем функциям:

- данные накапливаются постоянно;
- появление любых новых данных меняет результаты и методы их получения;
- получение результата;
- выбор метода получения результата;
- выбор обрабатываемых методом данных;
- выбор данных для накопления.

Эти шесть постулатов не имеют ограничения общности и относятся к любой задаче, связанной с обработкой больших объёмов данных, информации и знаний [Акоф, Сасиени, 1971]. Информационная система, в которой эти правила и функции реализованы и автоматизированы, имеет право называться интеллектуальной.

Этапы становления таких систем рассматриваются в рамках понятия «индекс цифровой зрелости» (Maturity Index) [Schuh et al , 2017]. Шесть стадий цифровой зрелости включают:

- информатизацию: сводится к пониманию, какие данные, информацию и знания необходимо накапливать для реализации информационной системы;
- связанность: сводится к пониманию связей между накапливаемыми данными, информацией и знаниями;
- наглядность: сводится к интеграции в информационную систему множества источников, из которых накапливаются данные, информация и знания;
- проницаемость: сводится к реализации всевозможных способов обработки данных, информации и знаний;
- предсказуемость: сводится к использованию имитационного моделирования процессов и прогнозирования возможных нештатных ситуаций;
- самокоррекцию: сводится к автоматизации реакции информационной системы на нештатные ситуации.

Отметим часто встречающуюся методологическую ошибку, допускаемую коллективами, начинающими знакомиться с парадигмой цифровых технологий. Довольно естественно начинать знакомство с

данной проблемой путём изучения научных статей на тему цифровой экономики, из которых можно получить набор ключевых понятий и далее сосредоточиться на некоторых из них. Этот процесс принято визуализировать с помощью технологий TextMining, применяемым к базам данных научных публикаций [van Eck, Waltman, 2011]. Например, с помощью компьютерной программы VOSviewer (<http://www.vosviewer.com/>) можно получить визуальные кластеризованные семантические схемы профилей научных учреждений, университетов или просто множества учёных, работающих в разных дисциплинах, но в рамках одной парадигмы [van Eck et al, 2010; Polley, 2016]. Фрагмент одной из таких схем, составленной по работам российских учёных, аффилированных с Высшей школой экономики, в парадигме цифровизации и Индустрии 4.0, приведён на рис. 9.

Жёлтыми прямоугольниками выделены понятия, которые в общественном сознании наиболее тесно ассоциируются с цифровыми технологиями, с использованием ИИ для выявления скрытых закономерностей и принятия решений. Однако практически все эти понятия не относятся к первым трём этапам цифровой зрелости. Для того чтобы начать пользоваться всеми этими методами, необходимо выстроить («аппаратно» и «программно») платформу для их использования на следующей, четвёртой стадии становления интеллектуальной информационной системы (ИИС) – с каким бы процессом и какой бы сферой деятельности она ни была связана. Не создавая платформу, можно изучать отдельные методы интеллектуальной обработки данных, информации и знаний с помощью специализированного программного обеспечения, снабжённого заранее подготовленными разработчиками тестовыми примерами. Используя его, можно получить представление о методе, но при этом ни на шаг не сдвинуться по шкале цифровой зрелости.

Упомянутые выше шесть стадий формируют маршрутную карту цифрового развития ПО ИИС в рамках любой сферы деятельности. Понятие «индекс цифровой зрелости» восходит к парадигме управления бизнес-процессами, в свою очередь выросшей из исследования операций, совмещённого с вычислительными возможностями современных компьютеров [von Brocke, Rosemann, 2015]. Парадигма управления бизнес-процессами включает шесть базовых элементов:

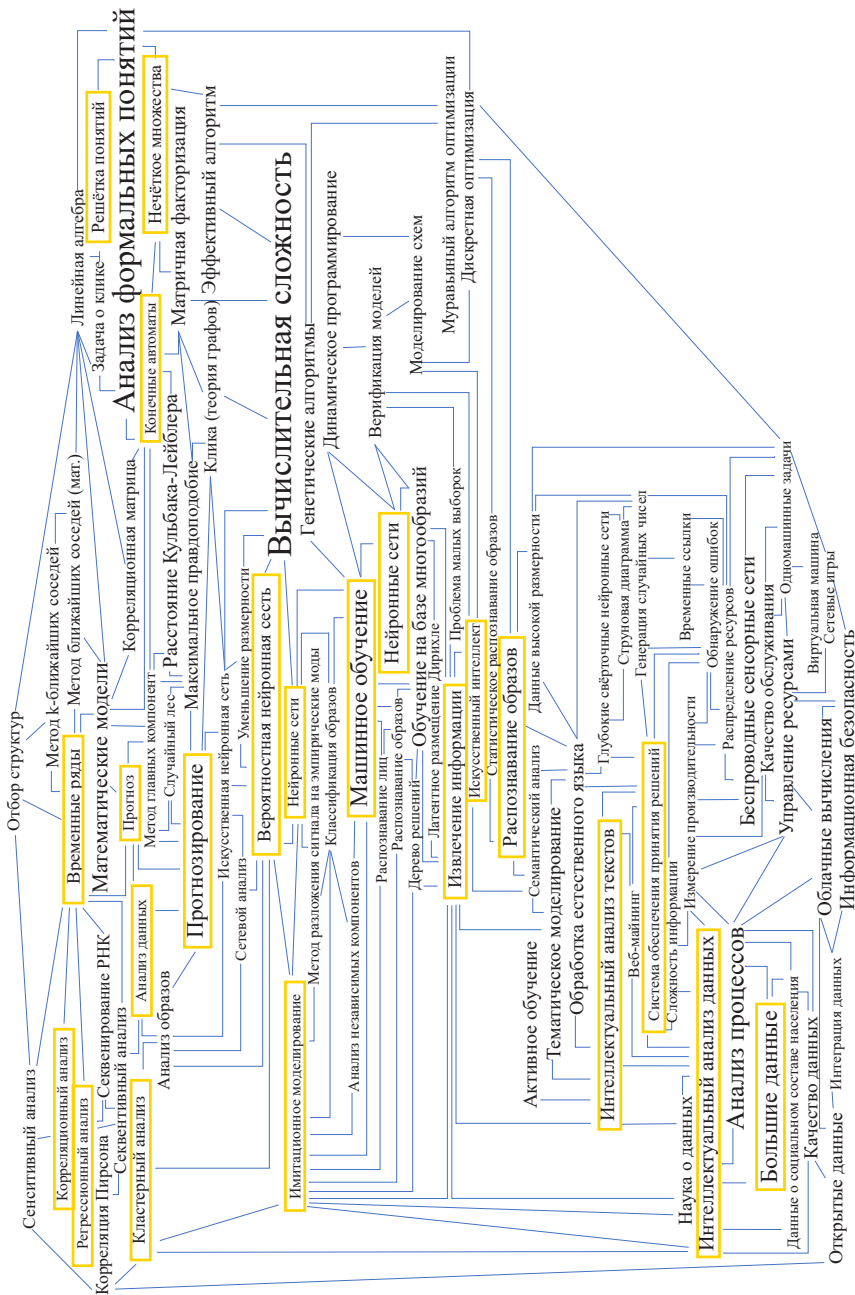


Рис. 9. Фрагмент семантической схемы работ учёных, аффилированных с Высшей школой экономики, за 2014–2018 гг. (источник: <https://www.hse.ru/info/competence>)

- стратегическое выравнивание (strategic alignment);
- управление (governance);
- методы (methods);
- информационные технологии (information technology);
- кадры (people);
- культуру (culture).

Эти шесть элементов задают среду (свойства команды, коллектива, лаборатории) реализации маршрутной карты становления цифровой интеллектуальной информационной системы.

Таким образом, для реализации цифровых технологий в управлении процессами сельского хозяйства нужно провести цифровизацию одного из стандартных для сельского хозяйства процессов, постепенно повышая его цифровую зрелость. В Лаборатории для практической цифровизации была выбрана ПО «процесс возделывания сельскохозяйственной культуры». С точки зрения исследования операций, этот процесс проходит в течение календарного года, состоит из этапов, выполняемых по стандартизированным маршрутным картам. Каждый этап предполагает набор предписанных действий – агроприёмов, применение которых может варьироваться в диапазонах изменения характеризующих их параметров. На выбор и параметры агроприёмов влияют две группы факторов: история предшествующего использования земельного участка и текущие случайные факторы технологического и погодного характера. Проблемы с первой группой факторов сводятся к недостаточности данных, проблемы второй – к принятию решения в условиях ограниченного времени и в обстановке, в которой невозможно дальнейшее следование заданной маршрутной карте.

Начало формирования цифровой зрелости процесса возделывания сельскохозяйственной культуры – это его информатизация, т.е. понимание того, какие данные, информация и знания необходимы в принципе для дальнейшего цифрового управления данным процессом. Чтобы в этом разобраться, необходимо выстроить картину мира [Чебанов, 2008] об этом процессе с учётом необходимости его дальнейшей автоматизации в некоторой ИИС. Картина мира – термин, используемый в герменевтике – науке, прикладное значение которой в том, чтобы методологически извлекать значимую информацию из неструктурированных текстовых данных [Кордонский, Бардин, 2018].

Герменевтика предвдвряет комплекс информационано-вычислительных методов интеллектуального анализа текста (Text Mining), в которых значимая по заданным критериям информация извлекается из текстов на электронных носителях автоматически, в задаче извлечения метазнаний. Метазнания – неявные знания экспертов ПО, используемые ими при решении практических задач на протяжении многих лет. Эти знания считаются экспертами само собой разумеющимся приложением к применению каких-либо формализованных методов решения стандартных задач ПО, допускающих большое количество траекторий движения к результату [Evans, Foster, 2011]. Для тех, кто этими неявными знаниями не обладает, они представляют огромную ценность, а выполнение задачи по их извлечению и структурированию является формализацией картины мира рассматриваемого процесса.

2.2. Использование элементов прикладного системного анализа

Системный анализ, с одной стороны, признается в настоящее время наиболее конструктивным из направлений системных исследований, с другой – не существует в виде строгой методологической концепции и тесно связан с различными направлениями современной науки. В связи с лавинообразным характером системных исследований предложено достаточно большое количество общих и специальных определений понятия «система». Не будем анализировать все определения, приведем лишь одно из них, данное Дж. ван Гигом (1981) и претендующее на универсальность: система есть совокупность или множество связанных между собой элементов [Гиг, 1981]. Практическое применение определений такого класса невелико. Однако свойство универсальности таких определений дает возможность развернуть знания о динамически сложной среде в любом направлении [Сорокин, 2017].

Добавим, что системный анализ в настоящее время представляет собой слабосвязанную совокупность приемов и методов неформального и формального характера. В системном мышлении чаще всего господствует интуиция (возможно, это и хорошо: излишняя формализация не всегда приводит к удовлетворительным результатам). Ситуация усугубляется еще и тем, что, несмотря на более чем полувекую

историю развития системных идей, нет однозначности в понимании самого системного анализа. Приведем для наглядности этого тезиса обзор, проведенный Ю.П. Сурминым (2003), который выделяет следующие варианты понимания сущности системного анализа:

- отождествление технологии системного анализа с технологией научного исследования. При этом для самого системного анализа в этой технологии практически не находится места;

- сведение системного анализа к системному конструированию, в этом случае системно-аналитическая деятельность отождествляется с системотехнической;

- сведение системного анализа к одной из его составляющих, например к структурно-функциональному анализу;

- отождествление системного анализа с системным подходом в аналитической деятельности;

- понимание под системным анализом совокупность математических методов исследования систем;

- сведение системного анализа к совокупности методологических средств, которые используются для подготовки, обоснования и осуществления решений по сложным проблемам.

Системный анализ в настоящее время вполне справедливо рассматривают как эффективный метод изучения объектов и процессов управления, однако методики системной аналитики в решении конкретных управленческих задач также практически отсутствуют. Как считает Ю.П. Сурмин (2003), «Системный анализ в управлении представляет ныне не развитую практику, а нарастающие ментальные декларации, не имеющие какого-либо серьезного технологического обеспечения» [Сурмин, 2003].

Однако при разработке технических систем, куда входят и ИИС, применение системного анализа достаточно стандартизировано и входит в учебную программу многих отечественных высших учебных заведений [ФГОС, 2015].

Научный подход к формализации картин мира, описывающих процессы в различных областях человеческой деятельности, зародился в течение нескольких десятков лет после Второй мировой войны, прежде чем выделиться в ряд конкретных научных направлений. Исследователи истории данного вопроса связывают это с усложнением

организации военных действий [Акоф, Сасиени, 1971], с требованием времени более эффективно управлять большими организациями [Гозман и др., 2008] и с отказом от закона исключённого третьего в математике [Колмогоров, 1925] в силу его неэффективности в реальных ситуациях сложного выбора [Клайн, 1984]. Перечисленные причины привели к появлению таких методологий решения практических задач как, например, исследование операций [Акоф, Сасиени, 1971], прикладной системный анализ [Тарасенко, 2017], метод тематического исследования (case study research method) [Gerring, 2007], теория серой информации [Liu et al, 2017], суперинтуиционистская логика [Клини, 1957; Гейтинг, 1965; Кошечева, 2016], теория случайных процессов [Булинский, Ширяев, 2005]. Применение каждой методологии требует обработки больших объёмов данных, информации и знаний, которые собираются по сформированным «под метод» картинам мира. Некоторые из перечисленных методологий можно называть гуманитарными, какие-то инженерными, другие можно отнести к прикладной математике. В парадигме цифровой экономики эти различия не имеют значения: любые данные, информация и знания, связанные с исследуемым объектом или процессом, должны быть формализованы и внесены в матрично-структурированные, облачно-цифровые хранилища без потери возможности использовать перечисленные методы обработки данных на дальнейших стадиях цифровой зрелости [Мошелла, 2020].

Продемонстрируем применение элементов технологии прикладного системного анализа к формализации картины мира для процесса возделывания сельскохозяйственной культуры. Согласно Ф.П. Тарасенко (2017), формализация картины мира происходит на пятом шаге алгоритма применения прикладного системного анализа и называется этот шаг «Определение конфигуратора». Пятый шаг предваряют шаги [Тарасенко, 2017]:

- 1-й шаг. Фиксация проблемы.
- 2-й шаг. Диагностика проблемы.
- 3-й шаг. Составление списка стейкхолдеров¹.
- 4-й шаг. Выявление проблемного месива.

¹(Англ. *stakeholder*), также заинтересованная сторона, причастная сторона, участник работ, роль в проекте – лицо или организация, имеющая права, долю, требования или интересы относительно системы или её свойств, удовлетворяющих их потребностям и ожиданиям (<https://ru.wikipedia.org/wiki/Стейкхолдер>)

Проблема в этой главе уже обозначена: для процесса возделывания сельскохозяйственной культуры необходимо построить рабочую цифровую модель, т.е. соблюсти шесть постулатов ИИС, пройти шесть стадий цифровой зрелости и реализовать шесть базовых элементов управления бизнеспроцессами.

Диагностика проблемы заключается в необходимости формализации картины мира для исследуемого процесса, чтобы в дальнейшем под эту картину мира эффективно собирать и оцифровывать данные, информацию и знания.

Стейкхолдеры – это, по определению, участники проблемной ситуации, имеющие разные наборы целей и интересов, в совокупности представляющие собой проблемное месиво. По технологии прикладного системного анализа со стейкхолдерами проводятся беседы и на основе бесед составляется конфигуратор – достаточный набор понятий, объединяющий различные профессиональные языки стейкхолдеров так, чтобы описание проблемной ситуации и её преобразование было понятно всем.

Сотрудниками Лаборатории в 2019, 2020 гг. проводились беседы – как в очной форме, так и путём знакомства с отчётами, научными статьями и книгами – с различными группами стейкхолдеров, связанных с процессом возделывания сельскохозяйственных культур: учёными, производителями сельскохозяйственной продукции, собственниками хозяйств, представителями государственных регулирующих органов, агрономами. На основании этих бесед и в рамках парадигмы цифровых технологий составлено описание проблемы сбора данных в сельском хозяйстве, демонстрирующее общее естественное желание всех стейкхолдеров построить некоторое цифровое хранилище данных таким образом, чтобы потом из него максимально эффективно можно было извлекать информацию и знания.

Отметим, что конфигуратор в качестве минимального набора профессиональных языков, позволяющий дать адекватное описание проблемной ситуации и ее преобразований, можно сформировать описанным ниже способом. Практики часто руководствуются интуицией, здравым смыслом, опытом, советами экспертов. Как всякое субъективное решение, оно может оказаться верным, но может содержать ошибки. Между тем объективная, не зависящая от чьих-то мнений информа-

ция для построения конфигуратора содержится в протоколах бесед со стейкхолдерами о проблемной ситуации. Дело в том, что каждый из них говорит только о том, что считает важным, т.е. на языках своего конфигуратора для данной ситуации. В связи с этим появляется возможность «вычислить» конфигуратор, анализируя его текст из проблемного мессива, но не на предмет того, о чем он говорит, а того, на каких языках он говорит. Кто-то обращает внимание на финансовые аспекты, в его конфигураторе – экономический язык. Другой упоминает проблемы адаптации агротехнологий к погоде и планирование технологических операций – агрометеорологический язык и язык планирования входят в его конфигуратор. Так поступают с каждым стейкхолдером и формируют набор конфигураторов всех стейкхолдеров. Конфигуратор ситуации в целом есть их объединение. В него может входить один, а может и много языков. Может быть язык, на котором говорят все, на другом языке – большинство, на третьем – меньшинство. Однако все языки необходимо использовать, нельзя выбросить ни один из них, иначе соответствующий аспект не будет учтен, что не позволит претендовать на построение адекватной модели поддержки принятия решений.

Пример работы системного аналитика с научным текстом по проблеме накопления данных о сельскохозяйственных объектах показан в табл. 9. В левой колонке описание проблемы сбора данных в сельском хозяйстве (традиционный текст написания, принятый в книгах и статьях по сельскому хозяйству, взятый из статьи [Каличкин, Куценогий, 2019]), в правом – его видение с точки зрения системного аналитика, выполняющий 5-й шаг алгоритма прикладного системного анализа. Для удобства комментариев текст в левой колонке пронумерован по абзацам.

Комментарии системного аналитика теперь следует кластеризировать, т.е. разбить на группы. При анализе комментариев достаточно быстро выделяются 6 групп понятий, характеризующих картину мира процесса возделывания сельскохозяйственной культуры. Приведём краткий обзор этих групп.

1. В первую очередь речь идёт о структуре данных: комментарии (1)–(3), (10)–(21), (23)–(24), (46), (55), (68), (73) указывают на наличие устоявшихся структур данных в сельском хозяйстве и существующих программных средств сбора этих данных. Всё это относится к первому этапу цифровой зрелости.

Анализ текста статьи по проблеме накопления данных в сельском хозяйстве

Оригинальный текст статьи	Комментарии аналитика
<p>1. Структура данных о сельскохозяйственных системах, а также их сбор и анализ, по нашему мнению, должен состоять из двух категорий: статистических (архивных) и текущих (мобильных). Статистические (архивные) данные необходимы для агроэкологической оценки земель, формирования систем земледелия, размещения по территории культур и севооборотов. Текущие (мобильные), часто в онлайн режиме, необходимы для сопровождения технологий возделывания культур с целью управления продуктивностью посевов.</p>	<p>2</p> <p>(01) Структура данных о сельскохозяйственных системах (02), (03); (02) статистические (архивные) данные; (03) текущие (мобильные) данные; (04) есть сельскохозяйственные задачи, в том числе (05)–(09); (05) агроэкологическая оценка земель; (06) формирование систем земледелия; (07) размещение по территории культур и севооборотов; (08) сопровождение технологий возделывания культур; (09) управление продуктивностью посевов.</p>
<p>2. <i>Статистические (архивные) данные.</i> В основе статистических данных должна лежать информация о земельных ресурсах определенной территории, почвенном покрове, рельефе, климате, урожайности сельскохозяйственных культур и др., позволяющая построить временные ряды. Лучшим инструментом для сбора этих данных выступают геоинформационные системы (ГИС). ГИС позволяют позиционировать участки земной поверхности любого экономического уровня, отразить его точное местоположение и описать это в пространственной (географической) базе данных.</p>	<p>(10) Категория (02) включает (10)–(14); (11) земельные ресурсы определённой территории; (12) почвенный покров; (13) рельеф; (14) климат; (15) урожайность сельскохозяйственных культур.</p>
<p>3. В ГИС все данные организируются в логические группы (тематические), называемые слоями, которые в свою очередь группируются в карты. Слоевые карты (тема) – совокупность однотипных пространственных объектов, определённых в одной модели данных на общей территории и в общей системе координат.</p>	<p>(16) ГИС – способ построить структуру (02).</p>

Продолжение табл. 9

1	2
<p>4. Цифровая модель, в свою очередь, может существовать, храниться и обрабатываться в рамках определенных моделей (представлений). К ним относят растровую, векторную, квадратометрическую и иные двух- и трехмерные модели данных, которым соответствуют некоторые форматы данных.</p>	<p>(17) Дальнейшие детали построения структуры (02) через (16). Пока что это всё относится к (02). То есть данные собираются и укладываются в структуру с помощью принятых технических средств.</p>
<p>5. Основные цифровые модели, используемые в ГИС: – цифровая модель местности (ЦММ; digital terrain model, DTM); – цифровая карта (ЦК; digital map); – цифровая модель рельефа (ЦМР; digital terrain model, DTM; digital elevation model, DEM; Digital Terrain Elevation Data, DTEM).</p>	<p>(18) Технические детали построения (02) через (16). Для системного анализа такая информация на данном этапе не представляет интереса.</p>
<p>6. Для ввода, хранения, манипулирования и вывода атрибутивной (непространственной) информации в ГИС используются стандартные системы управления базами данных (СУБД). При этом в число атрибутов в этом случае не включаются геометрические свойства, описывающие топологические характеристики географических объектов. Последние упорядочиваются и организуются с использованием особых свойств ГИС. Необходимая связь между геометрическим описанием объектов и их содержательными атрибутами устанавливается через идентификаторы – уникальные номера (коды) географических объектов (ИД). Атрибутивную базу данных (БД) можно создать автономно в любой доступной СУБД (например, в Microsoft Access или Microsoft SQL Server), а затем с помощью заполнения идентификационных полей атрибутивной таблицы кодами объектов и программных средств обмена данными связать ее с уже созданной географической БД.</p>	<p>(19) Технические детали; (20) добавляется ещё одно техническое средство для построения структуры (2) – СУБД.</p>

1	2
<p>7. Для агроэкологической оценки земель сельскохозяйственного назначения предлагается следующая атрибутивная БД ГИС. Ключевым компонентом предлагаемой структуры служит участок земной поверхности любого рода (рабочий участок пашни, пастбища, сенокоса и др.). Она должна включать, как минимум, 4 модуля (количество модулей при необходимости может быть расширено), каждый из которых предусматривает несколько позиций:</p> <ul style="list-style-type: none"> – рельеф – основные формы склонов и водосборов; преобладающие экспозиции; распределение углодий по крутизне склонов, выраженное в градусах (в интервалах 0–1, 1–3, 3–5, 5–7, 7–10, 10–15, > 15); степень вертикальной и горизонтальной расчленённости территории; прогноз развития эрозии почв; – почвы – тип; подтип; разновидность почвы; структура почвенного покрова; почвообразующая порода; мощность гумусового слоя; степень деградации; физическое состояние (гранулометрический и структурно-агрегатный состав, плотность, влагоёмкость, порозность, водопроницаемость и др.); содержание органического вещества; рН; ёмкость катионного обмена; содержание питательных элементов; биологическое состояние; фитосанитарное состояние участка; солонцеватость, заболоченность, карбонатность, гидроморфизм и др.; – климат – название метеостанции; ФАР; сумма активных температур >10 °С; минимальная, максимальная и среднегодовая температура воздуха; средние температуры воздуха по месяцам за вегетационный период; даты перехода среднеуточных температур воздуха через 0, 5, 10, 15 °С; даты перехода температуры почвы через 5°, 10°, 15°С; даты наступления поздних весенних и ранних осенних заморозков и продолжительность беззамерозкового периода (средняя, минимальная и максимальная); количество осадков за год, за вегетационный период, по месяцам вегетационного периода; запасы продуктивной влаги в метровом слое (мм); коэффициент увлажнения; тип водного режима; глубина промерзания почвы в зимний период; сроки и скорость снеготаяния и др.; – земельные ресурсы – структура посевных площадей и урожайность культур за последние 5–10 лет; система севооборотов и чередование культур во времени на данном участке; применение приемов обработки почвы; использование удобрений и средств защиты растений и др. 	<p>(22) Упоминается задача (05); (23) с задачей (05) связывается конкретная структура данных (02) с детализированным описанием компонентов, включая (11)–(14); (24) в целом абзац № 7 не богат метаэпиграмами, происходит дальнейшая детализация структур данных (02) в аспектах (11)–(14); (25) данные (24) предлагаются сохранять с помощью атрибутивной БД ГИС.</p>

2	
<p>(26) Термин «оценка» означает обработку данных, а не сбор: следовательно, существует алгоритм;</p> <p>(27) термин «установить» означает отыскание некоторого решения: следовательно, речь идёт об алгоритме;</p> <p>(28) из (26)–(27) делаем вывод, что есть операция обработки собранных данных и есть алгоритмы, включающие некоторое множество таких операций, выполняемых последовательно или параллельно; некоторые из них описаны в (29)–(34);</p> <p>(29) операция обработки показателей тепло- и влагообеспеченности территории;</p> <p>(30) алгоритм оценки биоклиматических показателей;</p> <p>(31) операция вычисления средних дат перехода температура через определённые температурные «пороги»;</p> <p>(32) операция учёта начальных температур прорастания семян;</p> <p>(33) алгоритм оценки потенциальных возможностей культур, сортов;</p> <p>(34) операция учёта свойств позднеспелых сортов;</p> <p>(35) определить целесообразность... – значит, целесообразность в чём-то измеряется, возможно, в баллах. Возможно, это важный показатель;</p> <p>(36) зная средние даты... – в некотором алгоритме используются обработанные данные;</p> <p>(37) из (35)–(36) делаем вывод, что есть важные для ведения сельскохозяйственной деятельности показатели, которые стейкхолдеры стремятся получить. Названия показателей в (38)–(44);</p>	<p>1</p> <p>8. Отдельным атрибутом, интегрирующим показатели тепло- и влагообеспеченности территории, может быть биоклиматический потенциал. Оценка биоклиматических показателей позволяет: 1) установить тепловые ресурсы и влагообеспеченность конкретного землепользования; 2) определить целесообразность возделывания в рассматриваемых условиях тех или иных культур и сортов с учетом их биологических особенностей; 3) зная средние даты перехода температур через определённые температурные «пороги», учитывая начальные температуры прорастания семян и устойчивость растений к заморозкам, определить сроки посева, позволяющие достигнуть посевом фазы полной спелости до наступления ранних осенних заморозков; 4) появляется реальная возможность оценить потенциальные возможности культур, сортов, учитывая, что позднеспелые сорта (гибриды) отличаются более высоким потенциалом урожайности. При этом следует учитывать, что высокий уровень урожайности достигается урожайности, соответствующей потенциальной возможности сортов. Однако урожайность всегда ограничивается климатическими условиями местности, то есть планируемую урожайность не должен превышать величину урожая, обеспечиваемую климатическими факторами. Под климатически обеспеченным урожаем (КОУ) понимают такой уровень урожайности, который можно достичь в идеальных почвенных и агротехнических условиях при ограничивающем действии различных метеорологических факторов. Уровень КОУ, как правило, меньше потенциально возможного.</p>

I	2
<p>9. <i>Текущие (мобильные) данные.</i> Как уже говорилось выше, текущие (мобильные) данные необходимы для поддержки принятий решений при выборе и сопровождении технологий возделывания сельскохозяйственных культур (агротехнологий) с целью управления продуктивностью посевов. Под технологией обычно понимается последовательность производственных процессов, способ использования материалов и машин (оборудования), а также конкретных действий, посредством которых производится конечный продукт. В отличие от промышленных технологий, агротехнологии характеризуются (в дополнение) использованием возобновляемых природных ресурсов и наличием изменчивых условий реализации, которые слабо поддаются управлению, но являются необходимым базисом. Следовательно, агротехнологии – это особый вид технологий, при реализации которых принцип «действуй по прописи» не всегда приводит к планируемому результату. Кроме того, агротехнологии реализуются в системе «человек – машина – растение» с незаменимым, активно действующим агентом. Управляющим звеном в этой системе является человек, но наличие в ней второго биологического агента – растения – преобразует всю систему в вероятностную, так как растения подчинены своим внутренним биологическим законам.</p>	<p>(38) тепловые ресурсы конкретного земельного зования; (39) влагообеспеченность конкретного земельного зования; (40) целесообразность возделывания в рассматриваемых условиях тех или иных культур или сортов; (41) биоклиматический потенциал; (42) биоклиматические показатели; (43) сроки посева, позволяющие достигнуть посевом фазы полной зрелости до наступления ранних осенних заморозков; (44) показатели потенциальных возможностей культуры; (45) приводятся детали алгоритма (33). (46) Разбор структуры данных (03); (47) упоминаются сельскохозяйственные задачи (08)–(09); (48) последовательность процессов означает последовательность изменения цифровых данных о сельскохозяйственном объекте; (49) тем не менее, изменение данных об объекте происходит в любом случае, именно это важно для автоматизации, а если изменение данных происходит, значит реализуются операции и алгоритмы их обработки; (50) безусловно, и растения, и другие природные объекты тоже являются стейкхолдерами со своими особенностями, но в рамках цифровых технологий их интересы на первой стадии цифровой зрелости сводятся к сбору данных, информации и знаний о них, а также о том, в каких случаях эти данные, информация и знания могут измениться;</p>

1	2
<p>10. В настоящее время в проектировании агротехнологий (составлении технологических карт) используются статические (статистические) подходы, как правило, полностью детерминированные (ориентированные на среднестатистическую изменчивость ресурсов). Целесообразно применять в дополнение к первому подходу также вероятностные (стохастические), когда при реализации агротехнологий ведется перманентный эксперимент, дающий информацию о непрерывно и неконтролируемо изменяющихся условиях. Второй аспект более важен, поскольку через него реализуется истинное содержание агротехнологий.</p>	<p>(51) из абзацев (1)–(9) можно сделать предварительные выводы. Есть совокупность сведений о сельскохозяйственном объекте, включающая статические и текущие данные о нём, сельскохозяйственные задачи, решаемые на этих данных, последовательности изменения данных (в силу естественных или рукотворных изменений), наборы операций и алгоритмов обработки этих данных, набор важных для принятия решений агропоказателей. Такой объект в рамках цифровых технологий принято называть цифровым объектом. Под такое определение подпадает любой сельскохозяйственный объект.</p> <p>(52) Статические (статистические) подходы – характеристики операций и алгоритмов обработки данных (28);</p> <p>(53) вероятностные (стохастические) подходы – характеристики операций и алгоритмов обработки данных (28);</p> <p>(54) речь о непрерывном и неконтролируемом изменении цифрового объекта (51).</p>

2	1
<p>(55) Абзац № 11 даёт детализацию структуры данных (3).</p>	<p>11. К текущим данным, поддерживающим наиболее распространенные в растениеводстве России агротехнологии, относятся следующие: перечень культур и сортов, определенных для возделывания в текущем полевом сезоне; содержание доступных элементов питания растений в почве на данном рабочем участке, запасы продуктивной влаги в метровом слое почвы, наличие сорняков и их видовой состав; предшествующие культуры и их урожайность; расчетная урожайность сельскохозяйственных культур, обеспеченная климатическими ресурсами (КОУ); расчетная величина действительно возможной урожайности (ДВУ); приемы обработки почвы, сроки, дозы, способы внесения удобрений и возможные варианты интегрированной защиты растений от вредных объектов (сорняков, болезней и вредителей); технологические карты, с учетом максимальной механизации всех процессов.</p>
<p>(56) К сельскохозяйственным задачам (04) также относится и эта задача – задача получения урожайности, близкой к ДВУ; (57) приёмы обработки почвы, удобрения и набор определённых средств защиты растений – всё это с точки зрения цифровых технологий суть варианты причин изменения данных цифрового объекта (51); (58) группа важных показателей (37) – вероятные отклонения в погодных и хозяйственных условиях; (59) потери урожайности – ещё один показатель (37); (60) значит, на оба направления должны существовать операции обработки данных и алгоритмы (28).</p>	<p>12. Процесс реализации агротехнологий в конечном счете должен решить задачу получения урожайности, близкой к ДВУ. В реальных условиях необходимо с помощью определенных управляющих воздействий на агроценоз (приемы обработки почвы, удобрения и набор определенных средств защиты растений) обеспечить преодоление вероятных минимуме потерь урожайности. Задача как бы делится на два основных направления: оценку оптимального развития посевов в средних для данного участка условиях и постепенное преодоление возможных отклонений от оптимума за счет оперативного вмешательства.</p>

1	2
<p>13. Текущие данные должны дать исчерпывающую возможность для формирования «технологического блока», который имеет целью спланировать комплекс агротехнических мероприятий для получения заданного урожая, включая сроки, объемы, последовательность операций и т.д.; дать набор технологических рекомендаций при возникновении нестандартных условий; дать контрольные параметры для проведения технологических операций; рассчитать календарные сроки проведения операций по уходу за посевами в связи с изменением условий; вести оперативный учет выполнения сроков и качества работ.</p>	<p>(61) Абзац № 13 даёт описание пространства изменений цифрового объекта (51).</p>
<p>14. Оценка развития посевов может осуществляться методами дистанционного зондирования Земли (ДЗЗ) космическими или иными летательными аппаратами. В частности, для этих целей можно использовать сервис ВЕГА, созданный в Институте космических исследований Российской академии наук (ИКИ РАН). Сервис ВЕГА – универсальный инструмент, ориентированный на работу со спутниковыми данными и результатами их обработки при решении задач исследования и мониторинга растительного покрова. Сервис был создан и введен в эксплуатацию в 2011 году. В настоящее время на базе него созданы системы ВЕГА-Science (http://sci-vega.ru/), которая в основном ориентирована на решение задач исследования растительного покрова в рамках различных научных проектов, и ВЕГА-PRO (http://pro-vega.ru/), в основном используемая для решения прикладных задач, связанных с мониторингом лесной и сельскохозяйственной растительности. Также на базе сервиса создано значительное число специализированных информационных систем, ориентированных на решение различных научных и прикладных задач.</p>	<p>(62) Оценка развития посевов подразумевает одновременно сельскохозяйственную задачу (04), показатель (37), алгоритм обработки данных (28), конкретный цифровой объект (51); (63) приводятся технические детали алгоритма (28), результаты работы которого относятся к конкретному цифровому объекту (51); (64) упоминаемые системы – это конкретные источники данных, которые собираются на 3-й стадии цифровой зрелости; (65) задача исследования растительного покрова – задача (04) внутри конкретного цифрового объекта (51); (66) говорится о некотором количестве научных и прикладных задач в пределах цифрового объекта (51).</p>

1	2
<p>15. Оперативное вмешательство в реализацию агротехнологий наиболее целесообразно осуществлять в системе точного земледелия (Precision Agriculture). Точное земледелие – это комплексная система сельскохозяйственного менеджмента, включающая технологию глобального позиционирования (GPS/ГЛОНАСС), ГИС (GIS), технологию оценки урожайности (Yield Monitor Technologies), технологию переменного нормирования (Variable Rate Technology) и технологии ДЗЗ (Remote Sensing of the Earth Technologies). Суть точного земледелия в том, что уход за посевами производится в зависимости от реальных потребностей возделываемых на данном поле культур. Эти потребности определяются с помощью современных информационных технологий, включая космическую и аэросъемки. Наиболее решенным вопросом в точном земледелии является определение оптимального уровня питания растений и применения средств их защиты от вредных объектов, а также определение доз внесения удобрений и препаратов (технология переменного нормирования), уменьшающих негативное воздействие на почву, растения и окружающую среду. ГИС в этом случае является интегрирующей основой для накопления, хранения, обработки, моделирования, интерпретации, анализа и отображения всех собираемых данных, характеризующих посевы, пашню, а также факторы и условия окружающей среды.</p>	<p>(67) Сельскохозяйственный менеджмент – набор задач (04) внутри цифрового объекта (51). Приведён пример технического средства для решения этих задач. Техническое средство по описанию претендует на реализацию задач 6-й стадии цифровой зрелости;</p> <p>(68) предложены технические средства для сбора данных в структуры (02)–(03) и их обработки (28) внутри цифрового объекта (51);</p> <p>(69) суть точного земледелия – особый порядок изменения цифрового объекта;</p> <p>(70) приведены 2 сельскохозяйственные задачи (04) в (64)–(65);</p> <p>(71) определение оптимального уровня питания растений и применения средств их защиты от вредных объектов;</p> <p>(72) определение доз внесения удобрений и препаратов, уменьшающих негативное воздействие на почву, растения и окружающую среду;</p> <p>(73) приводятся детали использования технического средства (16) для сбора данных в структуры (02)–(03).</p>

2. Немаловажную роль среди комментариев занимают сельскохозяйственные задачи: комментарии (04)–(09), (22), (47), (56), (62), (65)–(67), (70)–(72) свидетельствуют о некотором пространстве задач, связанных с процессом возделывания сельскохозяйственной культуры. Структурированная информация о задачах является знаниями и относится к первому этапу цифровой зрелости. В то же время сами задачи решаются начиная с четвёртого этапа цифровой зрелости (см. рис. 9).

3. Сельскохозяйственные задачи, как и любые другие, в рамках цифровых технологий решаются различными методами обработки данных, куда входят всевозможные функции, операции, приёмы и алгоритмы, характеризующиеся стандартной схемой движения информации в информационных системах: вход – обработка – выход. Комментарии (26)–(34), (45), (49), (52)–(53), (60), (62)–(63) свидетельствуют о существовании устоявшегося набора практик обработки сельскохозяйственных данных. Знания о всевозможных методах обработки данных – это первая стадия цифровой зрелости, этап информатизации, хотя сами методы обработки данных применяются на четвертой – шестой стадиях.

4. Завершая движение информации, обработка данных приводит к получению некоторых показателей, считающихся важными для стейкхолдеров процесса возделывания сельскохозяйственной культуры. Комментарии (35)–(44), (58)–(59), (62) указывают на набор показателей, которые связаны с фрагментами структур данных (эти данные нужны для их получения), сельскохозяйственными задачами (нужны для их решения), обработкой данных (вычисляются с её помощью). Знания об этих показателях, несомненно, важны на первой стадии цифровой зрелости, хотя сами эти показатели вычисляются, начиная с четвертой стадии.

5. Комментарии (48)–(51), (54), (57), (61)–(63), (65)–(69) свидетельствуют о необходимости конкретной практики – применения знаний о структурах данных, сельскохозяйственных задачах, методах обработки данных и рассчитываемых показателях к конкретному сельскохозяйственному объекту. Совокупность этих знаний и связанные с ними данные является цифровой информацией о реальном объекте. В частности, комментарий (64) свидетельствует о системах сбора данных,

используемых на конкретном объекте. Таким образом, можно говорить о понятии «цифровой сельскохозяйственный объект», которое находится на стыке парадигм цифровой экономики и аграрной науки. Полностью такой цифровой объект собирается к пятой стадии цифровой зрелости, однако знания о его свойствах и характеристиках – это приоритет первой стадии.

6. Комментарий (03) свидетельствует о том, что цифровой сельскохозяйственный объект не статичен и изменяется во времени вследствие как реальных действий на этом объекте, так и различных косвенных факторов. Это подтверждают комментарии (54), (57), (61), (67). Следовательно, существует совокупность способов изменения конкретного цифрового объекта, включая дополнение данных с помощью различных сервисов, источников и технологий, видоизменение приоритетов и последовательностей решения сельскохозяйственных задач, обработку различных комплектов данных в различных алгоритмах, сравнение различных вариантов обработки, поиск взаимосвязей между результатами обработки данных и выбором комплектов используемых данных с целью варьирования и верификации результирующих показателей. Часто все эти процессы тесно взаимосвязаны и переплетены, поэтому метазнания экспертов и специалистов содержат информацию о различных комплексах их одновременного применения как единого целого, в том числе и оценку результатов этого применения. Комплексы изменений цифровых объектов (причём, как в сторону прошлого, так и в сторону будущего) исследуются начиная с пятой стадии цифровой зрелости, но знания о траекториях их возможных изменений должны собираться на первой стадии.

2.3. Шаблоны знаний

Следующий шаг реализации первого этапа цифровой зрелости заключается в организации сбора данных, информации и знаний, связанных с шестью выделенными группами понятий картины мира процесса возделывания сельскохозяйственной культуры. Это структуры данных, сельскохозяйственные задачи, операции и алгоритмы обработки данных, агропоказатели, цифровые объекты, комплексы изменения цифровых объектов. В этом разделе подробно рассмотрим процедуры сбора знаний.

Начнём рассмотрение с первой группы понятий «структуры данных». По терминологии прикладного системного анализа совокупность понятий картины мира, основанная на шести выделенных группах, представляет собой модель реально действующей системы, а каждая группа – модель её подсистем, также являющимися самостоятельными системами. Ф.П. Тарасенко (2017) выделяет 12 абстрактных свойств системы, которые должны быть отображены в модели [Тарасенко, 2017]. Рассмотрим три из них, наиболее существенных для группы понятий «структуры данных»: состав, структуру и функциональность. Это означает, что при сборе знаний о структурах данных в сельском хозяйстве необходимо иметь набор собираемых данных в виде перечня (возможно, частично упорядоченного), понимать связи между различными элементами этого перечня и знать, зачем эти данные нужны – как по отдельности, так и в виде фиксированных наборов различных элементов перечня. При этом предполагается, что остальные 9 абстрактных свойств системы будут отражены на дальнейших этапах цифровой зрелости. В связи с этим в Лаборатории разработаны таблицы формата MS Excel, изображённые на рис. 10, 11.

ID	Имя	Атрибут 1	Атрибут 2	Атрибут 3	–	Метод 1	Метод 2	Метод 3	–

Рис. 10. Таблица «Классы» для группы понятий «Структуры данных»

ID	Название	Класс 1	Класс 2	Тип зависимости	Направление	Комментарий

Рис. 11. Таблица «Связи» для группы понятий «Структуры данных»

Обе таблицы разработали, руководствуясь двумя соображениями. Во-первых, следовало учитывать текущую цель – задать шаблон упорядочивания знаний о структурах данных в сельском хозяйстве, отразив три упомянутые выше абстрактные свойства систем. Во-вторых, необходимо было учитывать, какими именно техническими средствами и приёмами собранные знания будут использоваться на дальнейших этапах цифровой зрелости. Руководствуясь вторым соображением, выбрали достаточно распространённый язык абстрактной формализации моделей – UML [Фаулер, 2004], подробно рассматриваемый в следующей главе. В связи с этим используемая в таблицах терминология уже сейчас в основном совпадает с терминологией UML, хотя совмещение терминов можно было бы провести и позже, причём необязательно именно с этим языком.

Строка в таблице «Классы» на рис. 10 заполняется по колонкам. Приведём описания колонок с примерами заполнения. Обратим внимание на то, что на первом этапе цифровой зрелости неважно, какова именно система внесения знаний. К единым стандартам, удобным для дальнейшей обработки и связывания с данными и информацией в некотором едином хранилище больших данных, вносимые знания приводятся на следующих стадиях зрелости. Принцип соблюдается один – наличие единой системы, которая отражала бы понимание реальных объектов, описываемых теми или иными терминами. В свою очередь, понимание должно отражать метазнания экспертов.

ID – идентификатор для последующего управления базами данных и знаний через метаданные [Дадян, 2017]. ID должны различаться для любых вносимых знаний как этого, так и других шаблонов упорядочивания знаний.

Имя – отвечает на вопрос «что?», обозначает конкретный вид данных перечня собираемых данных. Такой конкретный вид данных будем в дальнейшем называть «класс» (терминология UML). Примеры: почвенно-климатическая зона, сельскохозяйственная культура.

Атрибут 1, Атрибут 2, Атрибут 3 (при необходимости можно добавить дополнительные колонки с атрибутами) отвечают на вопрос «из чего состоит», обозначают конкретные характеристики вида данных (класса) заполняемой строки. Называть такие характеристики и свойства конкретного вида данных (класса) будем атрибутами. Примеры

атрибутов для почвенно-климатической зоны: широта и долгота; почвенный покров, рельеф, климатические условия; водный бассейн; местоположение относительно городской агломерации; преимущественный агроландшафт и др. Примеры атрибутов для сельскохозяйственной культуры: вид (продовольственная, техническая, кормовая), сорт, урожайность и др. Каждый атрибут (характеристика, свойство) в чём-то измеряется, чем-то определяется. В примечании к ячейке атрибута предлагается это отметить. Например, для почвенно-климатической зоны географические координаты обозначаются в виде градусов долготы и широты в определённом диапазоне, поэтому в примечании к ячейке атрибута «географические координаты» можно записать этот диапазон одним из удобных способов: [30, 80] или 50–60. Водный бассейн определяется объектом из списка водных ресурсов, поэтому в примечании к атрибуту «водный бассейн» должен присутствовать список таких ресурсов, например (Обь, Енисей, Байкал) или более специфичный для территории.

Метод 1, Метод 2, Метод 3 (при необходимости можно добавить дополнительные колонки с методами) отвечают на вопрос, что с этим видом данных можно делать, подразумевая разумные действия по обработке этих данных в рамках решения сельскохозяйственных задач. Хотя сами задачи и алгоритмы обработки данных относятся к другим группам понятий конфигулятора, следует уже сейчас делать шаги в этом направлении для связывания групп понятий в единую модель. Более полно понятие «метод» раскрывается в главе 3. Примеры методов для почвенно-климатической зоны: учитывать или не учитывать её специфику, найти свободную или чем-то занятую зону с нужными характеристиками, классифицировать на предмет сельскохозяйственной пригодности. Примеры методов для сельскохозяйственной культуры: выбрать для возделывания, посеять, проводить уходы за посевом, убрать урожай. В примечаниях к ячейкам методов желательно указывать, как именно в каждом случае будут использоваться атрибуты этой же строки.

Отметим, что в терминологии главы 1 в приведённых примерах показан процесс получения знаний из понимания. Понимание понятий почвенно-климатической зоны и сельскохозяйственной культуры преобразовывается в упорядоченные знания с возможностью даль-

нейшего уточнения. Это несколько другой процесс, чем извлечение знаний из метазнаний эксперта, продемонстрированный выше при составлении конфигулятора.

Строка в таблице «Связи» на рис. 11 также заполняется по колонкам. Каждая строка призвана отразить связь между двумя понятиями, двумя видами данных, двумя классами.

ID – идентификатор с функцией, аналогичной одноимённой колонке таблицы «Классы».

Название. Название связи: должно быть коротким и отражать суть явления или структурного отношения, которое этой связью описывается.

Класс 1. Первый из классов, между которыми устанавливается связь. Рекомендуется не вписывать название класса в эту ячейку, так как названия классов вписываются в таблицу «Классы», и если допустить неточность, то при дальнейшей обработке ИИС будет различать два класса вместо одного. Вместо этого ставится ссылка на ячейку в строке соответствующего класса и столбце «Имя» таблицы «Классы», расположенной в том же файле на другом листе.

Класс 2. Второй из классов, между которыми устанавливается связь. Рекомендации те же, что и к предыдущей колонке.

Тип зависимости между классами с одной стороны ограничивает возможные связи (для этих случаев есть поле комментариев), с другой – позволяет эти связи быстрее выявлять, задавая конкретный шаблон. В таблице «Связи» предлагается 6 типов связей, используемых в UML [Фаулер, 2004] по причинам, описание которых приведено после рис. 11. Вот краткие описания связей.

1. Зависимость – направленное отношение в сторону главного класса (от которого зависит другой), характеристика отношения – изменение свойств зависимого класса при изменении свойств главного класса. Например, урожайность сельскохозяйственной культуры изменяется при изменении характеристик используемых агротехнологий, тогда главный класс – агротехнология, зависимый – культура.

2. Ассоциация – любая симметричная двусторонняя связь, её понимание соответствует термину русского языка «ассоциация». Например, по связи между сельскохозяйственной культурой и почвенно-климатической зоной можно заключить, что в одной зоне можно

выращивать разные культуры и одну культуру можно выращивать в разных зонах.

3. Агрегация – отношения вида «контейнер – часть», направление всегда в сторону контейнера, при этом уничтожение контейнера не влечёт за собой уничтожение части. Например, на одной территории посеяны несколько сельскохозяйственных культур, тогда территория – контейнер, культуры – части. Территорию можно разбить на несколько частей, переназначить, передать управление, но сами культуры так и останутся посеянными на своих участках.

4. Композиция – отношения вида «композит – часть», направление всегда в сторону композита, при этом уничтожение композита влечёт за собой уничтожение всех частей. Например, удобрения и заказ, заказ – это композит, удобрения – части. Удобрения приготовили для выдачи со склада по заказу потребителя, они отдельные части заказа, после того как заказ выполнили, он исчез, исчезли и отдельные объёмы заказанных удобрений, так как слились с общей массой удобрений на складе.

5. Обобщение – отношение «общее – частное» с сохранением свойств. Например, поле – часть сельскохозяйственного предприятия, сельскохозяйственное предприятие – часть почвенно-климатической зоны. При этом в частных классах (их нужно вписывать в таблицу «Классы») не указываются атрибуты общего класса, а только свои, специфические. Направление всегда от частного класса к общему.

6. Реализация – отношение между условиями и их выполнением. Главный класс создаёт условия, зависимый их выполняет, направление в сторону главного класса. Например, агротехнология в виде цикла агротехнических приемов по возделыванию сельскохозяйственной культуры в определённые моменты времени даёт сигнал, например по применению пестицидов: агротехнология – это характеристика сельскохозяйственной культуры, пестициды – это вид агротехнологии. Таким образом, сельскохозяйственная культура – класс, задающий условия, агротехнология – выполняющий эти условия.

Направление. Для ассоциации направление связи двустороннее, в остальных видах связи направление фиксировано. В связи с этим в колонке нужно просто указать направление удобным способом: «двустороннее», «вправо» или «влево».

Комментарий. Вносится дополнительная информация, которая по каким-то причинам не вписана в заданный шаблон описания связей. Комментарии по мере накопления регулярно анализируются и в совокупности могут в какой-то момент служить основанием для изменения шаблона упорядочивания знаний.

Приведённые шаблоны упорядочивания знаний по группе понятий конфигуратора «структуры данных» в рамках формализованной картины мира для процесса возделывания сельскохозяйственной культуры предназначены прежде всего для операторов, не обязательно являющимися экспертами в данной ПО. Это связано с тем, что, как правило, востребованность экспертов высока, и они не обладают временем заниматься заполнением шаблонов. Вместо этого они пишут научные статьи и книги, участвуют во внедрении изобретений и инновационных методов, решают организационные вопросы, требующие высоких компетенций. Для исследователей, начинающих свой научный путь в данной ПО, работа с шаблонами упорядочивания знаний имеет двойное назначение: с одной стороны, они формируют исследовательское мышление, связанное с систематической обработкой профильной информации, с другой – вносят свой вклад в создание ИИС. С точки зрения базовых элементов парадигмы управления бизнес-процессами, приведённых в начале главы, делается существенный вклад в кадры и культуру научно-исследовательского коллектива, ставящего целью разработку ИИС.

Ниже приведены некоторые рекомендации, позволяющие лучше ориентироваться в том, как именно работать с описанным шаблоном упорядочивания знаний о структурах данных в сельском хозяйстве.

1. Работа с информацией. Сначала следует внимательно прочитать научную статью или собственные отчеты о НИР, связанные с ПО. С точки зрения ИИ, это информация, но не знания. Далее из этой статьи или отчетов о НИР следует что-то для себя отметить, понять, взять на заметку. Тогда это уже знания, их можно и нужно вносить в шаблон упорядочивания знаний.

2. Работа со знаниями. В том, что было взято на заметку при работе с информацией, следует попытаться идентифицировать, определить, обнаружить отношение, связь между двумя понятиями, терминами, сущностями, классами. Для этого следует внимательно посмотреть

таблицы «Классы» и «Связи», включая комментарии к заполнению, приведённые выше. Разобравшись, что именно нужно выявлять, следует снова вернуться к тому, что взято на заметку. Скорее всего, это будут два класса и связь между ними. Тогда следует внести две строки в таблицу «Классы» и, как минимум, одну строку в таблицу «Связи».

3. Часто бывает так, что выявленная связь относится не ко всему классу, а к его части. В этом случае нужно предварительно создавать отдельный класс, связывать его с общим классом обобщением, и только затем обозначить связь, которую удалось выявить. По этой же причине часто получается, что атрибут некоторого класса требует дальнейшей детализации и поэтому должен вноситься дополнительно как отдельный класс.

Для сбора знаний, относящихся к группе понятий «сельскохозяйственные задачи», был построен шаблон, явно отражающий два из 12 абстрактных свойств систем: стимулируемость и целесообразность [Тарасенко, 2017]. Это означает, что при сборе знаний о связанных с процессом возделывания сельскохозяйственной культуры задачах необходимо раскрыть причины возникновения и существования задачи, а также разобраться в том, для чего она решается, какая практическая польза появляется в результате её решения. Важное обстоятельство заключается в том, что само по себе информационное поле, связанное с системой сельскохозяйственных задач процесса возделывания сельскохозяйственной культуры, непомерно огромно, об этом написаны десятки монографий практически на всех языках мира, но если ограничить его рамками парадигмы цифровых технологий и сценариями использования (use-case, [Jakobson et al, 2011]) стейкхолдерами ИИС, то получится гораздо более узкая картина мира. Эту картину мира можно охарактеризовать вопросом: «Как могут помочь цифровые технологии более эффективно решать поставленные задачи процесса возделывания сельскохозяйственной культуры?». Предполагая, что остальные 10 абстрактных свойств системы (в том числе напрашивающиеся, но преждевременные для рассмотрения состава и структуры) будут отражены на дальнейших этапах цифровой зрелости, сотрудниками Лаборатории предложена таблица формата MS Excel, изображённая на рис. 12.

Тема	Запрос госаппарата	IT-функция	Продукция агропродавца	Пакет агролабораторий	Услуга агроконсалтинга

Рис. 12. Таблица «Агротемы» для группы понятий «Сельскохозяйственные задачи»

Работа с этим упорядочивающим знания шаблоном требует от заполняющего гораздо более высокого уровня экспертизы по сравнению с предыдущей группой понятий конфигуратора. Однако достаточно быстро после заполнения нескольких строк эксперт сможет охарактеризовать источники информации, формирующие понимание сельскохозяйственных задач в рассматриваемой парадигме и передать работу по внесению знаний операторам, находящимся в начале знакомства с предметной областью. Таким образом, предварительно составленные экспертом примеры не только служат цели формирования базы знаний интеллектуальной информационной системы, но и развивают два базовых элемента управления бизнес-процессами: кадры и культуру.

Рассмотрим колонки таблицы подробнее.

Тема. Краткая характеристика рыночного или государственного интереса, подпадающего под высокий спрос, дотации и другие экономические стимулы, следовательно гарантированно интересные собственниками агрохозяйств – фермерам. Например, такой темой может быть биотопливо [Арутюнов, 2009; Булаткин, 2010]. В отличие от упорядочивающих знания шаблонов для группы понятий «Структуры данных», здесь не требуется специального ID, так как в роли уникального имени выступает это краткое название темы.

Запрос госаппарата. Развёрнутая характеристика темы с точки зрения государства, отдельного министерства или его регионального подразделения, специализированного государственного ведомства. В случае биотоплива развёрнутая характеристика темы может выглядеть так (обстановка на лето 2020 г.):

1. Есть требования ЕС к топливу пролетающих над его территориями самолётов, биотоплива должно быть 20%.
2. Российские компании, производящие топливо, должны добавлять в обычное авиационное топливо биодобавку.
3. Сейчас они вынуждены покупать биодобавку в ЕС за большие деньги.
4. Нужно, чтобы российские фермеры могли эффективно производить сырьё для производства такой добавки внутри страны.

IT-Функция. Ответ на вопрос: «Что будет рассчитывать интеллектуальная информационная система?». Формулировка может быть достаточно развёрнутой и отражать потребности фермеров в расчёте или оценке каких-то показателей. Расплывчатость формулировок также допускается, точность они приобретут при раскрытии других свойств системы сельскохозяйственных задач. Общая характеристика интересов фермеров как стейкхолдеров процесса возделывания сельскохозяйственной культуры может выглядеть так:

1. Фермеры имеют землю и технику, хотят зарабатывать.
2. В этом деле большие риски, самостоятельно делать это невыгодно.
3. Хочется иметь гарантированный доход, а также устойчивость рыночных цен.
4. Получать дотацию от государства, а лучше госзаказ на продукцию.

При заполнении строк отдельных тем проблема становится более специфической.

Продукция агропродавца. Агропродавец – ещё один стейкхолдер процесса возделывания сельскохозяйственной культуры. Агропродавцы поставляют в сельскохозяйственные предприятия «комплектующие», необходимые для проведения полного цикла процесса. Какая продукция фермеру нужна для реализации проектов по данной теме: семена, техника, горюче-смазочные материалы, удобрения, пестициды и др.

Пакет агролабораторий. Агролаборатории – ещё один стейкхолдер процесса возделывания сельскохозяйственной культуры. В ячейке нужно указать, какие конкретные научно-исследовательские коллективы, компании, аналитические лаборатории, заводы занимаются заданной темой (в том числе независимо от цели), какой комплект их

компетенций задействуется в экспертизе темы. В случае биотоплива это будут некоторые характеристики (возможно, ключевые слова) исследований или рекомендаций, связанные с возделыванием различных сортов масличных культур с нужными параметрами, а также информация о том, кто этим занимается. Далее по этим характеристикам можно будет (в том числе и автоматически) искать более специфическую информацию по теме.

Услуга агроконсалтинга. Агроконсультанты – ещё один стейкхолдер процесса возделывания сельскохозяйственной культуры. Компетенции агроконсультантов заключаются во владении цифровыми технологиями в сельском хозяйстве, интерес – в получении заказов на справки, прогнозы и рекомендации, связанные с сопровождением процесса возделывания сельскохозяйственной культуры. В этой колонке нужно охарактеризовать детали таких заказов. В случае биотоплива это может быть, например сравнительная характеристика рабочих участков для посева конкретного сорта масличной культуры или применение специальных агротехнических приемов для увеличения выхода масла.

Таблица «Агротемы» (см. рис. 12) качественно отражает связь между отраслью сельского хозяйства (колонки 1, 2, 4 и 5) и парадигмой цифровых технологий (колонки 3 и 6). Идея в том, чтобы под параметры темы (колонка 2) формировались комплекты продукции (колонка 4) и компетенций (колонка 5). Извлечение метазнаний эксперта в этот шаблон является отражением перебора по этим трём спискам, причём, элемент каждого списка – это набор понятий, который может быть довольно длинным и частично является результатом перебора всех понятий внутри колонки. Вписанные в колонки 2, 4, 5 знания, таким образом, будут задавать запрос на ПО в рамках цифровых технологий. Отражаемые в колонках 3 и 6 эти ПО аналогично отражают перебор уже с позиции интеллектуальных информационных систем. Идея дальнейшей обработки знаний этого упорядочивающего шаблона в том, чтобы автоматически перебирать обе группы списков и искать точки совмещения. Каждая отдельная тема – это набор таких точек.

Прежде чем перейти к разбору упорядочивающих знания шаблонов для понятий группы «Операции и алгоритмы обработки данных», отметим, что при составлении шаблонов, упорядочивающих знания для двух предыдущих групп, использовались стандартные для кибер-

нетики технологии и методики извлечения знаний [Минский, 1979; Осуга, 1989]. Для детализации картины мира по обработке данных этих кибернетических технологий оказывается недостаточно (ниже это будет показано) и в дополнение к ним используются принцип специальной технологии прикладной герменевтики – веерных матриц [Кордонский, Бардин, 2015].

Понятие веерных матриц введено в научную практику С.Г. Кордонским в 1977 г. [Кордонский, 1977]. Логика веерных матриц такова: если существует какая-то система понятий, то 1) эта система описывает какие-то явления; 2) кто-то занимается этим описанием. Таким образом, система понятий существует за счёт совмещения конкретных явлений и тех людей, которые эти явления рассматривают, исследуют. Если нарисовать таблицу, в заголовках колонок которой обозначить явления, а в заголовках строк обозначить их исследователей, то на пересечениях колонок и строк получатся системы понятий. Ясно, что если рассмотреть некоторое количество систем понятий, то именно столько будет строк и столбцов в таблице. Столько же будет систем понятий в диагональных элементах этой таблицы, все остальные клетки таблицы останутся пустыми. Работа прикладных герменевтиков заключается в том, чтобы эти пустые клетки заполнить по принципу совмещения исследователей, занимающихся некоторыми явлениями, с явлениями, которыми занимаются другие исследователи. Смысл этой работы в том, чтобы одной таблицей охватить смежные системы понятий, даже те, которых, возможно, фактически ещё нет. Такие таблицы и называются веерными матрицами. Веерные матрицы допускают вложенность любой глубины, т.е. систему понятий из любой ячейки можно детализировать в новой веерной матрице следующего уровня. Уровней может быть сколько угодно, благодаря чему структура смежных систем понятий приобретает вид пространственных данных, хорошо знакомый специалистам по ГИС-системам.

Теперь рассмотрим третью группу понятий конфигуратора для процесса возделывания сельскохозяйственной культуры – «Операции и алгоритмы обработки данных». Шаблон, упорядочивающий знания, связанный с обработкой данных, выстраивается по принципу веерных матриц, поскольку любая обработка данных является совмещением видов данных (выступают в роли явлений) и видов их обработки (выступают в роли изучающих явления исследователей). Только в случае

применения верных матриц существует гарантия невозможности исключения из рассмотрения какого-либо вида обработки данных.

При разработке упорядочивающего знания шаблона для рассматриваемой группы понятий конфигуриатора учтено отражение трёх из 12 абстрактных свойств системы операций и алгоритмов обработки данных в процессе возделывания сельскохозяйственной культуры: существование в изменяющейся среде, эмерджентность и ингерентность² [Тарасенко, 2017]. Это означает, что при сборе знаний о связанных с процессом возделывания сельскохозяйственной культуры возможностях обработки данных необходимо учесть все условия, в которых это происходит: постоянно изменяющиеся входные данные (один из постулатов ИИС), возникновение новых эффективных алгоритмов из фрагментов других, варьирование применяемых операций в зависимости от требуемых выходных данных. Предполагая, что остальные 9 абстрактных свойств системы будут отражены на дальнейших этапах цифровой зрелости, в Лаборатории разработали две связанные между собой таблицы формата MS Excel (рис. 13,14).

	АТРИБУТ	АТРИБУТ	АТРИБУТ	АТРИБУТ	–
МЕТОД	КЛАСС	КЛАСС	КЛАСС	КЛАСС	
МЕТОД	КЛАСС				
МЕТОД	КЛАСС				
МЕТОД	КЛАСС				
...	...				

Рис. 13. Таблица «Источники данных» для группы понятий «Операции и алгоритмы обработки данных»

Название	Тип	Источник данных 1	Источник данных 2	Источник данных 3	Источник данных 4	–

Рис. 14. Таблица «Последовательности» для группы понятий «Операции и алгоритмы обработки данных»

²(Англ. inherent – являющийся неотъемлемой частью чего-то) – свойство, характеризующее насколько система согласована, приспособлена к окружающей среде, совместима с ней. Степень ингерентности бывает разной и может изменяться в процессе жизненного цикла системы.

Таблица «Источники данных» заполняется следующим образом (см. рис. 13). В двойных заголовках столбцов проставляются класс и атрибут, наименования которых берутся из таблицы «Классы» (см. рис. 10). В двойных заголовках строк проставляются класс и метод, наименования которых также берутся из таблицы «Классы» (см. рис. 10). Ячейка, находящаяся на пересечении строки и столбца, заполняется только в том случае, если между классом столбца и классом строки существует связь в таблице «Связи» (см. рис. 11). Если класс столбца и класс строки один и тот же, то наличие связи подразумевается автоматически, но в этом случае можно обойтись и таблицей «Классы», поскольку назначения в ней колонок с методами и подразумевает обработку данных, соответствующих колонкам с атрибутами. Тем не менее, внесение в таблицу «Источники данных» отношений между атрибутами и методами одного класса ничему не противоречит. Возникающая в этом случае избыточная структура знаний может быть оптимизирована при обработке внесённых знаний на следующих этапах цифровой зрелости. Каждая ячейка несёт смысл источника данных, что становится ясным при рассмотрении примеров, приведённых к описаниям таблиц «Классы» и «Связи». Пусть требуемая к заполнению ячейка находится на пересечении класса «Почвенно-климатическая зона» с атрибутом «Климатические условия» и класса «Сельскохозяйственная культура» с методом «Убрать урожай». Тогда в этой ячейке нужно обозначить знания о данных по сбору урожая некоторой культуры в климатических условиях некоторой почвенно-климатической зоны: это может быть реестр, список, перечень и другое; возможно, эти реестры, списки и перечни где-то существуют и их можно найти, возможно, их придётся формировать.

Все эти соображения необходимо отразить в ячейке в свободной форме, руководствуясь общими принципами работы по преобразованию понимания в знания, сформулированными при описании таблиц «Классы» и «Связи». По мере заполнения новых ячеек в дальнейшем можно уже ориентироваться на понятие «Источник данных» по принципу: если некоторые учёные, авторы некоторых книг или научных статей что-то посчитали и сделали какие-то выводы, то они воспользовались какими-то источниками данных. Информацию о таких источниках и необходимо обнаруживать, понимать и вносить как

знания в упорядочивающий шаблон. Причём, если какой-то источник данных не вписывается в шаблон, то в таблицу «Классы» нужно добавить новые классы, атрибуты и методы, в таблицу «Связи» – новые связи. В предыдущий шаблон придётся добавлять новые классы.

Заметим, что составленные описанным образом ячейки источников данных отражают только элементарные операции обработки данных. Более сложные инструменты обработки данных, описываемые алгоритмами, допускают обращение к нескольким источникам данных в определённом порядке, который может быть последовательным, параллельными или и тем и другим в некотором соотношении. Это отражено в таблице «Последовательности» (см. рис. 14). Рассмотрим порядок её заполнения.

Название. Уникальное название последовательности, смысл этой колонки тот же, что и в предыдущих таблицах.

Тип. Вписывается один из 8 типов последовательностей. Во-первых, 6 вариантов из списка {alt, opt, par, loop, region, neg}. Эти обозначения используются в языке UML [Фаулер, 2004] и обозначают 6 типов последовательностей, характеризующихся следующим образом:

- 1) alt – альтернативная: выполняется одна из нескольких альтернатив;
- 2) opt – опциональная: выполняется только при определённых условиях;
- 3) par – параллельная: несколько параллельно выполняемых задач;
- 4) loop – циклическая: прогон по источникам данных осуществляется несколько раз;
- 5) region – критическая: если начата, должна быть доделана до конца, результат должен быть получен «за один присест»;
- 6) neg – отрицательная, показывает специальные случаи, как не должно быть или происходить.

Во-вторых, два базовых варианта {tree, collect} сценарных фреймов [Минский, 1979]:

- 7) tree – дерево: разветвление вариантов, исходящих из одной точки;
- 8) collect – сборка: множество вариантов сходятся в одну точку.

Источник данных 1, 2, 3, 4. В ячейки эти колонок вписываются ссылки на ячейки таблицы «Источники данных», содержащие знания об используемых в данной последовательности источниках данных.

Строк в таблице «Последовательности» может быть бесконечно много, процесс заполнения этого шаблона упорядочивания знаний

бесконечен. В связи с этим при внесении последовательностей нужно руководствоваться пониманием связанных с процессом возделывания сельскохозяйственной культуры задач из таблицы «Агротемы», прежде всего колонками «Тема и «Услуга агроконсалтинга». Конечно, операции и алгоритмы обработки данных в первую очередь должны иметь смысл в контексте существующих в картине мира задач и форм предоставления стейкхолдерам их решений.

Четвёртая группа понятий конфигуратора для процесса возделывания сельскохозяйственной культуры – «Агропоказатели». Шаблон, упорядочивающий знания, связанные с важными для стейкхолдеров аграрными показателями, разрабатывался с учётом важного абстрактного свойства систем – открытости [Тарасенко, 2017]. Открытость системы означает её непрерывную взаимосвязь с окружающим пространством. В случае цифровой модели процесса возделывания взаимосвязи отражаются через цифры и количественные показатели, представляющие смысл для стейкхолдеров. Подразумевая, что оставшиеся 11 абстрактных свойств системы показателей будут учтены на следующих этапах цифровой зрелости, в Лаборатории разработали табличный шаблон упорядочивания знаний (рис. 15).

№№ пп.	Наименование	Российское название	Пример запроса	Способ расчёта (модель)	Диапазон значений	Международное наименование

Рис. 15. Таблица «Показатели агроконсалтинга» для группы понятий «Агропоказатели»

Таблица предназначена для описания показателей, которые принято считать результатами в растениеводстве, к которым относится и процесс возделывания сельскохозяйственной культуры.

№№ пп. Номер показателя по порядку заполнения таблицы предназначен для удобства формирования ссылок на него, например «Агропоказатель № 75».

Наименование. Уникальное название показателя, смысл этой колонки тот же, что и в предыдущих таблицах.

Российское название. Название показателя, принятое в официальной печати, научных статьях и статистических отчётах государственных ведомств.

Пример запроса. В этой колонке в любой форме приводится запрос условного стейкхолдера (например, фермера или представителя госаппарата), в ответе на который должен фигурировать именно этот показатель. Можно приводить несколько примеров.

Способ расчёта (модель). В этой колонке в свободной форме должна быть размещена информация о том, какова устоявшаяся практика расчёта данного показателя.

Диапазон значений. Единицы, в которых измеряется показатель, и множество допустимых количеств этих единиц.

Международное наименование. Названия показателя на других языках.

Шаблон упорядочивания знаний по аграрным показателям показывает, чего в первую очередь следует добиваться при разработке ИИС, имеющей функцию поддержки принятия решений при сопровождении процесса возделывания сельскохозяйственной культуры. В связи с этим к таблице есть два дополнительных замечания.

1. Разумно разделить показатели на две группы по ориентации стейкхолдеров на коммерческие и научно-исследовательские. Это можно сделать, добавив дополнительную колонку с обозначением типа показателя или же создав ещё одну таблицу аналогичной структуры. Коммерческие показатели – это агропоказатели, которые ИИС предоставляет стейкхолдерам без раскрытия способа их получения (know how). Научные показатели – агропоказатели, имеющие ценность в научном сообществе, которые нужно стремиться получать в научных статьях.

2. Разумно добавить колонку «коэффициент ценности» или «индекс ценности», отражающую ценность показателя: какой коммерческий показатель более «престижно» уметь рассчитать с точки зрения стейкхолдеров или их отдельной группы. Аналогично и с научными показателями: какие вычисленные показатели рецензенты научных журналов (отечественных и зарубежных) считают более ценными, какие – менее ценными. Если опыт и знания эксперта позволяют делать такие оценки, то в ячейках добавленной колонки можно проставлять оценки от 0 до 5. Не исключено, что эти оценки в процессе эксплуатации ИИС изменятся, но ориентироваться всё равно следует на знания экспертов.

В таблице «Показатели агроконсалтинга» колонками, связывающими её с таблицами для предыдущих трёх групп понятий, конфигураторами процесса возделывания являются «Пример запроса» (группа понятий «Сельскохозяйственные задачи», таблица «Агротемы») и «Способ расчёта (модель)» (группа понятий «Алгоритмы и обработка данных», таблица «Последовательности»). В связи с этим при заполнении таблицы приоритет следует отдавать тем показателям, при описании которых в этих колонках возникнут уже используемые ранее знания. Если же этого не происходит, но показатель с точки зрения эксперта важен, следует вернуться к указанным таблицам и внести недостающие темы, источники данных и последовательности (или поставить такую задачу).

Группа понятий «Цифровой объект», пятая группа понятий для конфигуратора процесса возделывания сельскохозяйственной культуры, исследовалась в Лаборатории с точки зрения неразделимости на части – одного из 12 абстрактных свойств систем [Тарасенко, 2017]. В рассматриваемом случае это означает, что любые темы, проблемы, задачи, процессы и другое, выполняемые на реальном сельскохозяйственном объекте должны без всяких исключений быть частью фиксированной структуры знаний о его цифровом отражении. Добиться такого качества цифрового описания можно, приняв за аксиому вытекающее из шести постулатов интеллектуальных информационных систем утверждение: любые данные, попавшие в систему, обрабатываются (если допустить обратное, то это будет означать, что данные или в очереди на обработку, или уже просмотрены системой, но их обработка отложена). Обработка данных без ограничения общности описана в упорядочивающих знания шаблонах «Источники данных» и «Последовательности» (рис. 13, 14). Следовательно, через накапливаемые в этих таблицах знания можно выстраивать фрагменты цифровых объектов, моделирующие различные аспекты функционирования реальных сельскохозяйственных объектов. Один из вариантов такого выстраивания разработан в Лаборатории в виде двух таблиц формата MS Excel (рис. 16, 17).

	Последовательность 1	Последовательность 2	...
Последовательность 1			
Последовательность 2	Источник данных X		
...			

Рис. 16. Таблица «Пересечения последовательностей» для группы понятий «Цифровой объект»

Идентификатор	Конфигурация объектов	Общий источник данных	Объект 1	Объект 2	Объект 3	Объект 4	...	Описание конфигурации	Пример

Рис. 17. Таблица «Конфигурации объектов» для группы понятий «Цифровой объект»

Таблица «Пересечения последовательностей», как и таблица «Источники данных» (см. рис. 13), построена по принципу веерных матриц. В заголовки столбцов и строк вносятся последовательности из таблицы «Последовательности» (см. рис. 15). Ячейка на пересечении столбца и строки заполняется, если у двух последовательностей есть общий источник данных. Знания об источнике данных, напомним, берутся из ячеек таблицы «Источники данных», расположенных на пересечении класса и его атрибута и другого класса и его метода. Таким образом, неявно в каждой ячейке на пересечении столбцов последовательностей содержатся знания о четырех различных классах понятий, связанных с процессом возделывания сельскохозяйственной культуры. Такая взаимосвязь позволяет рассматривать конфигурацию классов, связанных через единый источник данных. Для внесения знаний об этой конфигурации классов разработана таблица «Конфигурации объектов». Классы в связях могут представлять их детализированные компоненты – подклассы, атрибуты, методы, объединить которые можно одним понятием «объект».

Заполнение таблицы «Пересечения последовательностей» ведётся в её части ниже диагонали, поскольку две одинаковые последовательности пересекаются во всех составляющих их источниках данных, а пересечение различных последовательностей не зависит от порядка рассмотрения (рис. 17).

Идентификатор. Уникальный номер для обозначения конфигурации объектов.

Конфигурация объектов. Название должно давать намёк на суть этой сложной связи (полностью связь должна быть описана в колонке «Описание конфигурации»), но быть достаточно кратким.

Общий источник данных. Ссылка на ячейку таблицы «Пересечения последовательностей».

Объект 1, объект 2, объект 3, объект 4. Перечисление классов (или их детализированных компонентов – объектов) пересекающихся последовательностей. Если пересекаемых последовательностей по одному источнику данных больше двух (это означает, что в таблице «Пересечения последовательностей» в разных ячейках будут одинаковые источники данных), то количество объектов вырастает пропорционально (при пересечении двух последовательностей получаем 4 объекта, при пересечении трёх – 6, и т.д.), при этом конфигурация и связанная с ней тема усложняется.

Описание конфигурации. В этой колонке должна быть развёрнуто отражена суть обусловленной пересечением последовательности сложной связи с использованием перечисленных примеров.

Пример. В этой колонке должна быть описана актуальность задачи, т.е. должно быть какое-то подтверждение, что она реально возникает. Это нужно для того, чтобы не множились несуществующие задачи.

Поясним некоторые дополнительные правила работы с пересечениями последовательностей, составленные с учетом оптимизации структуры знаний о цифровом объекте с ростом их количества.

1. Общий источник данных у двух последовательностей может быть только один.

2. Если у двух последовательностей два идущих подряд общих источника данных, то это сигнал на объединение этих двух источников в один с помощью правки таблицы «Источники данных» (см. рис. 13).

3. Если у двух последовательностей два общих источника данных и хотя бы в одной последовательности они идут не подряд, то это сигнал к изменению последовательностей (они в чём-то нелогичны, нерациональны) таким образом, чтобы общие источники данных в обоих шли подряд (таблица «Последовательности», см. рис. 14). Далее применяется правило 2.

4. Если возникла одна из ситуаций, описанных в правилах 2–3, то, возможно, это сигнал к удалению одной из последовательностей или к объединению двух последовательностей в одну с помощью правки таблицы «Последовательности» (см. рис. 14).

5. Если пересечений по источникам данных у какой-то пары последовательностей нет, это согласно методологии применения веерных матриц означает, что таблицы «Классы», «Связи» и «Источники данных» и «Последовательности» имеют пробелы или искажения в заполненных знаниях, требуется вмешательство эксперта. До времени этого вмешательства ячейку можно оставить пустой.

6. Заполнение пустых ячеек без вмешательства эксперта возможно в часто встречающемся случае, когда есть два источника данных, близких по своей сути. Тогда эти источники данных можно подкорректировать или расширить и сделать из них один. Это даст возможность пересечь две последовательности. Работа по оптимизации структуры таблицы «Источников данных» (см. рис. 13) является существенной для третьей стадии цифровой зрелости.

В таблице «Конфигурации объектов» строк должно быть ровно столько (см. рис. 18), сколько различных источников данных содержится в таблице «Пересечения последовательностей» (см. рис. 16).

Шаблон, составленный из двух таблиц, упорядочивающий знания о цифровом объекте, таким образом, даёт доступ к рассмотрению конкретных комплексов условий, в которых задаются вопросы прямой и обратной задачи процесса возделывания сельскохозяйственной культуры («Какую культуру размещать на данном участке?» и «На каких участках целесообразно размещать данную культуру?») и их производные. Совокупность таких комплексов условий и их различные комбинации как раз и задают цифровое отражение реального сельскохозяйственного объекта. Каждый комплекс условий как фрагмент знаний появляется на пересечении последовательностей операций или алгоритмов обработки данных, полученных из некоторых источников, когда у двух и более разных правильно составленных последовательностей обнаруживается один и тот же источник данных. Работает принцип: если в двух различных случаях обработки данных используется один и тот же источник данных, то речь идёт о сложной взаимосвязи, задающей комплекс условий выполнения операций процесса возделывания сельскохозяйственной культуры.

В таблице «Конфигурации объектов» должны быть отражены все темы таблицы «Агротемы» (возможно, по несколько строк на каждую). Формулировки в колонке «Описание конфигурации» должны быть похожими на названия научных статей – это ещё одно указание на заполнение таблицы знаниями. Если какая-то научная статья попала в поле внимания оператора, занимающегося внесением знаний, и эта статья связана с одной из строк таблицы «Агротемы», то под неё нужно выделить 4 объекта (4 компонента различных классов), построить две последовательности с одним и тем же источником данных, выделить скрещивающиеся классы так, чтобы они были разными. Тогда заголовок этой научной статьи можно отразить в таблице. Работает принцип: если какие-то учёные занимаются процессом возделывания в некотором комплексе условий, значит, эти условия достойны рассмотрения как конфигурация объектов.

Для группы понятий «Комплексы изменений цифровых объектов» разработаны упорядочивающие знания шаблоны для отражения последних двух абстрактных свойств систем – целостности и изменчивости во времени [Тарасенко, 2017]. Эти свойства означают, что различные варианты изменений условий существования и функционирования реального сельскохозяйственного объекта – какие бы они ни были: неучтённые, существовавшие ранее или появившиеся вновь – должны быть отражены в данной группе понятий так, чтобы конфигурация картины мира процесса возделывания сельскохозяйственной культуры не изменялся и система правил, по которым он выстроен, оставалась стабильной и неизменной. Любые детерминированные варианты обработки данных и их комбинации уже были учтены в предыдущих упорядочивающих знания шаблонах, поэтому описание конфигурации комплексов изменений цифровых объектов должно включать непредусмотренные этими вариантами ситуации и адекватные с точки зрения метазнаний эксперта способы реагирования на них. Предполагая, что оставшиеся 10 абстрактных свойств систем, применённых к комплексам изменений цифровых объектов, будут учтены на следующих этапах цифровой зрелости, в Лаборатории в 2019 г. разработали следующие две таблицы формата MS Excel (рис. 18, 19).

	Конфигурация объектов 1	Конфигурация объектов 2	...
Объект 1			
Объект 2			
...			

Рис. 18. Таблица «Применимость объектов» для группы понятий «Комплексы изменений цифровых объектов»

Идентификатор	Наименование	Объект	Конфигурация объектов

Рис. 19. Таблица «Прецеденты» для группы понятий «Комплексы изменений цифровых объектов»

Связь между таблицами «Применимость объектов» и «Прецеденты» заключается в том, что во второй таблице ровно столько строк (см. рис. 19), сколько заполненных ячеек в первой (рис. 18). Таблица «Применимость объектов» построена по принципу верных матриц. В заголовках столбцов вставляются названия конфигураций объектов из таблицы «Конфигурации объектов» (см. рис. 17), в заголовках строк – названия объектов, входящих в конфигурацию (из той же таблицы) или любых других объектов: классов, атрибутов и методов из таблицы «Классы» (см. рис. 10). Вносимый объект должен быть максимально детализирован (дополнительную информацию можно внести в примечание к ячейке): это должен быть конкретный представитель класса с набором имеющихся конкретные значения атрибутов и набором конкретных значений методов, применённых к этим конкретным значениям атрибутов. На пересечении столбца и строки выставляется число 0 или 1 по следующим правилам:

- 0 – данная конфигурация объектов неприменима к данному объекту;
- 1 – данная конфигурация объектов применима к данному объекту.

Единица ставится, если в конфигурации объектов есть данный объект или похожий на него: например принадлежит тому же классу, в методе получены схожие результаты, имеет общие атрибуты и др. Критерии «похожести» определяются на стадиях 4–6 цифровой зрелости на основании анализа накопленных к этой стадии данных, информации и знаний. Смысл заключается в том, что если есть похожий

объект, то с какими-то поправками данную конфигурацию можно применить к этому объекту, хотя изначально он ни в этой, ни в какой-либо другой конфигурации не состоял. Применяв конфигурацию к объекту, можно судить о его свойствах без проведения исследований. В этом и заключается принцип адекватного реагирования на нестандартные ситуации с использованием метазнаний экспертов.

Таблица «Прецеденты» – это журнал экспертных решений о совместимости отдельно взятого объекта из таблицы «Классы» и конфигурации объектов из таблицы «Конфигурации объектов» (см. рис. 19). Цель внесения знаний в эту таблицу – точное обозначение ситуации принятия экспертного решения (прецедента) о совместимости через её идентификатор (колонка 1) и наименование (колонка 2). В колонках 3 и 4 указываются объект и конфигурация объектов, совместимость которых определяется.

2.4. Накапливание данных и информации по сельскохозяйственным объектам

В научной литературе обсуждается вопрос организации сбора и накопления данных и информации по сельскому хозяйству [Ерешко, Мединников, 2018; Лазько, 2018]. В частности, предлагается первичную учётную информацию о производственной деятельности отдельного сельскохозяйственного предприятия формировать в виде универсальной структуры (кортежа): вид операции, объект операции, место проведения, кто проводил, дата, интервал времени, задействованные средства производства, объём операции, вид потребленного ресурса, объём потребленного ресурса. Таким образом, вся первичная учётная информация любого предприятия может храниться в единой базе данных в виде указанного кортежа. Более того, учитывая современные возможности облачного хранения информации на основе СУБД, первичная учётная информация всех предприятий может храниться в единой базе данных. При этом также может использоваться единая система классификаторов, справочников, нормативов на основе предварительного онтологического моделирования данных видов информационных ресурсов.

Представленные в предыдущем подразделе главы 2 упорядочивающие знания шаблоны обеспечивают сбор знаний о процессе возделывания сельскохозяйственной культуры в целом. Помимо этого, к первому этапу цифровой зрелости – информатизации – относится и сбор данных с конкретных сельскохозяйственных объектов, на которых осуществляется процесс возделывания сельскохозяйственных культур и могут быть применены данные, связанные с ним. Виды программного обеспечения для такого сбора данных уже упоминались при составлении конфигуратора (комментарии (16) и (20) в табл. 9): ГИС и СУБД. Использование ГИС в сельском хозяйстве подробно описано [Каличкин, Павлова, 2018]. В этом разделе рассмотрен вопрос о структуре данных, подлежащих сбору с каждого конкретного сельскохозяйственного предприятия.

Необходимой предпосылкой для использования данных в ИИС является их формализованное и структурированное представление в формате, пригодном к анализу на ЭВМ. В связи с большим разнообразием возможных источников данных и отсутствием в настоящее время стандартов их сбора и хранения отдельной задачей становится разработка подхода к формализованному описанию, формату обработки, структурированию и хранению данных. В общем виде структура массива данных может быть представлена следующим образом (табл. 10).

Основной массив – это ряды данных об урожайности культур с привязкой ко времени (сезону), месту (полю) и применяемых (примененных) агротехнологиях.

Таблица 10

Общий вид структуры массива данных

Поле	Время	Управляющее воздействие	Культура	Результат
Идентификатор из рубрикатора, включающий данные об истинном географическом положении объекта	Сельскохозяйственный сезон	Агротехнология (комбинация агротехнологий) из классификатора	Сорт культуры в соответствии с классификатором	Урожайность по результатам сельскохозяйственного сезона, выраженная в абсолютных величинах

Объекты исследования – конкретные поля (участки) и культуры, возделываемые на данных участках.

Процессы исследования:

– природные, внешние факторы, не зависящие от действий конкретного субъекта сельскохозяйственного производства;

– агротехнологии, или управляющие воздействия, планируемые и осуществляемые конкретным субъектом сельскохозяйственной деятельности;

– урожайность, являющаяся ключевым индикатором успешности применяемых стратегий и выбора сельскохозяйственной культуры.

Для учета всей доступной информации об объектах и процессах необходимо создать систему сквозной индексации. Каждый из индексов кодирует электронную ссылку на ячейку электронной таблицы, где располагается искомая информация. Одна из главных задач в этом направлении – разработка форматов данных для структурирования информации. Генерация индексов и ссылок происходила в ходе изучения и формализации ПО «агроэкологическая оценка земель», «выбор и размещение культуры», «выбор агротехнологий».

В процессе формализации ПО «выбор и размещение культуры» создается пополняемый список культур и сортов. В процессе формализации ПО «агроэкологическая оценка земель» создается список свойств земельного участка, которые (свойства) могут быть присвоены каждому конкретному объекту (земельному участку) в соответствии с его расположением и результатами натурных обследований данного, соседних или аналогичных (по выбранным признакам) участков. В процессе формализации ПО «выбор агротехнологий» создается пополняемый список применяемых приёмов и агротехнологий. Возможно также формирование комбинаций применяемых агротехнологий в случае необходимости группировки данных в более крупные массивы при недостаточном количестве данных для детального анализа. По мере накопления данных количество комбинаций может увеличиваться, сами приемы детализироваться.

Для структурирования, систематизации и организации хранения информации по ПО используются таблицы в формате MS Excel. Первичные хранилища – память рабочих станций Лаборатории.

На первом этапе создания названных выше таблиц (рис. 20) используется следующая структура: состав полей (столбцы в таблице называются полями, строки – записями), их имена, последовательность их размещения в таблице, тип данных для каждого поля, размер полей, ключевые поля. Также важно было задать параметры, в соответствии с которыми MS Excel автоматически поддерживает целостность данных. Для этого при определении структуры таблиц были указаны ограничения на допустимые значения данных. Поля, по которым часто приходится проводить поиск или сортировать информацию, индексируются. Ключевое поле индексируется автоматически.

В приведенной таблице (см. рис. 20) представлены 23 поля, которые включают практически всю необходимую информацию об объектах и процессах в растениеводстве (год, культура, сорт, урожайность, ID (уникальный номер) рабочего участка, площадь, тип почв, крутизна склонов, экспозиция склонов и др.). Через общие поля можно связать таблицы друг с другом и далее работать с ними, как с одной большой таблицей.

На втором этапе производится создание записей таблицы и заполнение их данными, как представлено на рис. 21.

На следующем этапе для того, чтобы в дальнейшем смоделировать какой-либо процесс управления сельскохозяйственными объектами, с помощью формализованного описания имеющиеся данные преобразуются в специальные формальные объекты матрицы (рис. 22). Это необходимо для более удобного проведения над ними вычислительных и имитационных экспериментов на ЭВМ.

В качестве примера представлена матрица ПО «выбор и размещение культуры», которая включает следующие классы: культура, биология культуры, условия роста и развития, управляющие воздействия.

Матрица представляет собой таблицу (42 × 15), где представлена взаимосвязь между объектами и атрибутами. В столбцах прописаны управляющие воздействия: выбор сорта, оптимальный предшественник, уровень плодородия почвы, система основной обработки почвы, способ посева, сроки посева, качество посевного материала, норма высева, глубина заделки семян, система удобрений, система защиты растений, сроки и способы уборки, мелиоративные мероприятия, экспозиция склонов, крутизна склонов. В строках прописаны биология,

условия роста и развития культуры: равномерность высеваемых семян, прорастание семян и дружность всходов, фазы роста растения от всходов до полной спелости, длина вегетационного периода, густота стояния растений, поражение растений болезнями, вредителями, засорённость посевов сорняками, световой режим посевов, тепло- и влагообеспеченность посевов, эродированность почвы, устойчивость растений к засухе, заморозкам и затоплению, ветровой режим, процесс фотосинтеза, накопление биомассы, обеспеченность растений элементами питания, микроэлементами, количество и качество урожая, биологическая активность, кислотный режим почв и др.

В каждой ячейке ставится 0, если блок класса «Управляющие воздействия» не влияет на факторы того или иного класса, и число от 0 до 1, если влияет, так, чтобы сумма в колонке блока равнялась 1. Эти значения выставляются по известным справочным или экспериментальным данным, а также с привлечением экспертов.

Таким образом, для каждой культуры заполняется таблица из 630 чисел от 0 до 1, сумма которых равна 15. Сравнивая эти таблицы по культурам или их отдельные строки, можно осуществить предварительный выбор культуры, наиболее приспособленной к заданным условиям произрастания.

Для реализации системы сбора, хранения и обработки информации необходимо обеспечить инфраструктуру (рис. 23). Для дублирования (резервирования) информации необходимо применение Backup-Ур – сервера или формирование пространства в облачном хранилище данных. Наиболее ценными данными являются данные сельхозпроизводителей (агропредприятий) по конкретным полям, сезонам, культурам, применяемым агротехнологиям и полученным урожаям. Данная информация подлежит обязательному резервированию.

Используемые данные делятся на три категории:

- данные, получаемые через web-интерфейс от пользователей сервисов и потребителей экспертного ресурса;
- данные, генерируемые экспертным путем, например по анализу литературных источников (публикаций);
- данные из сторонних хранилищ, например данные о погоде.

Возможно несколько ключевых источников данных, необходимых для дальнейшей работы создаваемой ИИС. Первый источник – не-

Данные (сбор, получение данных)			
Web-интерфейс сервиса «агропредприятия»	Экспертная работа		Внешние данные (доступ к внешним хранилищам данных): 1. Климат 2. Погода 3. Почвенные карты
	Программа анализа текстов	Работа эксперта	
Хранение данных			
Рабочие станции обработки данных		Back-Up сервер резервного хранения	
Обработка данных			
Загружаемая на рабочие станции обработки данных экспертная система (ИИ)		Внешняя экспертная система, или система моделирования и визуализации	

Рис. 23. Структура системы сбора, хранения и обработки информации

посредственно субъекты сельскохозяйственной деятельности; за счет вовлечения их в работу обеспечивается набор «опыта» и интерактивная система взаимодействия – «обучение» ИИС. Кроме того, конечной целью ИИС является создание экспертного инструмента в интересах сельскохозяйственных производителей (стейкхолдеров). Для этой цели необходим web-интерфейс (сервер СФНЦА РАН), который будет отражать страницу пользователя для сельскохозяйственного предприятия. По мере формирования сервисов и размещения их на web-интерфейсе возможно увеличение числа пользователей ресурса.

Второй источник – данные, полученные экспертным путем. Большой объем первичной информации присутствует в научных статьях, посвященных интерпретации сельскохозяйственных данных и проверки существующих моделей сельскохозяйственной деятельности. Возможна автоматизация извлечения подобной информации с использованием адаптированных программ анализа текстов.

Третий источник данных – данные из сторонних хранилищ. Так, данные о погоде в привязке к географическим координатам участка и интересующим периодам сельскохозяйственной деятельности возможно получить из открытых ресурсов, таких как <https://meteoinfo.ru/archive-pogoda>; http://pogoda-service.ru/archive_gsod.php и др. В атри-

бутах исследуемых сельскохозяйственных участков присутствует привязка к реальным физическим координатам (широта и долгота), поэтому возможно обращение к существующим ГИС, содержащим данные о земельных участках, включая описание почвенных структур и рельефа местности. Также необходимо обращение к ресурсам, содержащим данные агрохимического, агрофизического и других анализов почв на данном участке с целью формирования базы данных этих параметров.

Таким образом, необходимым элементом IT-инфраструктуры является набор программ с автоматизированными запросами и маршрутизаторами, связывающими сервер хранения данных для дальнейшей обработки с интерфейсом по сбору данных, рабочими станциями экспертов и сторонними серверами, представляющими требуемую информацию через систему автоматически генерируемых запросов.

Данные, собираемые через web-интерфейс, предварительно хранятся в выделенном серверном пространстве (сервер первичных данных) в качестве исходников и дублируются для возможной организации доступа к оригинальному архиву в случае потери данных или эволюции инструментария обработки данных. Данные перенаправляются на серверы сопутствующих сервисов (1С, статистика и др.) а также в обработку для формирования таблиц классификаторов.

Для взаимодействия со сторонними экспертными системами и системами моделирования и визуализации на рабочих компьютерах Лаборатории формируются данные в требуемом формате и набор алгоритмов. В частности, для работы с системой моделирования и визуализации True World для каждой из обрабатываемых моделей создается блок необходимой информации для загрузки в систему. Выгрузка результатов моделирования производится на персональные компьютеры Лаборатории.

Для реализации одного из вариантов накопления данных в Лаборатории в 2019, 2020 гг. создана «Анкета фермера», которая позволяет в режиме онлайн осуществлять сбор информации о сельскохозяйственных предприятиях в систематизированном виде. «Анкета фермера» размещена на странице Лаборатории на сайте СФНЦА РАН. Анкета фермера состоит из 10 таблиц, содержащих 89 параметров, по которым предприятие может представлять максимально полную информацию. На основе полученной информации в «Анкету фермера» формируются следующие базы данных:

- типы и подтипы почв (для территории Западной Сибири);
- виды и сорта сельскохозяйственных растений (для территории Западной Сибири);
- минеральные и органические удобрения;
- геопараметры и агрохимические свойства почв;
- свойства посевного материала возделываемых культур;
- агротехнологии;
- сельскохозяйственные машины и оборудование.

2.5. Ко второму этапу цифровой зрелости

В рамках информатизации процесса возделывания сельскохозяйственной культуры составлена картина мира, полноценность которой подтверждается тем, что шаблоны, упорядочивающие знания по всем шести выделенным из бесед со стейкхолдерами группам понятий конфигулятора, составлены с учётом всех 12 абстрактных свойств систем [Тарасенко, 2017].

Фиксированные упорядочивающие знания шаблоны можно было бы назвать фреймами, по принципу которых разработаны 7 из 10 таблиц («Классы», «Связи», «Агротемы», «Последовательности», «Агропоказатели», «Конфигурации объектов», «Прецеденты»), если бы не методология веерных матриц, используемая при разработке остальных трёх таблиц («Источники данных», «Пересечения последовательностей», «Применимость объектов»). В связи с этим для данного комплекта связанных между собой таблиц, через которые возможна организация сбора знаний по картине мира процесса возделывания сельскохозяйственной культуры, в Лаборатории сложилось конкретное название: «Шаблоны знаний».

Процесс разработки шаблонов знаний объединил в себе методы прикладной кибернетики и прикладной герменевтики в рамках реализации пятого шага прикладного системного анализа для решения проблемы становления интеллектуальной информационной системы с функцией поддержки принятия решений в сфере сельского хозяйства на первом этапе её цифровой зрелости – информатизации. По такому же принципу могут быть построены шаблоны знаний для любой другой области сельского хозяйства, требующей полноценного вписывания в цифровую экономику и индустрию 4.0.

По мере заполнения шаблонов можно организовывать и заполнять базы данных и информационные базы. Базы данных – данные, собираемые в соответствии со структурами данных, применёнными к конкретным агрохозяйствам; информационные базы – базы научных статей и статистических материалов, из которых можно извлечь знания, вписывающиеся в сформированную картину мира. С помощью шаблонов знаний можно автоматически получать элементы семантических сетей и визуализированные представление знаний в виде графов. Знания, собранные с помощью шаблонов знаний, можно автоматически переводить на любой формальный язык, например UML.

Принятие решений – цель построения любых экспертных информационных систем. Решения всегда принимаются на основе выбора альтернатив. Выбор происходит на основе некоторого критерия, возможно многомерного. Критерий – это всегда функция, результат вычисления которой – число. Критерий применяется к альтернативам, получается набор чисел, числа сравниваются, наиболее близкое к заранее заданному эталону соответствует альтернативе, подлежащей выбору. Эталон определяется заранее экспертом ПО.

В области сельского хозяйства любая альтернатива – процесс, поддающийся регулированию в дискретном наборе точек времени. При этом процесс применяется к материальному объекту: земельному участку с конкретными координатами, сельскохозяйственному предприятию с конкретными культурными растениями и породами животных, техническим возможностям с конкретными единицами сельскохозяйственной техники в гараже и удобрениями на складе.

Такой сельскохозяйственный материальный объект функционирует независимо от того, знает ли о нём некоторая ИИС или нет. Постепенно знания об объекте начинают накапливаться через шаблоны знаний, формируется массив данных. Данные начинают обрабатывать некоторыми элементарными функциями обработки данных, в результате получают новые данные – те, которые не удавалось получить непосредственно измерениями. Применение нескольких независимых функций, объединённых в некоторый алгоритм или метод, привязанный к некоторой, обоснованно применяемой с сельскохозяйственной точки зрения математической модели, даёт дополнительную информацию, причём в пределе, при использовании всех возможных для данного сельскохозяйственного объекта моделей, – максимально

полную в рамках, ограниченных применением функций обработки данных и математических методов в сельском хозяйстве.

Для такого информационного отображения реального сельскохозяйственного объекта и требуется принимать решения с помощью экспертной информационной системы. Получившееся информационное отображение максимально заполняется реальными данными, альтернативы подвергаются имитационному моделированию и сравниваются по критерию, затем выбирается одна из них.

В связи с изложенным выше независимо от назначения ИИС (или системы искусственного интеллекта) обладают едиными свойствами в части логики принятия решений.

Например, на заре разработки шахматных программ рассматривались три стандартных алгоритма принятия решений [Карпов, Гик, 1982].

1. Дерево перебора – перебор всех возможных вариантов в порядке адекватности. В применении к сельскохозяйственным системам это означает необходимость разработки пространства агроприёмов – действий агрономов в сельскохозяйственных процессах при наступлении ключевых запланированных или критических (незапланированных) событий. Любые предпринимаемые действия должны быть элементами этого пространства, из множества элементов пространства должен осуществляться выбор. Знания об агроприёмах должны присутствовать в структурах данных. Любые изменения на реальном объекте должны отражаться в цифровом виде, и таким образом каждому агроприёму будет соответствовать изменение данных, информации и знаний в цифровом описании процесса на объекте. Один из путей реализации этого подхода может быть сведён к перебору известных детерминированных моделей, где функция ИИС заключается в применении критерия их адекватности, и далее предлагать результаты фиксированного количества наиболее адекватных моделей.

2. Метод граней и оценок – «умный» перебор, основанных на эвристиках. На основе шаблонов знаний в Лаборатории разработаны три схемы такого перебора.

Схема 1. Выделяем из условий задачи конфигурацию объектов. Формируем окрестность конфигураций объектов по некоторой мере похожести конфигураций объектов. Выявляем из окрестности те конфигурации объектов, которые соответствуют некоторому критерию качества. Тогда выявленную из задачи конфигурацию объектов мож-

но изменить в соответствии с этими конфигурациями так, что качество данной конфигурации повысится. Выявленная закономерность заключается в правиле, по которому некоторая конфигурация объектов переформируется в оптимальный вид.

Схема 2. Выделяем из условий задачи объект. Формируем окрестность объектов по некоторой мере похожести объектов. Выявляем из окрестности те объекты, которые входят в конфигурации объектов заданного качества. Тогда для данного объекта могут существовать конфигурации объектов заданного качества, которые могут быть составлены по образцу выявленных конфигураций, в соответствии с отдалённостью по мере похожести. Выявленная закономерность будет заключаться в правиле, по которому некоторый объект оптимально достраивается до конфигурации объектов.

Схема 3. Выделяем из условий задачи последовательность получения объекта из класса. Формируем окрестность последовательности по некоторой мере похожести последовательностей. Среди последовательностей, попавших в окрестность, выбираем те, которые соответствуют заданному качеству получаемых с помощью них объектов. Тогда с помощью данной последовательности можно получать более качественные объекты, которые могут быть составлены по образцу получаемых через выявленные последовательности объектов, в соответствии с отдалённостью по мере похожести. Выявленная закономерность будет заключаться в правиле, по которому следует получать объект оптимального качества из класса.

Критерии оптимальности во всех трёх схемах связаны с условиями задачи.

Отметим, что все три схемы разработаны в рамках модели клеточных автоматов [Тоффоли, Марголус, 1991]. Клеточные автоматы лежат в основе фундаментальных концепций цифровой физики [Фролов, Ежов, 2000] и цифровой философии [Ллойд, 2014]. Структуры шаблонов знаний (см. рис. 10–19) дают основания полагать, что на основе клеточных автоматов возможно построение фундаментальной концепции цифрового сельского хозяйства.

3. Форсированный вариант. Ситуации, когда работают незыблемые правила, накопленные практическим опытом агрономов, – аксиомы агрономии (см. главу 4). Например, к приемам регулирования режима питания растений относится внесение всех видов удобрений (органических, минеральных, сидеральных, бактериальных) и сокра-

щение потерь питательных элементов. Посевы многолетних трав, сидеральных культур, особенно бобовых – люпина, донника и других, способствуют пополнению почвы азотом и органическим веществом. Специальные приемы обработки почвы, особенно рыхление, повышают биологическую активность почвы и количество легкодоступных элементов питания для растений. Посевы растений, способных усваивать элементы из труднорастворимых соединений в почве, например фосфатов, превращают их в легкодоступные формы. К таким растениям относятся люпин, донник, гречиха, горчица и др. Примером форсированного варианта может служить и технологическая карта, если нет отклонений от запланированного процесса возделывания сельскохозяйственной культуры под влиянием внешних условий.

Таким образом, данные, информация, знания, а также условия их применения в конкретном хозяйстве, собранные в шаблоны знаний и анкету фермера, завершают этап информатизации (первый этап цифровой зрелости) для ИИС, осуществляющей поддержку принятия решений при управлении процессом возделывания сельскохозяйственной культуры. На этапе связанности (второй этап цифровой зрелости) предстоит разобраться, как эти данные, информацию и знания наиболее полноценно использовать при их дальнейшей машинной обработке. Основной критерий полноценности – отсутствие каких-либо потерь на любом этапе обработки, поскольку любая единица данных, информации или знаний может пригодиться на любой стадии формирования решения. Потери исключены при обработке данных анкеты фермера, поскольку эти данные укладываются в стандартную структуру мультитабличной базы данных [Дадян, 2017]. Информацию и знания из шаблонов знаний необходимо автоматически преобразовывать в абстрактный язык моделирования UML.

В главе 3 настоящей монографии представлены примеры использования языка UML для описания информации и знаний, собранных в таблицы «Классы» и «Связи» (см. рис. 11, 12). Это происходило в ручном режиме силами сотрудников Лаборатории. Решение о написании автоматизированной программы преобразования шаблонов знаний в формат UML будет приниматься после оценки применимости UML к описанию остальных таблиц (см. рис. 12–19).

ГЛАВА 3

КОНЦЕПТУАЛЬНОЕ МОДЕЛИРОВАНИЕ

Концептуальное моделирование – способ зафиксировать отражение метазнаний эксперта в некоторой ПО на абстрактном формальном языке для дальнейшей машинной обработки. В предыдущей главе было показано, как это можно сделать с использованием разговорного языка, однако принятого стандарта для обработки знаний, накапливаемых с помощью предложенных шаблонов, не существует. Отсюда вытекают две возможности: разработать новый стандарт или же вписаться в один из существующих. В рамках второго этапа цифровой зрелости ИИС в Лаборатории освоен язык абстрактного моделирования UML.

При программно-аппаратной разработке ИИС менее формализованное построение картины мира о том или ином сельскохозяйственном процессе на втором этапе цифровой зрелости превращается в строго алгоритмизированный процесс концептуализации ПО, в результате которого синтезируется концептуальная модель (КМ) подлежащего цифровизации процесса или явления. В общем виде КМ представляет собой множество объектов, связанных отношениями. Эти объекты, называемые «концепт», «концепция», «понятие», могут иметь различную природу. КМ инвариантна к природе данных, что позволяет применять эти модели в самых разных областях. Отметим также, что КМ определяет не просто понятия ПО, а структуры представления информации об объектах ПО и спецификации поведения для манипулирования объектами [Никаноров, 2009; Вовченко и др., 2010; Богатырев, Вакурин, 2013; Скворцов и др., 2016].

В рамках известных парадигм программирования наиболее распространено объектно-ориентированное (ООП), основными концепциями которого являются понятия объектов и классов [Тузовский, 2019]. Отношения между объектами описываются при помощи наследования и инкапсуляции. Сами объекты вступают в такие отношения (ассоциации), как часть – целое (агрегация), использование и

др. [Буч и др., 2006]. Для создания качественного программного обеспечения целесообразно использовать языки, поддерживающие парадигму ООП, например UML. Использование нотации UML позволяет увеличить скорость разработки программного продукта и уменьшить количество синтаксических, семантических и других видов ошибок [Dobing, Parsons, 2006; Крэг, 2006; Воевода, Зимаев, 2008; Марков, 2013; Сорокин, 2017].

3.1. Абстрактный язык UML

Применение абстрактно-логических языков для формализации знаний необходимо при составлении технического задания на разработку информационных систем интеллектуального анализа данных. Общая идея абстрактно-логических языков – классификация объектов и процессов ПО, их свойств и характеристик. Классификация необходима для разработки структуры хранилища данных и структур отдельных баз данных для разделения программных модулей и функций на отдельные библиотеки. В хорошо разработанной ИИС обе эти задачи должны выполняться автоматически. Для автоматизации требуется поддержка абстрактно-логического языка специальными программами, проектирующими базы данных и структуры программных комплексов. Всеми перечисленными качествами обладает язык UML (Unified Modeling Language) [Rumbaugh et al., 1999].

Язык UML задаёт стандарт описания любой ПО. Стандарт включает набор абстрактных классификационных объектов, а также способы их отображения на плоскости в виде блочных диаграмм. Всего таких диаграмм около двух десятков видов, в каждом из них существуют стандартизированные обозначения [Рей, 2001]. Рисуются диаграммы в специальных визуальных редакторах, разработанных для использования на рабочих станциях ЭВМ. Сохраняются диаграммы в файлах специального формата UML так, чтобы их в дальнейшем можно корректировать в других подобных редакторах, открывать и просматривать в табличном виде, использовать в СУБД.

Существуют два способа проектирования базы данных, связанной с ПО и лежащей в основе ИИС, которая обеспечивает интеллектуальную обработку данных и поддержку принятия решений.

Первый способ – полностью ручной, мануальный. Эксперт описывает то, что хочет, чтобы выполняла программа ЭВМ, т.е. представляет свои знания в виде набора функций обработки данных. Исходя из этого описания разработчики ИИС определяют количество и структуры баз данных, необходимых для выполнения заданного списка функций. Далее происходит программная реализация ИИС, при этом для добавления новых функций весь цикл повторяется заново.

Второй способ – полуавтоматический. Эксперт сразу описывает ПО в формате UML в одной или нескольких диаграммах одного или нескольких видов. Описания в формате UML автоматически переносятся в СУБД, формируя структуру баз данных, и в программную среду разработки ИИС, задавая список функций обработки данных, подлежащих реализации, включая их входные и выходные параметры. Особенности программной реализации в данном случае будут настройка реакции СУБД на новые UML-файлы, а также программирование функций на одном из стандартных языков программирования без необходимости знания особенностей оболочки используемой программной среды.

Последний фактор особенно важен для экспертов, так как освоить один из языков программирования в части реализации математических функций и элементарных алгоритмов достаточно просто. Гораздо более трудоёмким является освоение программных пакетов, в которых эти функции уже написаны, вместе с их оболочками, часто реализованными на иностранном языке. Но ещё более трудоёмкой задачей для эксперта является передача своих знаний специалистам, владеющим языками программирования и функционалом оболочек разработки программного обеспечения.

У языка UML есть обратная функция. По заданному комплекту баз данных можно составить диаграмму ПО, которую этот комплект описывает. То же самое можно делать и с комплектом функций обработки данных. Если известны форматы входных и выходных данных каждой из функций комплекта, то можно построить соответствующий им вариант комплекта баз данных, по которому уже определить ПО, которую описывает данный комплект функций. Благодаря стандарту UML и другим (скрытым от эксперта) технологиям автоматизированного проектирования систем, всё это можно реализовать автоматически [Воевода, Марков, 2014; Токмаков, 2015; Гома, 2016].

Остановимся подробнее на наборе функций обработки данных. Согласно последней строке табл. 1 раздела 1.1 (см. главу 1), знания – это в том числе информация о составе и структуре множества программно-стандартизированных функций, способных эффективно обрабатывать данные, связанные с ПО или её фрагментом. На рассматриваемой ниже в разделе 3.2 диаграмме (см. рис. 25) представлены свойства и характеристики материальных объектов и явлений, имеющих отношение к агроэкологической оценке земель. Связи и наборы атрибутов на этой диаграмме позволяют судить о наборе функций обработки данных, т.е. о знаниях, которые в этой ПО могут существовать в принципе. Из свойств и характеристик какого-либо объекта или явления могут быть получены дополнительные свойства и характеристики того же объекта или явления или другого в соответствии с выставленной связью. Тогда общее количество функций обработки данных для объекта или явления, содержащего n свойств и характеристик, будет

$$n \sum_{k=1}^n C_n^k ;$$

если данный объект или явление оказывает влияние на другой объект или явление, содержащий m свойств или характеристик, то общее количество функций обработки данных, описывающих оба объекта или явления, будет

$$(n + m) \sum_{k=1}^n C_n^k + \sum_{k=1}^m C_m^k ;$$

при взаимном влиянии

$$(n + m) \left(\sum_{k=1}^n C_n^k + \sum_{k=1}^m C_m^k \right).$$

Из формул видно, какова цена как в единицах информации, так и в единицах труда специалистов, разрабатывающих ИИС, полных знаний о наборе взаимно влияющих друг на друга объектов и явлений с их характеристиками и свойствами. Если предположить, что на внедрение одной функции уходит 1 час времени специалиста-програм-

миста, то последняя формула отражает время, которое необходимо затратить на реализацию всего количества функций обработки данных для двух взаимно влияющих друг на друга объектов или явлений. Уже при $n = m = 10$ получаются огромные цифры. Кроме того, способов вычислить одну характеристику по нескольким другим может быть несколько в соответствии с избираемыми парадигмами и моделями. Если не каждому человеку под силу владеть таким количеством знаний, то возможности современных ИИС это допускают. Главное, чтобы знания – информация о функциях обработки данных – вносились и накапливались.

В языке UML всё это многообразие функций описывается абстрактным объектом «метод», являющимся составным элементом абстрактного объекта «класс». Например, представленная ниже в разделе 3.2 табл. 11 содержит методы – функции обработки данных для класса «Рельеф» и классов, на которые этот объект влияет.

Стандарт UML 2.5.1 включает 14 видов диаграмм, состоящих из блоков и связей между ними со специальными обозначениями [OMG®, 2017, с. 685]. Для формально-логического описания процесса возделывания сельскохозяйственной культуры на языке UML в Лаборатории использовалось программное обеспечение Lucidchart (lucidchart.com). Для отражения собираемых знаний об этом процессе через таблицы шаблонов знаний (см. главу 2, рис. 10–19) достаточно использовать четыре вида UML-диаграмм, некоторые результаты освоения одной из которых продемонстрированы в разделах 3.2 и 3.3. Диаграммы последовательностей на языке UML в Лаборатории составлялись на основе алгоритмов машинного обучения, подробно рассмотренных в разделе 5.1.

Структуры таблиц шаблонов знаний отражают следующие соответствия:

- таблицы «Классы», «Связи» – диаграмма классов (Class Diagram);
- таблицы «Источники данных», «Последовательности» – диаграмма последовательностей (Sequence Diagram);
- таблицы «Пересечения последовательностей», «Конфигурации объектов» – диаграмма объектов (Object Diagram);
- таблицы «Применимость объектов», «Прецеденты» – диаграмма прецедентов (Use Case Diagram);
- таблицы «Агротемы» и «Агропоказатели» используются для вы-

страивания очередности построения диаграмм указанных видов на основании приоритета тех функций ИИС, реализация которых наиболее важна для стейкхолдеров.

Несмотря на то что язык UML призван моделировать систему в рамках некоторой ПО, понятие системы в его спецификации обеднено по сравнению с используемым в системном анализе. Согласно UML, существуют три системные категории [OMG®, 2017, с. 12]:

- 1) классификация – множества объектов и связей между ними;
- 2) события – множество возможных событий и их связей между собой и с объектами;
- 3) действия – множество возможных действий и их связей между собой с событиями и с объектами.

Однако при использовании UML для моделирования представленной таким образом системы в семантическом аспекте применяются лишь две категории [OMG®, 2017, с. 13]:

- 1) структурная (статическая) семантика – всё, что может существовать в системе в некоторый конкретный момент времени;
- 2) поведенческая (динамическая) семантика – всё, что относится к последовательности изменения любых элементов системы во времени.

Как видим, эти две категории соответствуют двум категориям структур данных – первой группы понятий конфигулятора картины мира процесса возделывания сельскохозяйственной культуры, выявленной путем анализа экспертного текста [Каличкин, Куценогий, 2019] (см. главу 2). Это одна из причин выбора языка UML в качестве основного для этапа концептуального моделирования в процессе разработки отдельных функций сельскохозяйственной ИИС.

3.2. Концептуальная модель ПО «агроэкологическая оценка земель»

Комплекс предиктивных функций сельскохозяйственных ИИС связан с проектированием адаптивно-ландшафтных систем земледелия (АЛСЗ). Проектирование АЛСЗ начинается с параллельной концептуализации шести ПО (рис. 24).

Совокупность факторов	1	2	3	4	5	6
1. Общественные (рыночные) потребности (рынок продуктов, потребности животноводства, требования переработки продукции)						
2. Агроэкологические требования культур и их средообразующее влияние		22				
3. Агроэкологическая оценка земель (природно-ресурсный потенциал)		2–3	33			
4. Производственно-ресурсный потенциал, уровни интенсификации						
5. Хозяйственные уклады, социальная инфраструктура						
6. Качество продукции и среды обитания, экологические ограничения						

Рис. 24. Совокупность факторов при моделировании АЛСЗ

В рамках этой модели анализируют биологические особенности сельскохозяйственных культур (требования к условиям произрастания) и их средообразующее влияние. Данные требования сопоставляют с агроэкологическими параметрами земельных участков, а затем на основании этого делают вывод о степени пригодности их для использования под ту или иную культуру [Агроэкологическая оценка..., 2005].

Агроэкологическая оценка земель сельскохозяйственного назначения – одна из шести совокупностей факторов, которые необходимо учитывать при моделировании АЛСЗ.

Это моделирование включает следующие задачи:

- систематизацию совокупности факторов влияния на свойства модели АЛСЗ: концептуализация ПО или построение КМ ПО – 6 задач;
- анализ совместимости совокупностей факторов влияния на свойства модели АЛСЗ: попарное совмещение предметных областей – 15 задач.

Предметом рассмотрения в настоящем разделе является задача [33] (см. рис 24). Структурирование информации и построение модели природной характеристики земель (природно-ресурсного потенциала) осуществлялось на языке UML с помощью диаграммы классов, включающей в себя структуру иерархии классов системы, и их атрибуты, а также взаимодействие между классами [Каличкин и др., 2020].

Концептуализация ПО осуществлялась путём выбора ограниченного количества абстрактных объектов из некоторой принятой дескриптивной системы, достаточной для описания ПО. Для задачи [33] выбраны три абстрактных объекта – класс, атрибут и связь, по которым распределены 33 понятия, характеризующие базовые природные характеристики земельного участка. Для 6 понятий выбран абстрактный объект «Класс», для 27 – «Атрибут». Выбор понятий ПО осуществлялся экспертом по АЛСЗ на основании заполненных таблиц шаблонов знаний «Классы» и «Связи» (см. главу 2, рис. 10, 11).

Диаграмма классов для ПО «агроэкологическая оценка земель» содержит 13 связей (рис. 25; обозначены цифрами). Связи отражают взаимозависимости классов между собой вследствие происходящих вне рассматриваемой ПО процессов как внутри соседних ПО из рис. 24, так и вне АЛСЗ. Однако с точки зрения методологии использования UML, количественный состав и качественные характеристики этих процессов не имеют значения и вся совокупность их влияния на элементы ПО описывается классификацией связей (классификация связей кратко рассматривалась при описании таблицы шаблонов знаний «Связи» (см. главу 2, рис. 11).

Так, связи 1, 3–8 (обозначение линий на рис. 25) называются в UML зависимостью (зависимость показывает направленное отношение в сторону главного класса, от которого зависит другой вид отношения – изменение свойств зависимого класса при изменении свойств главного класса), связь 2 – ассоциация (любая связь, характеризующаяся практически любым глаголом русского языка), связи 9–13 – композиции (отношения вида «композит – часть», направление всегда в сторону композита, при этом удаление композита влечёт за собой удаление всех частей).

Используемые типы связей обусловлены следующим: данные классы по своей сути полностью или отчасти – природные характеристики определенного участка земли, который помещен в центре и является композитом. Все происходящие естественные процессы (варьирование агрометеорологического ресурса, формирование рельефа, почвообразование и др.) тесно связаны между собой и их взаимодействия могут быть описаны зависимостью или ассоциацией.

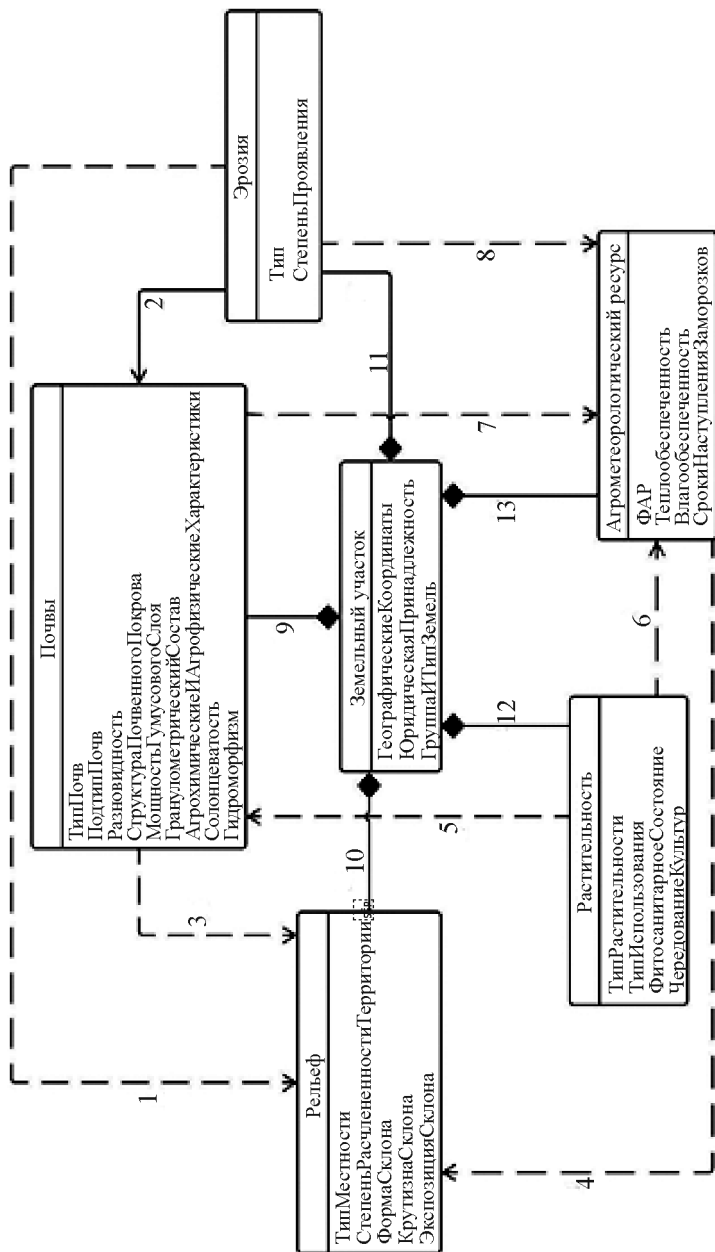


Рис. 25. Диаграмма классов ПО «агроэкологическая оценка земель»

Связь «зависимость» характеризует отношение между двумя классами, когда при изменении одного фактора (класса) изменяется и другой. Например, связь «зависимость» между классами «Рельеф» и «Почвы» показывает, что характер рельефа местности влияет на почвообразовательные процессы (главный класс «Рельеф»). Существуют также методы определения значений атрибутов «Тип» и «Степень проявления» класса «Эрозия» по одному или нескольким значениям атрибутов класса «Рельеф». Эти методы отражают фундаментальные законы воздействия экзогенных и эндогенных процессов на земную кору, а также зависимость степени проявления эрозии почв от рельефа. Например, смыв почвы увеличивается прямо пропорционально уклону. Увеличение уклона почвы от 2 до 4° приводит к увеличению смыва почвы в 1,8 раза, от 4 до 8° – в 7,2 раза. Протяженность склона также влияет на водную эрозию. Согласно данным А.Д. Орлова, смыв возрастает при удвоении линии стока от 50 до 100 м в 2,9–3,7 раза [Орлов, 1983].

Связь «ассоциация» показывает отношения объектов одного класса с объектами другого, например связь между классами «Почва» и «Эрозия» показывает, что эрозия является одной из характеристик почв. Такой связи можно присвоить имя «реализуется на» (эрозия реализуется на почве). Связь 2 говорит об ассоциации или о совмещении значений атрибутов классов «Эрозия» и «Почвы». Совмещение означает, что можно говорить, например об отражении типа эрозии в показателях плодородия почв или степени проявления эрозии в конкретном типе почв. При необходимости дальнейшей детализации ПО для каждого такого сочетания можно формировать отдельный класс, связанный с двумя данными.

Связь 9 указывает на бессмысленность рассмотрения каких-либо значений атрибутов класса «Почвы» без привязки к одному или нескольким атрибутам класса «Земельный участок». Формулировка может быть такой: зачем узнавать параметры почвы, если неизвестно, где она располагается, кому принадлежит и как используется в отнесении земель к конкретной группе и типу по пригодности к использованию в сельском хозяйстве.

Добавим, что атрибуты являются связующим звеном между моделью, методами, потоками данных и запросов системы, поскольку

ку, с одной стороны, они входят в состав классов, участвующих в сценариях диаграмм последовательности, с другой – принадлежат к внешней оболочке модели (числовые параметры, математические модели).

Важно, что при построении диаграмм соблюдается принцип исключения повторности, так как класс не допускает возможности наследования у самого себя прямо или косвенно одинаковых атрибутов, а класс-ассоциация не может иметь те же атрибуты, что и связанные с ним классы, атрибуты не должны пересекаться и т.д. Кроме того, классы – это комплекты характеристик без привязки к обладающим этими характеристиками объектам. Связи между классами указывают на ограничения, в рамках которых характеристики могут меняться. Диаграмма классов является одним из основных этапов в проектировании объектно-ориентированной системы. Описание классов используется на разных этапах проектирования системы и включает разную степень детализации.

Для усложнения и оптимизации модели, введения новых конструкций или их замену на более подходящие можно использовать бинарные диаграммы. Вопрос по оптимизации конкретной модели, которая описывается с помощью диаграмм классов, всегда остается открытым. Со временем с целью усовершенствования и более глубокой детализации могут быть добавлены новые классы и элементы.

Представленный результат решения задачи [33] является промежуточным. Для того чтобы его можно было применить к задаче [2–3], несомненно относящейся к процессу возделывания сельскохозяйственной культуры, необходимо учесть остальные 5 групп понятий конфигулятора, отраженных в восьми оставшихся таблицах шаблонов знаний (см. главу 2, рис. 12–19). В первую очередь:

- ввести в перечень понятий ПО набор алгоритмов, с помощью которых возможно получение расчетных величин и экспертных заключений, в классы диаграммы добавить описывающие этот набор алгоритмов методы;

- ввести в перечень понятий ПО внешние процессы, косвенно влияющие на характеристики элементов ПО, и связать их с обозначенными на диаграмме 13 связями;

– для всех 27 атрибутов классов указать числовые диапазоны значений или списки возможных текстовых значений.

Перечисленные шаги относятся к динамической семантике UML, самый простой компонент которой – метод. В связи с этим КМ ПО «агроэкологическая оценка земель» должна содержать ещё один тип принятых в UML абстрактных объектов – методы. Методы включены в таблицу шаблона знаний «Классы» (см. рис. 10), в UML относятся к диаграммам классов и отражают всевозможные способы преобразования выявленных через атрибуты классов структур данных в другие имеющие смысл в рамках ПО, т.е. прикрепленные к конкретным абстрактным объектам её КМ данные.

Абстрактный объект UML «метод» – наиболее простой и интуитивно понятный среди объектов, используемых для реализации процедур решения конкретных задач (группа понятий «Операции и алгоритмы обработки данных» в главе 2). Другие виды таких объектов, например *последовательности*, используются для более глубокой детализации и усложнения структуры классов и связей КМ ПО.

Методы в UML привязаны к классам и отражают возможности использования атрибутов этого класса для получения значений других атрибутов этого же класса или атрибутов других классов через связи между классами. Получение значений выполняется через известные или пока не известные процедуры решения конкретных задач: это могут быть детерминированные или стохастические алгоритмы, статистические и другие обрабатывающие ряды данных методы, методы анализа данных и искусственного интеллекта. Иногда реализацией метода может быть устоявшаяся форма расчёта одного параметра по другим, известная в виде конкретной формулы.

Таким образом, в UML метод класса – вычислительная процедура, аргументами которой являются значения (или одно значение) атрибутов этого же класса. На выходе такой процедуры может получиться другой атрибут этого же класса или атрибут класса, имеющего связь с данными. Графическая нотация класса состоит из двух частей – заголовка с именем класса и тела с описанием его полей (атрибуты – в терминах UML) и методов (операций – в терминах UML) [Буч и др., 2002; Фаулер, 2004; OMG®, 2017].

В рамках ПО «агроэкологическая оценка земель» основная цель применения методов – восполнить пробелы в имеющихся данных о землях и природных условиях не измерительными методами, а с помощью автоматических расчётов на ЭВМ и в ИИС.

Прежде чем переходить к описанию методов расчёта одних характеризующих природные объекты и комплексы величин через другие, заметим, что все они обусловлены нахождением земельного участка в природной среде, значит, выполняются под влиянием и с учётом внешних природных процессов. Набор природных процессов – важная часть ПО, которая сформируется после понимания того, какие комплекты внешних природных процессов учитывает каждый из выбранных методов.

Для отражения практики применения методов в рассматриваемой ПО в качестве примера выбран класс «Рельеф» (табл. 11). Строки – наборы атрибутов класса «Рельеф», подаваемые на вход метода, в столбцах – значения выдаваемого на выходе атрибута этого же или другого класса.

Цифрами 1–5 обозначены атрибуты класса «Рельеф» сверху вниз согласно их порядку на диаграмме (см. рис. 25). Они дают 31 комплект входных параметров, из которых с помощью методов можно получить или 5 атрибутов класса «Рельеф», или 18 атрибутов четырёх связанных классов (связи 1, 3, 4, 10, см. рис 25). Аналогичны другие обозначения: э1, э2 (2 атрибута класса «Эрозия», связь 1); п1, п2, п3, п4, п5, п6, п7, п8, п9 (9 атрибутов связанного класса «Почвы», связь 3); а1, а2, а3, а4 (4 атрибута связанного класса «Агрометеорологический ресурс», связь 4); з1, з2, з3 (3 атрибута связанного класса «Земельный участок», связь 10). В ячейках таблицы цвет означает наличие метода, разные цвета – разные виды методов. Расшифровка цветов – внизу таблицы.

Пример 1. Для отражения связи между крутизной склонов (р4) и степенью проявления эрозии (э2) – в таблице это фиолетовая клетка на пересечении строки с обозначением «4» и столбца с обозначением «э2» – можно использовать статистический метод, предложенный М.Н. Заславским [Заславский, 1986] (табл. 12).

Класс «Рельеф»: методы

	1	2	3	4	5	п1	п2	п3	п4	п5	п6	п7	п8	п9	э1	э2	з1	з2	з3	а1	а2	а3	а4	
1		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
2	■		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
3	■	■		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
4	■	■	■		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
5	■	■	■	■		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
1.2		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
1.3		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
1.4		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
1.5		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
2.3	■		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
2.4	■		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
2.5	■	■		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
3.4	■	■	■		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
3.5	■	■	■	■		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
4.5	■	■	■	■	■		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
1.2.3		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
1.2.4		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
1.2.5		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
1.3.4		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
1.3.5		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
1.4.5		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
2.3.4	■		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
2.3.5	■		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
2.4.5	■		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
3.4.5	■		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
1.2.3.4		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
1.2.3.5		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
1.2.4.5		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
1.3.4.5		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
2.3.4.5	■		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
1-5						■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■

Расшифровка цветов:

■	– метод невозможен, так как параметр находится среди входных;
■	– существует математическая формула;
■	– существует статистический метод;
■	– существует метод математического моделирования;
■	– существует метод искусственного интеллекта;
■	– логика предметной области требует разработки метода;
■	– метод неизвестен

Таблица 12

Интенсивность эрозии в зависимости от крутизны склона

Крутизна склонов, град.	Интенсивность эрозии, т/га в год
0–1	От незначительной (<0,5) до средней (1–5)
1–3	От слабой (0,5–1) до сильной (5–10)
3–5	От средней (1–5) до очень сильной (> 10)
5–7	От сильной (5–10) до очень сильной (> 10)

Пример 2. Для отражения связи между рельефом (р1–5) и почвенным покровом (п1–п4) – в таблице это желтые клетки на пересечении строки с обозначением «1–5» и четырех столбцов с обозначениями «п1», «п2», «п3», «п4» – необходимо разработать метод.

Для разработки метода может быть использован закон аналогичных топографических рядов [Захаров, 1927]. Сущность этого закона заключается в том, что в любой зоне распределение почв на элементах рельефа имеет аналогичный характер: на возвышенных элементах залегают почвы, генетически самостоятельные (автоморфные), которым свойственны вынос подвижных продуктов почвообразования и аккумуляция малоподвижных; на пониженных элементах рельефа (шлейфы склонов, днища низин и западин, приозерные понижения, пойменные террасы и др.) расположены генетически подчинённые почвы (полугидроморфные и гидроморфные) с аккумуляцией подвижных продуктов почвообразования, приносимых с поверхностным и внутрпочвенным стоками с водоразделов и склонов; на склоновых элементах рельефа залегают переходные почвы, в которых по мере приближения к отрицательным формам рельефа возрастает аккумуляция подвижных веществ.

Пример 3. Для отражения связи между типом местности, экспозицией склона и влагообеспеченностью – в таблице это голубые клетки на пересечении строки с обозначением «1,5» и столбца с обозначением «а2», «а3» – можно использовать методы искусственного интеллекта (ИИ). Тип местности подобно ландшафту или району является одним из наиболее распространенных и важных понятий ландшафтной географии. Эту категорию классифицируют также по условиям увлажнения: «переувлажненные», «влажные», «умеренно увлажненные», «засушливые» и «острозасушливые». Для использования одного из методов ИИ необходимо базовую диаграмму UML (см. рис. 25) дополнить, а именно в классе «Рельеф» создать подкласс «Тип местности» (на основе одноимённого атрибута) и добавить в этот подкласс атрибуты «по условиям увлажнения», «по рельефу», «по ветровому режиму». Дифференциация подкласса «Тип местности» может быть продолжена до создания атрибутов по условиям увлажнения (рис. 26). Аналогичным образом детализируем класс «Агрометеорологический ресурс».

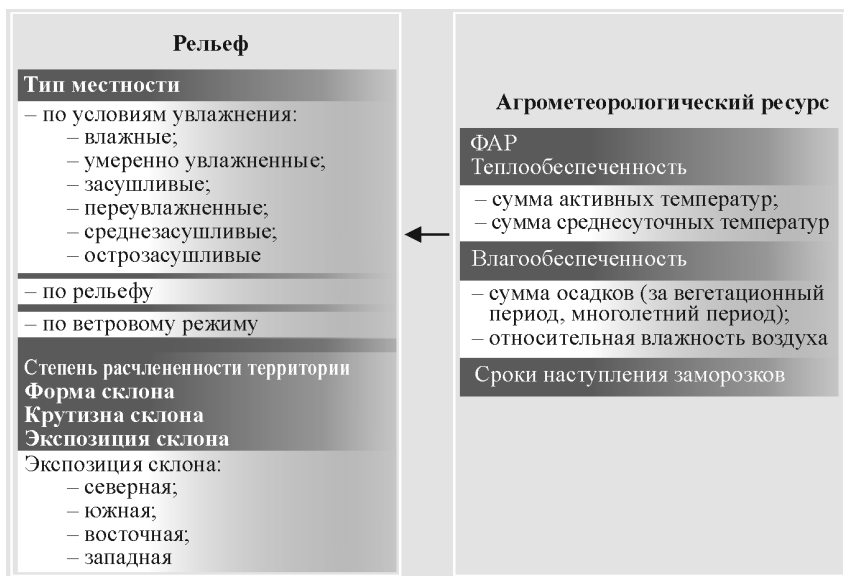


Рис. 26. Детализация классов – «Рельеф» и «Агрометеорологический ресурс»

Определение типа местности по условиям увлажнения возможно на основе использования искусственной нейронной сети (ИНС) и дерева моделей M5 Tree (которое также применяется для прогнозирования прогревания почвы) [Мансуров, 2006]. Построение модели начинается со сбора первичных данных. В нашем случае – это гидрометеорологические параметры, так как они составляют основу водного (сумма осадков) и теплового балансов (испаряемость). Для более детального анализа и прогноза список используемых атрибутов может быть расширен и задействованы такие параметры, как форма склона и угол наклона рельефа, условия дренированности, гранулометрический состав почв и др. Это позволит получить оптимальную модель на основе ELM (Extreme Learning Machine) [Huang et al., 2006] для оценки влажности исследуемого участка, учитывая экспозицию, рельефно-ландшафтные характеристики и сопоставляя с соответствующей комбинацией метеорологических данных.

Алгоритм обучения ИНС можно провести методом обратного распространения ошибки. Обучение ИНС содержит три основных ста-

дии: первая – подача на входы сети обучающих данных (X_1 – данные по рельефу, X_2 – данные по экспозиции склона, X_3 – данные по теплообеспеченности, X_4 – данные по влагообеспеченности), вторая – противоположное распространение погрешности, третья – корректирование весов. На выходе получаются векторы ответов по определению типу местности по условиям увлажнения. В результате благодаря полученной зависимости можно по данным, связанным с атрибутами «Тип местности» и «Экспозиция склона» класса «Рельеф», оценивать данные, связанные с атрибутами «Теплообеспеченность» и «Влагообеспеченность» класса «Агрометеорологический ресурс», реализуя заявленное предназначение UML-метода как абстрактного объекта в концептуальной модели предметной области.

Представленный в табл. 11 фрагмент знаний о взаимозависимостях данных ПО «агроэкологическая оценка земель» – важный инструмент в ИИС, одной из функций которых является заполнение пробелов в информации. Например, «Аграрная интеллектуальная система» [Каличкин и др., 2019], подробно рассматриваемая в главе 4 настоящей монографии, имеет компонент «Доформализация входных данных». На практике любая задача земледелия (растениеводства) связана с конкретным земельным участком, а значит, применима концептуальная модель ПО, отображаемая на диаграмме (см. рис. 25). Однако про этот участок известна не вся информация, и даже при наличии большого количества данных нет никакой гарантии, что они не устарели. Для заполнения пробелов в знаниях ИИС обращается в таблицы методов, находит в них комплекты входных данных, соответствующие уже имеющимся по участку данным, по этим комплектам определяет методы для расчёта недостающих величин. Если метода нет, сигнализирует о необходимости его разработки. Для некоторых задач нужны не все данные, поэтому доформализация входных данных может происходить по разным сценариям с использованием разных комплектов методов. Методология применения и практическая реализация одного из методов, соответствующего группе методов голубого цвета описана в разделах 5.3 – 5.6.

Ещё один пример концептуализации сельскохозяйственной ПО с помощью языка UML рассматривается в следующем разделе.

3.3. Концептуальная модель ПО «выбор и размещение культуры»

Для построения концептуальной модели ПО «выбор и размещение культуры» также использовались диаграмма классов UML (рис. 27) и заполненные таблицы шаблонов знаний «Классы» и «Связи». Центральным классом в модели является «Культура», остальные четыре класса характеризуют факторы, влияющие на принятие решения о выборе и размещении культуры на данной территории (Каличкин и др., 2019а).

Концептуализация двух смежных предметных областей «Агроэкологическая оценка земель» (раздел 3.2) и «Выбор и размещение культуры» позволяет углубиться в описание картины мира процесса возделывания с помощью оставшихся 8 таблиц шаблонов знаний. В задаче совмещение культуры и территории (задача |2–3| в таблице на рис. 24) заполнение таблицы «Применимость объектов» (см. рис. 18) позволяет выделить и доработать до использования в машинной обработке данных её фрагмент, связанный с критерием* [сноска: в описании этой таблицы в главе 2 использовалось обусловленное контекстом выражение «мера похожести»] влияния атрибутов класса «Управляющие воздействия» (УВ) на атрибуты классов «Биология культуры» (БК), «Условия роста и развития» (УРР) и «Размещение» (Р). Этот доработанный фрагмент таблицы «Применимость объектов» в работе (Каличкин и др., 2019а) получил наименование «Управляющая матрица культуры» (табл. 13).

Таблица 13

Управляющая матрица культуры

	УВ 1	...	УВ7
БК 1			
...			
БК 6			
УРР 1			
...			
УРР 11			
Р 1			
...			
Р 5			

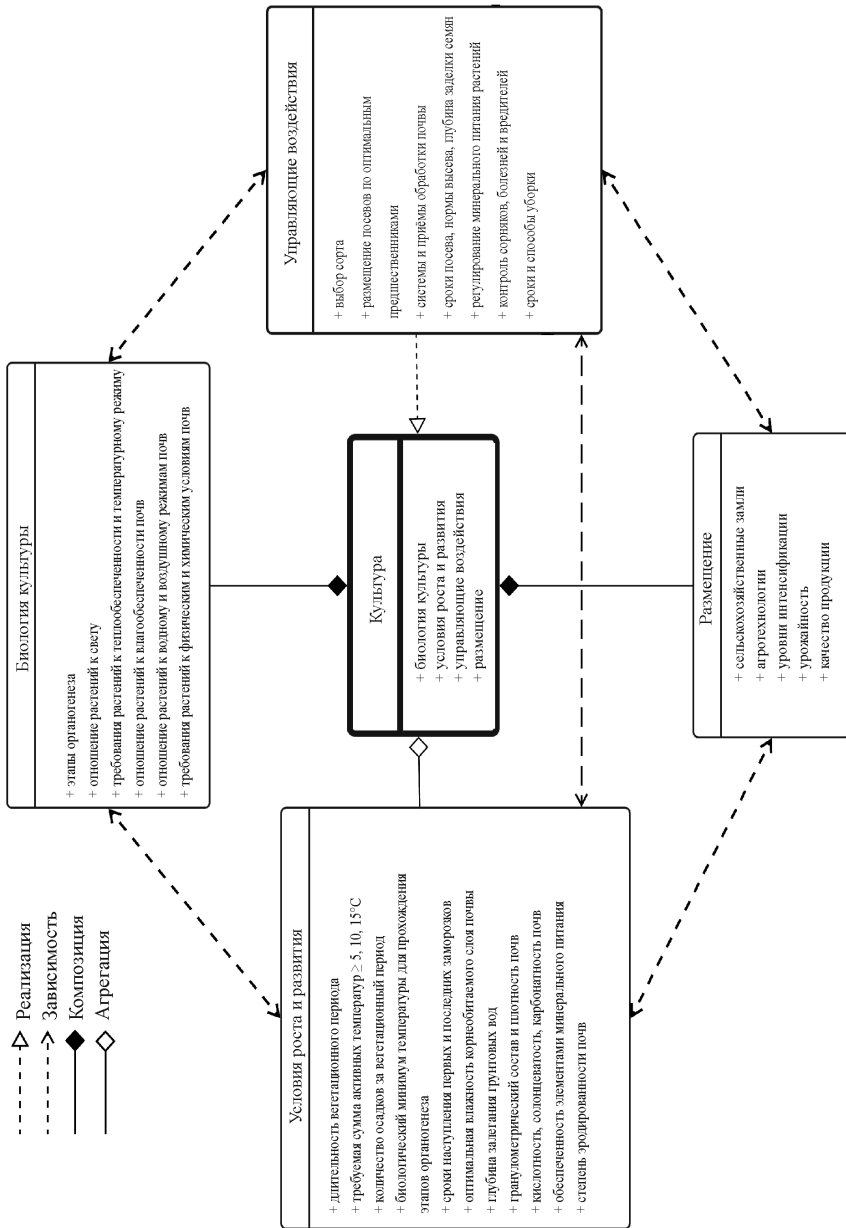


Рис. 27. Диаграмма концептуальной модели ПО «Выбор и размещение культуры»

При управлении процессом возделывания сельскохозяйственной культуры с помощью 7 блоков, перечисленных в классе УВ, осуществляется влияние на 22 фактора, содержащихся в БК, УРР и Р. Управляющая матрица культуры увязывает 7 блоков управления развитием растения (УВ 1 – УВ 7) и остальные 22 фактора (БК 1 – БК6, УРР 1 – УРР 11, Р 1 – Р 5) в 154 ячейки со значениям от 0 до 1 (в отличие от таблицы «Применимость объектов», где допускаются значения 0 или 1) так, чтобы сумма в каждой колонке равнялась 1. Эти значения выставляются по известным справочным или экспериментальным данным с привлечением экспертов. Например, если к началу вегетации весной плотность стеблестоя недостаточная или наблюдается задержка развития посева (поздний сев, поздние всходы из-за недостатка влаги, вымерзание озимых и др.) целесообразно в первую подкормку дать повышенную дозу азотных удобрений и провести ее как можно раньше, что дает возможность стимулировать кущение. Этой же цели служат вторая азотная подкормка и ранняя (в начале допустимых сроков) обработка ретардантами, которые, обеспечивая устойчивость к полеганию, стимулируют кущение. Данное положение иллюстрирует влияние удобрений, упомянутых в классе «Управляющие воздействия», на рост и развитие растений.

Сравнивая управляющие матрицы различных культур (а также их отдельные строки или наборы строк), можно осуществить выбор культуры. Одной из мер может быть сравнение плотностей в строке урожайности. Существенное значение в этом подходе имеет корректное заполнение управляющих матриц и выбор критерия их сравнения. На практике для этого сначала привлекается эксперт, затем применяются технологии машинного обучения [Alpaydin, 2009; Witten et al., 2016]. Имея верхние и нижние оценки показателей для каждой культуры, можно с помощью того или иного критерия выбрать культуру для размещения на данной территории. Сравнение управляющих матриц с точки зрения UML является методом класса «Культура», поскольку при сравнении управляющих матриц используются атрибуты классов, являющихся его подклассами.

Ещё один метод класса «Культура», также возникший при заполнении таблицы шаблонов знаний «Применимость объектов», требует расширения диаграммы классов для КМ ПО «выбор и размещение

культуры» и представляет собой схему подготовки принятия решения по управлению процессом возделывания культуры (рис. 28). Перед применением того или иного управляющего воздействия проводится имитационное моделирование с целью прогнозирования эффектов этого воздействия. Это стандартная функция системы поддержки принятия решений [Heinemann, 2010; Jakku, Thorburn, 2010; Zhu et al., 2018]. Представленная на рис. 28 схема позволяет проследить возможности управления процессом возделывания культуры от класса «Управляющие воздействия» (УВ) до класса «Обоснование и формулировка оптимального конечного результата» (ОФОКР).

Звенья управления обозначены цифрами 1–4 в красных рамках. Вся цепочка управления, закладываемая в основу имитационного моделирования, проходит через классы УВ (7 атрибутов), «Культура» (4), «Оперативное управление» (3) и ОФОКР (1). Легко видеть, что всего цепочек управляющего воздействия будет 84 ($7 \times 4 \times 3 \times 1$). Каждую из этих цепочек – обозначать их можно первыми тремя компонентами, так как последнее звено будет для всех цепочек одинаковым из-за отсутствия атрибутов класса ОФОКР – можно оценить баллом от 0 до 1, руководствуясь критерием соответствия некоторой идеальной агрономической стратегии.

Имея рейтинги цепочек управляющего воздействия, можно получить самую сильную и самую слабую стратегию лица, принимающего решение (ЛПР) – агронома. Самая слабая стратегия даёт наименьший эффект управления, самая сильная – наибольший. Самая слабая стратегия – хотя и ограниченный, но гарантированный фактор управления: если возможно управлять с помощью этой стратегии, то сработают и другие, более сильные стратегии. Тогда показатели, полученные или оценённые при этой стратегии, будут нижними оценками. Аналогичный результат получится и для самой сильной стратегии. Если при достижении некоторых показателей не срабатывает самая сильная стратегия, то не сработают и более слабые. За счёт этого принципа можно получать верхние оценки для показателей. С помощью подобного механизма Ю. фон Либих [Либих, 1936] сформулировал закон ограничивающего фактора.

Качество этого подхода зависит от корректного оценивания стратегий агронома – цепочек управляющего воздействия. Машина легко

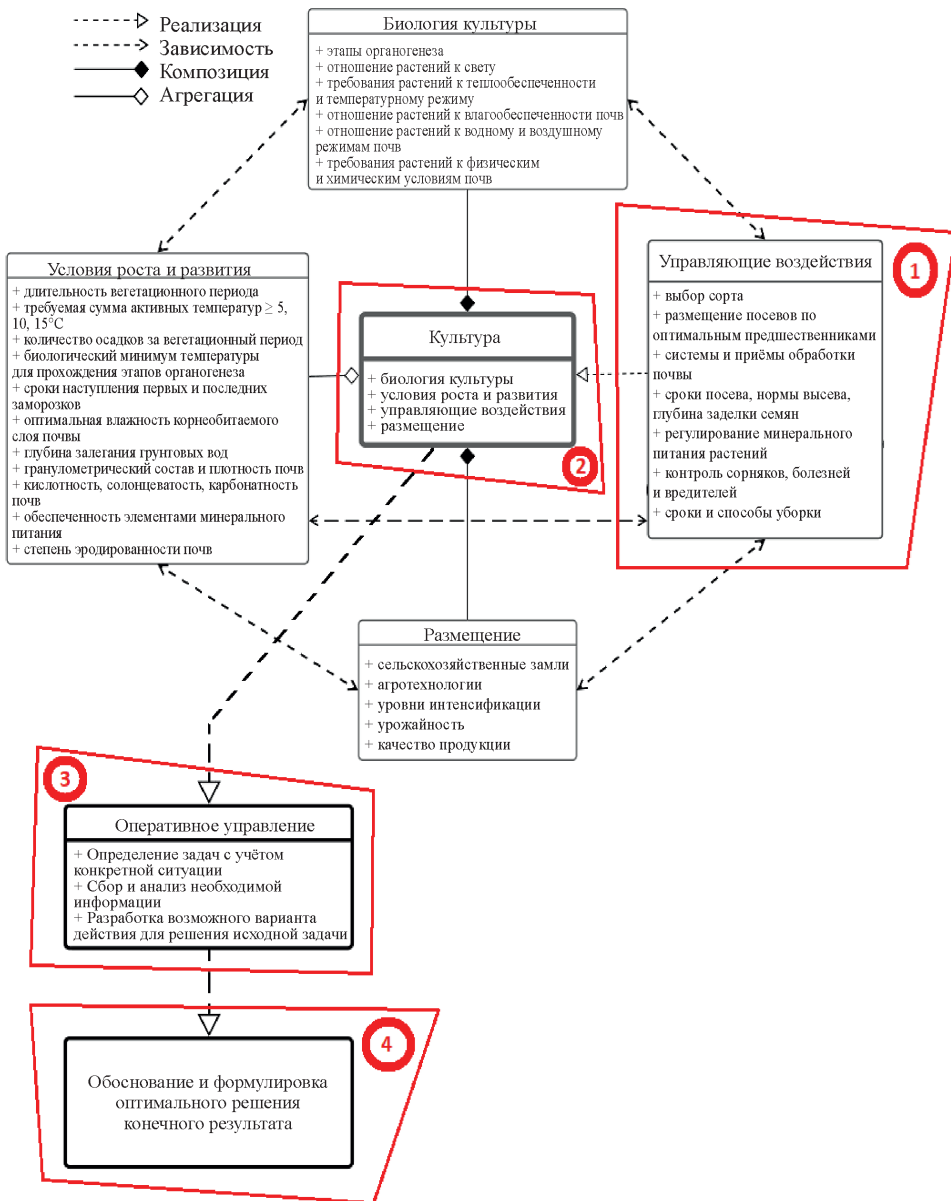


Рис 28. Схема процесса подготовки принятия решений

выявит весь набор стратегий, но алгоритм их оценивания, назначение рейтинга должны быть разработаны совместно с экспертом. Не исключено, что рейтинг стратегий будет постоянно обновляться в зависимости от различных случайных факторов, а также от появления новых знаний о применении этих стратегий на других участках и регионах.

Представленные два метода класса «Культура», реализующие структурное управление выбором культуры, заключаются не в изменении моделей, рассчитывающих числа в управляющих матрицах или цепочках управляющего воздействия по каждой культуре (хотя это и важно, но делается один раз и далее остаётся неизменным), а в разработке критериев присутствия в таблице элементов ПО. За счёт добавления или удаления элементов для каждой культуры будут меняться итоговые показатели. Следует учитывать, что исключение из анализа малосущественного фактора для одной культуры может сильно исказить итоговую картину для другой. В связи с этим необходимо сделать вывод, что факторы, элементы ПО, должны присутствовать в управляющей матрице или цепочке управляющего воздействия культуры все без исключения. Для современных вычислительных средств это не проблема. Концептуализация ПО или даже очерченной по некоторым критериям её части – задача бесконечная, с ростом объёма структурированной и детализированной информации риск не учесть какие-то данные стремится к нулю. Однако всегда остаётся риск при всём множестве исходных данных решать ненужную задачу. Именно поэтому следует рассмотреть и систематизировать все возможные методы класса «Культура», как это было продемонстрировано для ПО «агроэкологическая оценка земель».

Представленные методы – конкретные функциональные блоки системы поддержки принятия решений (СППР). На выдаваемые такими блоками рекомендации с некоторой долей доверия опирается ЛПП, осуществляя выбор и размещение конкретной культуры на данной территории.

3.4. Математические основания системы знаний

Существенная роль математики в разработке ИИС отражена в позиции отечественных учёных. «Одним из основных видов математической деятельности в современном мире является построение и реализация математических алгоритмов (в широком смысле этого слова) в формализованных системах – языках программирования, языках проектирования электронных устройств, другими словами – в проектировании элементов информационных технологий» [Семёнов, 2013].

Функционирование какой бы то ни было ИИС происходит благодаря выполнению набора запрограммированных операций согласно заданному набору логических схем. Исследователи выделяют 8 базовых направлений логики [Карпенко, 2003]:

- 1) логики теории моделей;
- 2) логики теории множеств;
- 3) логики теории рекурсии;
- 4) логики теории доказательств;
- 5) интуиционистские и суперинтуиционистские логики;
- 6) модальные и временные логики;
- 7) многозначные и нечёткие логики;
- 8) релевантные и паранепротиворечивые логики.

Любые другие логики – это некоторые конструкции, которые в каком-то отношении сходны с эмпирическими или теоретическими описаниями указанных базовых направлений (возможно, сходны только в отношении некоторых значков) и чаще всего строятся путём их «модернизации» [Ивлев, 2018]. Прочие математические приёмы используются в рамках отдельных программируемых операций обработки данных и сводятся к различным прикладным дисциплинам двух базовых направлений: математической кибернетики и теории вероятностей.

Любые логики состоят из символов и операций над ними. Символы и операции – всегда элементы некоторого языка. Применение логик в ИИС сводится к соотношению символов и операций этого языка с абстрактными объектами ПО и их взаимодействиями. В рамках посвящённого Дню логики [Азуле, 2020] круглого стола «Математическая логика: современное состояние и перспективы», прошедшего

в Новосибирском Академгородке, о роли логики в ИИС было сказано [Баженов, 2020]: «Одним из ключевых методов информатики является постановка и решение задач через формальные, компьютерные языки. Для интересующей предметной области выбирается формальный язык, в котором можно описать класс задач, подлежащий решению. После этапа формализации возможность решения класса задач изучается на уровне формального языка уже с помощью методов математики и компьютерных наук – фактически исследуются свойства языка как математические объекты».

Стандартный математический объект для исследования свойств языка – алгебраическая система, которая может быть построена в рамках любой логики. Простейшие алгебраические системы, например группы движений, лежат в основе алгоритмов последовательностей изменений экранных изображений в компьютерных играх и других приложениях с высоким уровнем визуализации. Если математикам удаётся построить алгебраическую систему, полностью покрывающую некоторую систему знаний, то полученные с помощью фундаментальных математических законов утверждения можно преобразовывать обратно в знания и тем самым получать формулировки закономерностей, справедливых в соответствующей ПО. Это подтверждают многочисленные примеры получения известных ранее фундаментальных законов в различных научных дисциплинах путём логико-лингвистической обработки внесённых в ИИС данных, информации и знаний.

Представленные в разделе 2.3 (см. главу 2) таблицы шаблонов знаний (рис. 10–19) дают возможность построить алгебраическую систему, покрывающую систему знаний о процессе возделывания сельскохозяйственной культуры. Достаточный объём собранных в эти таблицы знаний ставит вопрос о корректном к ним обращении функциональных блоков ИИС для дальнейшей машинной обработки. Корректное обращение к знаниям возможно с помощью специального языка поисковых запросов (ЯПЗ), задача о разработке которого была поставлена в Лаборатории в 2019 г. Описание ЯПЗ (пп. I–III ниже) базируется на шаблонах знаний (раздел 2.3), терминологии языка UML (раздел 3.1) и общих лингвистических правил построения языков [Мозговой, 2006].

I. Ядро ЯПЗ включает 7 базовых единиц: конфигурация объектов; объект; последовательность; связь; атрибут; метод; класс.

II. Цель использования ЯПЗ – стандартизация выявления скрытых закономерностей в условиях сельскохозяйственной задачи. При этом схема выявления скрытых закономерностей состоит из пяти шагов:

1) условия задачи представляются как комплект фрагментов собранных знаний;

2) комплекты фрагментов собранных знаний преобразуются в базовые единицы языка ЯПЗ;

3) в базовых единицах ЯПЗ формулируется стандартный набор возможных закономерностей (3 сценария их получения расшифрованы в разделе 2.5);

4) каждая из закономерностей преобразуется из базовых единиц ЯПЗ в комплект фрагментов знаний;

5) комплект фрагментов знаний интерпретируется в терминах условий задачи.

В результате знания, получаемые на шаге 5, являются дополненным, более качественным комплектом знаний, чем комплект, имеющийся на шаге 1.

III. Любой комплект условий сельскохозяйственной задачи имеет представление в виде связанной структуры фрагментов собранных знаний – таким образом, пункты II.1 и II.5 к правилам языка ИПЗ не относятся. В соответствии с пунктами II.2–II.4 сформирована таблица служебных команд на ЯПЗ с расшифровкой их значений и возможностей применения, используемые в схемах 1–3 из раздела 2.5 (табл. 11).

Развитие ЯПЗ заключается в его дальнейшей формализации до уровня регулярного языка [Гладкий, 1973] и разработки формальной грамматики его использования. С помощью таких грамматических правил сформулированные схемы и сценарии выявления скрытых закономерностей в массивах данных сельскохозяйственных задач примут вид фрагментов формализованного описания некоторого конечного детерминированного автомата [Хопкрофт и др., 2002] – стандартного математического объекта, который легко запрограммировать и тем самым автоматизировать соответствующий логический элемент сельскохозяйственной ИИС.

Служебные команды на ЯПЗ

Обозначение	Расшифровка	Применение
$o(c_1, \dots, c_k)$	Найти объект, входящий в каждую конфигурацию объектов из заданного набора	Схема 2
$m(o)$	Найти все объекты, попадающие в окрестность данного по заданной мере	Схема 2
$s(q_0)$	Найти последовательности, соответствующие заданному качеству объекта	Схема 3
$o(q_0, k)$	Найти объект заданного качества из заданного класса	Схема 3
$s(m_1, \dots, m_k)$	Найти последовательности, содержащие методы из заданного набора	Схема 3
$c_0(c)$	Найти улучшенный вариант данной конфигурации объектов	Схема 1
$c(q_{\min}, o)$	Найти конфигурации объектов с минимальным уровнем заданного качества для заданного объекта	Схема 1

3.5. К созданию сельскохозяйственной ИИС

Функции сельскохозяйственной ИИС нужны стейкхолдерам для решения тех задач, с которыми текущая конфигурация метазнаний доступных профильных экспертов не справляется. Эти функции формулируются при заполнении соответствующего шаблона знаний (раздел 2.3, рис. 12, см. главу 2) и учитываются при составлении диаграмм UML (разделы 3.2–3.4 этой главы). Иерархия функций сельскохозяйственной ИИС должна восходить к служебным командам ЯПЗ (см. табл. 11), список которых (включая схемы реализации) может быть при необходимости расширен. Роль функций сельскохозяйственной ИИС должна сводиться к уточнению деталей формулировок служебных команд ЯПЗ в каждом конкретном случае решения практической

задачи на реальном сельскохозяйственном объекте: заданные наборы конфигураций объектов и методов, заданная мера различия между объектами, заданное качество объекта, мера качества конфигурации объектов – все эти формулировки должны обрести конкретный математический смысл, реализованный в программном коде.

Использование функций сельскохозяйственной ИИС возможно только тогда, когда собраны (необязательно полностью) данные, информация и знания, связанные с решаемой задачей, поскольку именно к ним происходит обращение на ЯПЗ. В рамках второй стадии цифровой зрелости (раздел 2.1, см. главу 2) необходимо объединить в единую систему все связанные с актуальными для стейкхолдеров сельскохозяйственными задачами данные, информацию и знания, на основании которых будут выполняться соответствующие функции сельскохозяйственной ИИС. Такая единая система должна предусматривать подключение разноформатных источников данных, информации и знаний, распределение потоков которых по элементам системы должно быть строго регламентировано (и реализовано на третьей стадии цифровой зрелости).

Задача проектирования системы с указанными свойствами и назначением была поставлена в Лаборатории в 2019 г. В главе 4 настоящей монографии детали проекта этой системы раскрыты в трёх связанных между собой темах:

1. Аграрная интеллектуальная система (раздел 4.1) – общая логика системы.
2. Хранилище данных (раздел 4.2) – логика размещения и перераспределения данных, информации и знаний в сельскохозяйственной ИИС.
3. Задачи организации ХД и использование накопленных агрономических знаний (раздел 4.3) – логика организации и использования знаний, в том числе создание аксиоматики в агрономии.
4. Репозиторий знаний (раздел 4.4) – логика организации знаний ПО решаемых задач посредством сельскохозяйственной ИИС.

Разработка указанных логических схем неявно содержит целый комплекс научных и практических задач, возникающих в цифровых технологиях независимо от сферы применения.

Например, методы UML нужны для устранения пробелов в данных и знаниях. Анализ научной и методической литературы по дан-

ной теме показал, что существуют только общие рекомендации для составления дидактических и корректирующих методов, но сфера разработки новых, более оптимальных и эффективных методов (в том числе с использованием адаптивного подхода) для определения и исправления пробелов в знаниях недостаточно исследована. В связи с этим актуальной задачей является разработка эффективной методики для определения и исправления пробелов в знаниях, на основании которых сельскохозяйственная ИИС с функцией поддержки принятия решений может давать прогнозы и рекомендации принимающим решения лицам, занятым непосредственно в сельскохозяйственном производстве.

Не менее важная задача – формирование комплектов знаний и прослеживание их взаимосвязи с практическими задачами и возникающими при их реализации моментами необходимости поддержки принятия решений.

ГЛАВА 4

ЭЛЕМЕНТЫ СЕЛЬСКОХОЗЯЙСТВЕННОЙ ИНТЕЛЛЕКТУАЛЬНОЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ

В последние годы в научной литературе обсуждается парадигма новой промышленной революции Индустрия 4.0 [Гуськова, Неретина, 2017; Зозуля, 2018; Никонова, 2019]. В рамках этой парадигмы развивается автоматизация производственных технологий, включающая киберфизические системы [Radanliev et al., 2019; Napoleone et al., 2020], интернет вещей [Wortmann, Flüchter, 2015], искусственный интеллект [Hwang, 2016; Lee et al., 2018; Özdemir, Hekim, 2018], когнитивные вычисления и большие данные [Wang L., Wang G, 2016]. Интеллектуальное сельское хозяйство также использует новые технологии, к которым относятся различные датчики и устройства, роботы, GPS-ГЛОНАСС, информационные технологии, которые позволяют сельскохозяйственным предприятиям быть более прибыльными и экологически безопасными [Trivelli et al., 2019; Варганова, 2019; Огнивцев, 2019]. Прогнозируется, что искусственный интеллект к 2026 г. в четвертой технологической волне станет одним из основных инструментов для принятия решений [Лясников, 2018].

4.1. Аграрная интеллектуальная система

Аграрная интеллектуальная система (АИС), реализуемая Лабораторией в рамках парадигмы Индустрия 4.0, – экспертная подсистема сельскохозяйственной ИИС, включающая задачи, данные, информацию, знания, модели и результирующие показатели, предназначенные для лиц, принимающих решения (ЛПР) в сельском хозяйстве [Каличкин и др., 2019]. АИС разрабатывалась для поддержки принятия решений в управлении процессом возделывания сельскохозяйственной культуры, но может быть применена и к другим ПО сельского хозяйства.

Данные, информация и знания должны накапливаться по группам в рамках иерархии АИС. Все цепочки преобразования теоретических знаний в практические результаты должны быть прозрачны.

Для АИС, выступающей основой для самообучения сельскохозяйственной ИИС, на иллюстрации продемонстрирована архитектура, призванная учитывать все этапы предварительного анализа: от формулировки сельскохозяйственной задачи до выдачи аналитической справки, на основе которой не имеющий специального образования человек может принимать решение, руководствуясь только здравым смыслом. Компоненты архитектуры изображены как элементы блоков и связаны между собой концептуальной моделью системы в виде схемы для разработки конкретных требований. Такое представление материала применяется для отображения информационной архитектуры любой разрабатываемой системы (рис. 29). Нотация обозначений приведена на иллюстрации, цифрами отмечены ключевые элементы, которые рассмотрены ниже.

Пространство сельскохозяйственных задач (1). Нижняя плоскость на иллюстрации с коричневым кружком – задачей (7) внутри. Сетка в плоскости демонстрирует наличие в этом пространстве некоторой топологии. Для разработки топологии требуется концептуальное и онтологическое моделирование ПО. Результаты такого моделирования могут быть представлены по-разному, поэтому необходимо наращивание объёма концептуализированных и онтологизированных знаний по определённому алгоритму, в конкретном направлении. Сельскохозяйственные задачи могут формулироваться в различных направлениях, например агроэкологическая оценка земель, выбор и размещение культур на определенной территории, выбор и сопровождение агротехнологий и др. В разделе 2.2 настоящей монографии знаниям о сельскохозяйственных задачах соответствует таблица шаблонов знаний «Агротемы» (см. рис. 12).

Пространство источников данных (2). Каждый конкретный источник обозначен прямоугольником фиолетового цвета с закруглёнными углами. Речь идёт об источниках нетривиальных данных, которые где-то и кем-то собраны или собираются на регулярной основе, но нет ясности, как именно их оцифровать и перевести в понятный для интерпретации вид. Полноценное применение данных, информации и знаний, регулярно получаемых из источников, реализуется на третьем этапе цифровой зрелости ИИС.

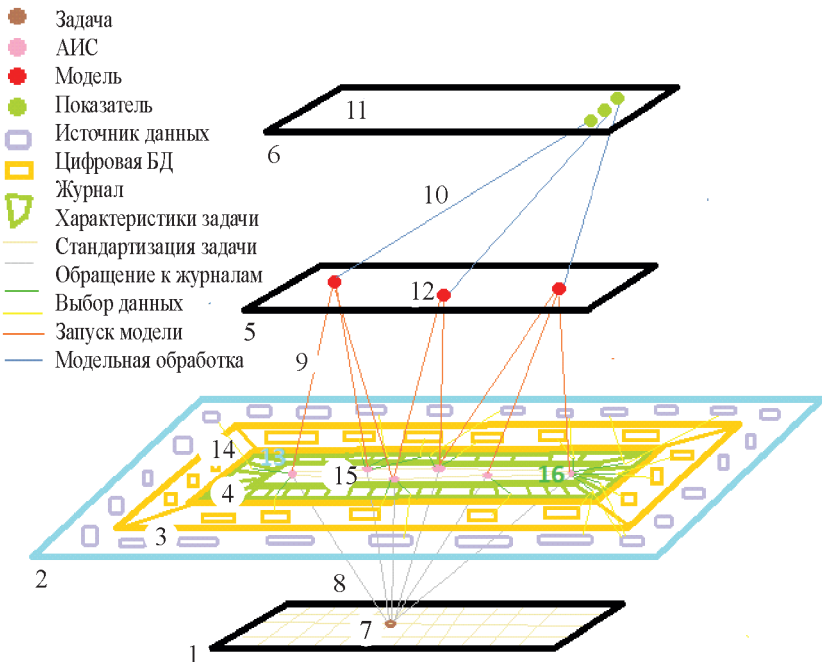


Рис. 29. Архитектура аграрной интеллектуальной системы

Хранилище данных (3), или пространство баз оцифрованных данных. Это данные, которые находятся в хранилище данных (ХД) и прикреплены к АИС в виде, в котором их можно использовать для различных математических, статистических и других расчётов. Не представляется рациональным брать данные из источников (2) и переводить их в понятный формат ради того, чтобы они просто были в ХД. Данные и так есть в источниках (2), а забирать и обрабатывать их оттуда (автоматизация этого процесса для каждого источника – не всегда простая задача) имеет смысл по запросу, основанному на условиях задачи (1, 7). Если для решения задачи понадобились данные из внешних источников (2), то эти и только эти данные можно перевести в оцифрованный формат и поместить в ХД. Выглядеть это будет в виде присоединения некоторого массива данных к уже существующей в ХД (возможно, пустой) базе данных. При этом во внешнем источнике продолжают накапливаться данные, поэтому он тоже не исчезает, а

остаётся в контуре (2). Структура и свойства ХД детализируются в разделах 4.2 и 4.3 настоящей монографии.

Журналы (4) – способ хранения информации об источниках внешних данных (3) и базах оцифрованных данных (4). Журналы разбиты по темам, соответствующим структуре пространства сельскохозяйственных задач (1). При этом в каждый журнал не только может попасть (14) несколько источников данных (2) или баз данных (3), но и каждый источник данных (2) или база данных (3) может попасть (14) в несколько журналов. Через журналы с помощью ЯПЗ (разделы 3.4, 3.5) АИС взаимодействует с ХД. Журналы являются способом организации данных, информации и знаний в ХД в соответствии со структурой ЯПЗ и другими метакатегориями ПО, встраиваемых в сельскохозяйственную ИИС в виде семантической сети (раздел 4.5).

Пространство моделей (5). Вторая сверху плоскость с тремя красными кружками – моделями (12) – внутри. Модели – это средства обработки данных, дающие результаты, по которым можно судить о связанной с задачей ситуации и принимать решения. Без моделей не обойтись, ведь даже подсчёт среднего арифметического некоторого ряда исходных данных уже является моделью, не говоря уже о сравнении рядов статистическими и математическими методами. Моделей придумано много, и каждая из них может сработать в определённый момент в определённых условиях, а может и не сработать. Модели, разработанные отдельными лабораториями и учёными, могли бы быть сервисными элементами АИС, выступая в качестве чёрных ящиков, в которые отправляются специально структурированные входные данные (структурирование входных данных (9) – отдельная и трудоёмкая работа, выполняемая прежде всего в интересах авторов моделей), из которых получают показатели (11), имеющие ту или иную ценность. Множество используемых моделей как способов получения решений – неотъемлемая часть любой ИИС (см. раздел 2.1). Первоначальный набор моделей формируется через таблицу шаблонов знаний «Агротемы» (см. рис. 12). Формализация знаний из таблиц «Классы» и «Связи» (см. рис. 10, 11) с помощью языка UML позволяет составить таблицы методов (пример такой таблицы – табл. 11 в разделе 3.2), каждая ячейка в которых задаёт множество моделей, объединённых комплектами входных и выходных данных.

Пространство документов (6). Верхняя плоскость с тремя кружками – показателями (11) – внутри. В результате обработки соответствующих задаче (1, 7) данных (15), собранных из ХД (3) и из внешних источников (2) моделями (5, 12), получены результирующие показатели (11), по значениям которых в автоматическом режиме составляются документы: справки, отчёты, аналитические записки, статистические отчёты, рекомендации и др. При выборе моделей в пространстве (5), отборе (14) источников данных (2) и оцифрованных баз данных (3) через систему журналов (4,13) необходимо чётко понимать условия конкретной задачи и цель использования АИС. Утилитарная задача – это всегда документ с цифрами и их краткой, ориентирующей в пространстве выбора интерпретацией. Следовательно, формат таких документов должен быть известен заранее. Пространство документов – это полный комплект документации поддержки принятия решений. Первоначальный комплект документации формируется через таблицы шаблонов знаний «Агротемы» (см. рис. 12) и «Агропоказатели» (см. рис. 15). Детализация документов формируется через специальные карточки документов и другие компоненты ХД (раздел 4.2).

Задача (7) – элемент пространства (1). Формализация понятия «задача» для всех пользователей АИС необходима для оптимизации структуры журналов (4), правильной работы используемого контура АИС (15) и её отдельных узлов (16), поскольку именно эти компоненты отвечают за выбор данных (13, 14), необходимых для корректного и полного формирования входа (9) моделей (5,12) и отображения в документах (6) необходимых фактических или прогнозных показателей (11), связанных с задачей.

Предполагается 8 групп стейкхолдеров – пользователей АИС. На основе таблицы шаблонов знаний «Агротемы» (рис. 12) для них составляется рейтингованный список свойств и функций ИИС, на основе которого концептуализируется пространство задач для каждой группы. Например, концептуализация пространства сельскохозяйственных задач может осуществляться для группы пользователей «Сельскохозяйственные предприятия» с выделением задач «Выбор и размещение культуры» с её подзадачами «Оценка соответствия требований культур свойствам территории», в рамках модели управления агротехнологиями может формулироваться задача «Управление

процессом возделывания культуры» с подзадачей «Пространство состояний и поля выбора». Первоначальный список сельскохозяйственных задач формируется через таблицу шаблонов знаний «Агротемы» (см. рис. 12). Детализация каждой задачи формируется через специальные карточки задач и другие компоненты ХД (раздел 4.2). Важный навык, которым должна обладать сельскохозяйственная ИИС, – формулировка условий любой задачи в терминах базовых единиц ЯПЗ (раздел 3.4).

Формализация данных пользователя (8). Серые линии, идущие от коричневого кружка – задачи (7) – к розовым кружкам – узлам АИС (16). Даже точная идентификация в пространстве (1) задачи (7) не отменяет разнородности и разноформатности получаемых от стейкхолдера – пользователя АИС – данных. Например, нет никакой гарантии, что данные, собираемые через Анкету фермера (раздел 2.4), окажутся полными и достоверными. Данные стейкхолдера необходимо формализовать и оцифровать так, чтобы при стыковке с отобранными (14) через систему журналов (4, 13) данными внешних источников (2) и оцифрованных баз данных (3) в узлах (16) и контурах (15) АИС корректно формировались фрагменты входных массивов данных (9) для моделей (5, 12).

Этот этап обработки данных пользователя АИС тесно связан с задачей (7), в связи с этим осуществляется этап «докомплектация данных». Необходимо разработать сценарии работы с данными на этом этапе, которые реализуются в процессе разработки тестовой модели движения информации в АИС (см. раздел 4.3). С другой стороны, к данному вопросу осуществляется подход в рамках задачи расчёта ресурсных затрат на отдельные операции работы ИИС и её операторов (см. раздел 2.1). Формализация данных пользователя – полное встраивание получаемых от стейкхолдеров данных, информации и знаний в компоненты сельскохозяйственной ИИС. Итоговые комплекты данных, информации и знаний пользователя должны соответствовать структурам знаний, собранным в таблицы шаблонов знаний «Классы» (см. рис. 10), «Связи» (см. рис. 11), а также в случае использования внешних источников данных – в таблицы «Источники данных» (см. рис. 13) и «Последовательности» (см. рис. 14).

Формирование массива входных данных для применения модели (9). Оранжевые тонкие линии, идущие от розовых кружков – узлов

АИС (16) – к красным кружкам – моделям (12). В основном на вход моделям поставляются экспертные или справочные данные, часто неполные или устаревшие, т.е. фрагментарно связанные с источниками данных (2). Медленное получение необходимой и полной входной информации – основной недостаток каждой конкретной модели (12) и одновременно интерес группы стейкхолдеров «модельные сервисы», о которых кратко сказано в (5). Пространство массивов входных данных зависит от используемых АИС моделей (5), и его концептуализация должна проводиться с оглядкой на группы пользователей системы и группы связанных с ними сельскохозяйственных задач, о чём шла речь в (7). Комплекты входных данных для применения модели формируются в специальном реестре ХД (раздел 4.2) и являются фрагментами структур данных, построенных на основании заполненных таблиц шаблонов знаний «Классы» (см. рис. 10) и «Связи» (см. рис. 11).

Выходные данные моделей (10). Тёмно-голубые тонкие линии, идущие от красных кружков – моделей (12) к светло-зелёным кружкам – показателям (11). Существующие модели (12) из пространства моделей (5) выдают некоторые наборы данных, но не все из них нужны в пространстве документов (6) в качестве показателей для принятия решений. В связи с этим из выходных данных моделей можно сформировать минимально необходимый комплект показателей для того, чтобы иметь возможность достаточно полно судить, например, обо всех стадиях процесса возделывания сельскохозяйственной культуры, на пятой стадии цифровой зрелости сельскохозяйственной ИИС осуществлять имитационное моделирование этого процесса. Выходные данные моделей должны соответствовать ячейкам веерной матрицы, заложенной в основу таблицы шаблонов знаний «Источники данных» (см. рис. 13) и соответствовать структурам данных, построенным на основании заполненных таблиц шаблонов знаний «Классы» (см. рис. 10) и «Связи» (см. рис. 11).

Показатели (11). Светло-зелёные кружки на плоскости документов (6). Формируются обработкой выходных данных моделей (10) и размещением их в заданных местах документа из пространства (6). Первоначально множество показателей формируется через заполненную таблицу шаблонов знаний «Агропоказатели» (см. рис. 15). Deta-

лизация каждого показателя формируется через специальный перечень агропоказателей и другие компоненты ХД (раздел 4.2).

Модели (12). Красные кружки на плоскости моделей (5). Отдельные элементы пространства (5). Нужны для преобразования фрагментов данных, формируемых (14) в контурах (15) и узлах (16) АИС на основании взаимодействия через журналы (4, 13) с источниками данных (2) и базами оцифрованных данных (3), в демонстрируемые в агроконсалтинговых документах (6) показатели (11). Исследование и встраивание этих моделей в систему должно проводиться на основании характеризующих конкретную задачу (7) из пространства (1) показателей (11). Детализация моделей формируется через специальные карточки моделей и другие компоненты ХД (раздел 4.2).

Обращение к журналу (13). Изображено на иллюстрации в виде тонких зелёных линий. Происходит в момент, когда нужно совместить, состыковать данные пользователя (8) с данными входа моделей (9) с учётом неизвестных пользователю, но доступных АИС данных внешних источников (2) и баз оцифрованных данных (3). Таким образом, через обращение к журналу происходит совмещение одновременно четырёх ключевых компонентов системы: (2), (3), (7) и (12), вследствие этого работа над структурой журналов приобретает стратегическое значение. Состав и структура журналов (4), их связи с источниками данных (2) и базами оцифрованных данных (3) могут существенно облегчить взаимодействие АИС с ХД при правильной их реализации или серьёзно её осложнить при неправильной. В связи с этим основное назначение обращения к журналу – его преобразование в ЯПЗ (разделы 3.4, 3.5), на котором происходит указанное взаимодействие.

Отбор данных (14). Изображён на иллюстрации в виде тонких жёлтых линий. Осуществляется через систему журналов (4) путём обращения к одному или нескольким журналам (13) для выделения из источников данных (2) и баз оцифрованных данных (3) тех фрагментов данных, которые, с одной стороны, соответствуют задаче (7), с другой – предоставленным пользователем АИС (8), с третьей – входным данным (9) моделей (12), с четвёртой – независимо от пространства моделей (5) являются необходимым и достаточным объёмом данных для выдачи адекватных и достоверных показателей (11) в доку-

ментах пространства (6). Ключевой блок АИС и сельскохозяйственной ИИС: обращение из системы журналов в ХД формулируется на ЯПЗ (разделы 3.4, 3.5), в связи со структурой которого эффективный отбор данных возможен только в том случае, если ХД организовано на основании шаблонов знаний (раздел 2.2).

Активный контур АИС (15). Идея активного контура – отобразить подмножество пространства задач (1) во множестве отобранных (14) через журналы (4) данных, т.е. каждой задаче (7) представителя одной из 8 групп стейкхолдеров – пользователей сельскохозяйственной ИИС – должен соответствовать свой активный узел (16), который отбирает (14) очерченный областью пространства задач (1) набор связанных с задачей (7) данных. Каждый активный контур моделирует в АИС так называемого агента – наблюдателя за одними и теми же данными при решении одних и тех же задач в течение длительных периодов времени. Предложена идея использовать концепцию понятия «агент», известную в агентном моделировании. Многие современные зарубежные системы предиктивных технологий в области сельского хозяйства основаны именно на таком агентном подходе [Le et al., 2012; Groeneveld et al., 2017]: отбираются (экспертно или статистически) данные для наблюдения, моделируется многолетнее наблюдение за этими данными, многолетние временные ряды наблюдаемых данных обрабатываются математически, результат обработки сравнивается с реальными данными. Недостаток подхода использования одного агента в том, что используется экспертное мнение только одного наблюдателя. В связи с этим была выдвинута идея разделять агентов на биоценозные, или внутренние (зерно, растение), и внешние (агроном, биотехнолог, фермер). В результате возникает научная задача коцептуализации пространства агентов. Первоначальные знания об агентах извлекаются из таблицы шаблонов знаний «Применимость объектов» (см. рис. 18) в виде комплектов взаимосвязанных через общие объекты конфигураций объектов.

Узлы АИС (16). Обозначены розовыми кружками в пространстве, расположенном с внутренней стороны блока журналов (4). Составная часть активного контура (15). Моделирует источники наблюдения агента. Агент наблюдает в течение длительного времени за несколькими источниками данных. Эти временные ряды формируют

его «интуицию», на основе которой он принимает решение. Каждый узел может входить как составная часть в несколько активных контуров (15), так же, как и каждый активный контур (15) может содержать несколько узлов. Терминами «узнаваемость» и «различимость» понятие «интуиция» формализовано в парадигме суперинтуиционистской логики, выросшей в научной школе А.Н. Колмогорова [Колмогоров, 1985] и разработанной основоположником сибирской школы алгебры и логики А.И. Мальцевым [Мальцев, 1970] и его учениками [Мардаев, 1987; Лавров, Максимова, 2002; Максимова, Юн, 2019]. Необходимо отметить, что структура активных контуров (15) и узлов (16) независима от структуры журналов (4), т.е. каждый узел может обращаться (13) к нескольким журналам (4) и несколько узлов могут обращаться (14) к одному журналу (4), что показано на иллюстрации зелёными тонкими линиями. На данный момент понятие агента как активного контура АИС (15) не формализовано: эта задача отнесена к следующему этапу исследования, следовательно, и узлы АИС как составные части контуров отдельно не рассматриваются. Независимость от структуры журналов (4) позволит в будущем смоделировать агентов как множества составленных в активные контуры (15) узлов и встроить в работу АИС, используя при необходимости и другие направления логических исследований (раздел 3.4).

Представленная архитектура АИС покрывает все возможные сценарии её применения, однако эффективность имплементации в значительной степени зависит от полноты и ценности экспертных метазнаний, собираемых через шаблоны знаний (раздел 2.2) и закладываемых в основу топологии онтологической модели языка ИПЗ (разделы 3.4–3.5). Решающую роль играют экспертные знания, их представление в концептуализированном виде, в основе которого лежит идея анализа и проектирования конструкции моделей не только ПО, но и самих баз знаний в виде системы инвариантных конструкторов – моделей и задания метода их интеграции в некоторую единую понятийную структуру [Болотова и др., 2011]. Такой структурой является представляемая в настоящей монографии концепция сельскохозяйственной ИИС, разработанная в Лаборатории и используемая в настоящее время для формализации любого вида данных, информации и знаний, связанных с сельским хозяйством.

4.2. Структура и организация Хранилища Данных

Первый шаг реализации сельскохозяйственной ИИС – её концептуальное описание как единой ПО, включая её статические и динамические свойства. Статические свойства – архитектура АИС и её 16 основных компонентов – представлены в разделе 4.1. Подход к описанию динамических свойств, характеризующих движение в сельскохозяйственной ИИС данных, информации и знаний, представлен в настоящем разделе.

Концептуальное описание динамических свойств сельскохозяйственной ИИС должно давать представление о том, какая именно информация, из каких именно источников, в каких именно форматах и по каким именно траекториям проходит через статические компоненты АИС, прежде чем предстать перед ЛППР в виде прогнозов и рекомендаций – продуктов экспертной ИИС с функцией поддержки принятия решений – в рамках конкретной сельскохозяйственной задачи.

Дальнейшие шаги реализации сельскохозяйственной ИИС предполагают создание структурированного ХД на основе статических и динамических свойств; наполнение ХД данными; автоматизацию применения математических, логических, семантических методов ИИ для обработки собранных данных; автоматизацию использования моделей сельскохозяйственных процессов; разработку механизмов верификации используемых методов и моделей для оценки качества выдаваемых прогнозов и рекомендаций.

По мере наполнения ХД и выявления скрытых взаимосвязей между статическими и динамическими свойствами ИИС базы данных в ХД превращаются в базы знаний. Благодаря использованию баз знаний система «учится» работать с неполными данными, выдавая приемлемые результаты. На основе собранных данных, информации и знаний должна быть сформирована «интуиция» системы, определяемая как возможность находить приемлемые решения из массивов их несвязанных и неполных комплектов.

В разделе 4.1 вместо понятия ХД акцент делается на отдельном компоненте АИС – оцифрованных базах данных, тем самым игнорируются способы, методы и инструменты, с помощью которых данные, полученные из внешних источников или условий задачи, приобретают таблично-цифровой вид, пригодный для обработки программами

ЭВМ. Здесь этот пробел устраняется: в структуре ХД выявляется 22 компонента, только одним из которых является множество оцифрованных баз данных. Через взаимный обмен данными между этими компонентами выстраиваются процессы обработки данных в сельскохозяйственной ИИС. Важно понимать, что 22 компонента ХД суть структурированная информация о совокупности реальных сельскохозяйственных объектах, а также об их отношениях, связях и характеристиках.

Отметим, что среди 22 представленных компонентов ХД встречаются как базы данных, так и базы знаний, содержащие структурированную информацию о самом ХД, уточняемую и дополняемую в процессе работы ИИС. По мере эксплуатации «опыт» системы наращивается, соотношение содержащейся в ней информации эволюционирует в пользу знаний.

ХД состоит из 22 компонентов, посредством которых осуществляется хранение, упорядочивание, структурирование, движение данных в АИС. В рамках этой взаимосвязи ХД и АИС, компоненты ХД делятся на три группы: карточки, перечни и реестры (табл. 12).

Карточка – структурированное описание объекта с выделением его значимых свойств и характеристик, отражаемых в специально выделенных полях. Поля карточки имеют стандартизированные форматы для автоматизации обмена данными. В структуре ХД выделено четыре карточки: Карточка модели <4>, Карточка агросервиса <1>, Карточка документа <2> и Карточка задачи <3>. Объекты, описываемые этими компонентами ХД, являются структурными единицами четырёх пространств АИС, на которых строится её работа в рамках стандартной парадигмы экспертных систем: проблема (постановка задачи) <3> – подготовка данных (сбор необходимой информации для решения) <1> – решение (модели и методы) <4> – результат (документы для ЛПР) <2> (указанная траектория видна слева на рис. 1 ниже). Например, компонент ХД Карточка модели <4> является структурированным описанием компонента АИС Модель (12) и предназначена для хранения информации о каждой модели, отобранной для использования в АИС. Содержит 17 полей со свойствами и характеристиками модели, а также её связями с другими компонентами ХД.

Состав компонентов ХД и АИС

Компоненты ХД			Компоненты АИС
карточки	перечни	реестры	
<1> Карточка агросервиса	<5> Перечень агропоказателей	<9> Реестр агросервисов	(1) Пространство задач
<2> Карточка документа	<6> Перечень геопа- раметров	<10> Реестр внешних источников	(2) Внешний источник данных
<3> Карточка задачи	<7> Перечень мето- дов обработки гео- данных	<11> Реестр входных форматов мо- делей	(3) Оцифрованные базы данных
<4> Карточка модели	<8> Перечень функ- ций ГИС-систем	<12> Реестр выходных форматов агросервисов	(4) Журнал
		<13> Реестр документов	(5) Пространство моделей
		<14> Реестр задач	(6) Пространство документов
		<15> Реестр мер для сравнения тер- риторий по гео данным	(7) Задача
		<16> Реестр моделей	(8) Формализация данных задачи
		<17> Реестр отношений модель – документ	(9) Выходные данные модели
		<18> Реестр оцифрованных баз данных	(10) Выходные данные модели
		<19> Реестр регламентов импорта данных внешних источников	(11) Агропоказатели
		<20> Реестр регламентов формали- зации данных задач	(12) Модель
		<21> Реестр форматов данных внешних источников	(13) Обращение к журналу
		<22> Реестр шаблонов формализа- ции данных задачи	(14) Отбор данных
			(15) Активный контур
			(16) Узел

Перечень – структурированное представление данных об информационных единицах, участвующих в работе АИС. Как правило, представляет собой обычную таблицу объект – свойство. В качестве свойств используются как индивидуальные характеристики, так и связи с другими компонентами ХД. Например, Перечень агропоказателей <5> содержит список рассчитываемых в АИС для различных сельскохозяйственных процессов ключевых показателей (11), на которые ориентируются ЛПР. В свойствах каждого показателя указывается, при помощи каких моделей (5) <16> и агросервисов <9> они могут быть рассчитаны, в каких документах (6) <13> присутствуют, какие данные необходимы для расчёта агропоказателя (10) <17>.

Реестр – структурированное представление группы объектов или характеристик объектов с выделением значимых свойств этой группы и отражением связей с другими группами. В реестре в отличие от перечня содержатся объекты, каждый из которых может представлять собой некоторую информационную структуру, содержащие разные массивы данных, не все из которых обязательно имеют табличную форму. Например, компонент ХД Реестр моделей <16> представляет собой набор карточек моделей <4> для различных используемых в АИС моделей (12). По мере накопления количества моделей в реестре возникает возможность их классификации, уточнения структуры пространства моделей (5). Всего в ХД выделено 14 реестров. На основании реестров происходит движение информации по маршрутам. Одно из предназначений реестра – быстро находить близкие по свойствам объекты. Например, для данной модели можно выделить набор похожих моделей, используя близость характеристик, внесённых в соответствующие поля их карточек <4>.

Важная роль среди компонентов ХД реестра оцифрованных баз данных <18> показана выше. Каждый элемент этого реестра – отдельная база данных, содержащая в индивидуальном формате таблично-цифровые данные, полностью готовые к обработке математическими функциями (алгоритмами), в том числе реализованными в готовых программах ЭВМ. Важным разделом реестра <18> является множество баз данных ГИС. Особенность баз данных ГИС в том, что они являются массивами пространственных данных, привязанных к конкретным территориям по некоторому набору геопараметров –

Для каждого маршрута движения данных должно быть создано описание, однозначно отражающее, из каких именно компонентов ХД происходит отбор информации, какие характеристики и свойства содержащихся в ХД объектов необходимо извлечь, какие методы и при каких условиях будут использованы, какова точная траектория движения данных по элементам АИС и ХД, каковы изменяющиеся во времени форматы перемещающихся по нему данных.

Свойства маршрутов и их компонентов неразрывно связаны с четырьмя компонентами АИС: пространствами задач (1), моделей (5) и документов (6), а также множеством журналов (4). Вопрос о первичности этих онтологических единиц по отношению к маршрутам остаётся открытым. Описание маршрутов существенно обогатит и уточнит онтологии указанных компонентов, но и сами маршруты выстраиваются на базе именно этих онтологий. Скорее всего, это процесс параллельный, взаимопроникающий. Представляемый в работе подход к концептуальному описанию динамических свойств сельскохозяйственной ИИС реализован именно со стороны маршрутов.

В Лаборатории в рамках работы над автоматизацией взаимодействия АИС и ХД выявлены и описаны семь маршрутов движения информации, к ним составлены схемы. В качестве примера приведём полное описание одного из маршрутов.

Маршрут 1. Агросервисы для расчёта агропоказателей. Показатель, значимый для ЛПП и представленный в некотором документе или «фразе», реализованной функцией поддержки принятия решений, получается в результате цепочки процессов обработки формализованных данных. Порядок последовательности – обратный, т.е. в этом маршруте лучшее понимание работы АИС даёт движение не от данных, а от показателя, для получения которого нужно собрать данные, использовать модели. Акцент в этом маршруте делается на сторонних моделях неизвестной внутренней структуры – агросервисах (12) <9>. Агросервисы – это скачиваемые или работающие в режиме онлайн приложения, позволяющие ввести некоторые входные данные и получить рассчитанные моделью агропоказатели. Приоритет изучения сторонних агросервисов – посредством составления реестра <9> – должен заключаться в том, чтобы понять, какие из них наиболее подходят для предиктивных сельскохозяйственных технологий в условиях Сибири.

Последовательность компонентов АИС. Агропоказатели (11) → Выход модели (10) → Модель (12) → Вход модели (9) → Узел АИС (16) → Оцифрованные БД (3).

Компоненты ХД. Перечень агропоказателей <5> составляется экспертом. К каждому показателю должны быть привязаны те агросервисы из реестра агросервисов <9> и те модели из реестра моделей <16>, с использованием которых возможно получение этого показателя.

Реестр выходных форматов агросервисов <12> должен связывать набор агропоказателей из <5> с выходными данными агросервиса из <9>, обеспечивая полное понимание того, как и где брать, в том числе аппаратно и программно, нужную цифру и как из неё получить агропоказатель.

Агросервисы из <9> не попадают в реестр моделей <16>, так как их внутренняя структура неизвестна, хотя каждый из агросервисов присутствует как отдельная (возможно, единственная) функция обработки данных в некоторой модели (12) пространства моделей (5) АИС. Для каждого агросервиса через представляющую его модель и реестр отношений модель – документ <17> должен быть представлен набор агропоказателей из перечня <5>, при этом карточка агросервиса <1> связывается карточкой модели <4>.

Реестр форматов входных данных моделей <11> содержит описание порядка формирования входных данных каждой конкретной модели из данных, содержащихся в различных компонентах ХД. Отметим, что для моделей пространства (5) существуют реестры <11> и <16>, для агросервисов – реестры <9> и <12>. Это сделано из-за того, что входные данные моделей и агросервисов имеют одну и ту же природу и собираются из оцифрованных БД. Выходные данные агросервисов изначально находятся вне АИС, поэтому они сначала должны быть перенесены в оцифрованные БД <18>, а затем задействованы в реестре <17>. Мотивация использования агросервисов определяется разницей в трудоёмкости между автоматизацией работы с реестрами <9>, <12> и включением модели агросервиса (если она известна) в АИС через реестры <11>, <16>.

Реестр <18> необходим в конце этого имеющего обратный порядок маршрута, чтобы окончательно определить множество оцифрованных БД, из которых собираются данные для расчёта заданного множества агропоказателей.

Последовательность компонентов ХД. В табл. 13 представлены задействованные в маршруте компоненты АИС и ХД.

Таблица 13

Состав компонентов АИС и ХД маршрута 1

Компонент ХД			Компонент АИС
карточки	перечни	реестры	
<1> Карточка агросервиса	<5> Перечень агропоказателей	<9> Реестр агросервисов	(3) Оцифрованные базы данных
		<11> Реестр входных форматов моделей	(5) Пространство моделей
		<12> Реестр выходных форматов агросервисов	(9) Входные данные модели
		<16> Реестр моделей	(10) Выходные данные модели
		<17> Реестр отношений модель – документ	(11) Агропоказатели
		<18> Реестр оцифрованных баз данных	(12) Модель

На основании изложенного выше можно составить последовательность компонентов ХД, задействованных на каждом шаге маршрута: (11) → (10): <5>, <17>; (10) → (12): <16>; (12) → (9): <11>; (9) → (16): <1>, <9>, <12>; (16) → (3): <18> (рис. 31).

Графическое представление маршрута. Информацию из пунктов «Последовательность компонентов АИС», «Компоненты ХД» и «Последовательность компонентов ХД» следует использовать с учётом ограничений, накладываемых взаимосвязями между компонентами ХД (см. рис. 30). На диаграмме, представленной на рис. 31, жёлтым цветом отмечены связи между компонентами ХД, отражающие маршрут 1. Отсутствующая на рис. 1 связь <11> – <18> – упрощённое представление более сложной связи <11> – <21> – <19> – <12> – <1> – <9> – <18>, отражающей работу узла АИС (16), отвечающего за алгоритм подготовки входных данных модели с использованием агросервиса как внешнего источника. Жирными оранжевыми линиями на рис. 31 показана логика работы компонентов АИС.

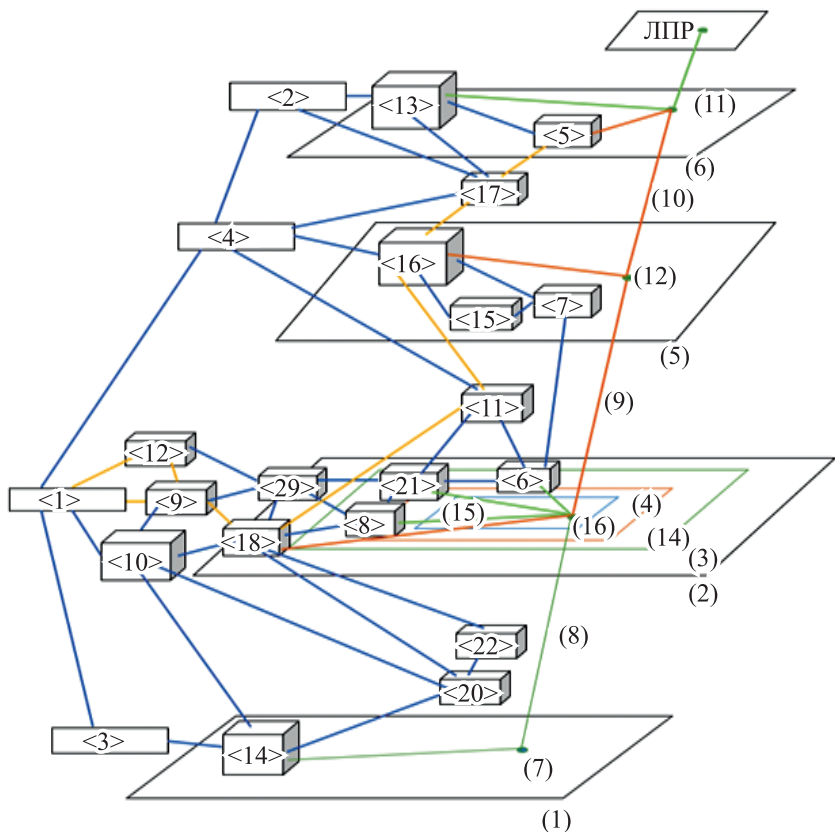


Рис. 31. Графическое представление маршрута 1

Описание движения данных маршрута 1 и сопутствующих компонентов ХД даёт представление о трудоёмкости концептуального описания сельскохозяйственной ИИС. По этому шаблону составляются более детализированные маршруты, связывающие агросервисы и агропоказатели с конкретными интернет-адресами и форматами данных, а также тестовыми комплектами БД – упрощёнными аналогами ХД. Далее эта информация входит в техническое задание для отправки на реализацию специалистам по разработке программного обеспечения.

4.3. Задачи организации ХД и использование накопленных знаний в агрономии

Существует определённый порядок взаимодействия между АИС и ХД независимо от того, какое именно аппаратное и программное обеспечение используется для их поддержки. Основным принципом взаимодействия между отдельными подсистемами экспертных ИИС и данными – задействование исключительно цифровых данных, готовых к обработке с помощью операций и алгоритмов обработки данных (раздел 2.2). Отсюда следует, что все данные для экспертной ИИС делятся на три условные группы:

- 1) цифровые данные;
- 2) нецифровые данные;
- 3) отсутствующие данные.

В связи с этим задач организации ХД существует тоже три.

1. Организация и ведение системы управления базами данных (СУБД) для управления цифровыми данными. Помимо само собой разумеющейся аппаратно-программной части основной задачей здесь является наполнение данными. Из схемы АИС (раздел 4.1) видно, что к этой задаче в первую очередь относятся следующие блоки: (3) оцифрованные базы данных; (9) входные данные модели; (10) выходные данные модели.

2. Преобразование нецифровых данных в цифровые. Данных, не готовых к обработке программными функциями, может быть три вида. В каждом случае первоочередной задачей является преобразование этих данных к текстово-табличному виду и внесение в ХД в виде оцифрованной базы данных. Это могут быть данные из внешних интернет-ресурсов – тогда требуется автоматизация их единовременного или регулярного (если данные на ресурсе обновляются) извлечения из этого источника. В последнем случае требуется автоматизированная программа, отслеживающая появление обновлений на ресурсе или обращающаяся к нему через фиксированные промежутки времени.

Извлечение информации – термин, обозначающий наличие в ХД необработанных данных, – как правило, в виде изображений различных форматов, для которых требуется провести распознавание. Изображения могут быть или результатом сканирования каких-то архивных данных на бумажных носителях, или результатом применения

средств дистанционного зондирования территорий, или снимками беспилотных летательных аппаратов. Распознавание каждого вида данных такого типа – это отдельный проект.

Третий вариант нецифровых данных – данные на бумажных носителях. Как правило, они сканируются и переходят в категорию электронных изображений, требующих обработки.

В структуре АИС все эти виды данных включены в компонент (2) внешние источники данных. О комплексе задач, связанных с организацией, управлением и преобразованием данных разных форматов, принято говорить как о технологиях Big Data [Wu X. et al., 2013; Wang, Wang, 2016].

3. При наличии весомого количества отсутствующих данных возникает невозможность добиться ясности в некотором экспертном вопросе. Это означает, что все доступные к привлечению оцифрованные данные и данные из внешних источников уже рассмотрены, и их всё равно не хватает. В этом случае ставятся задачи интеллектуальной обработки данных (Data Mining), предназначенные для заполнения пробелов в данных или для обработки данных с пробелами [Hand D.J., Adams, 2014].

Три задачи организации ХД начинают решаться задолго до возможности их постановки научно-исследовательскому коллективу (или команде разработчиков ИИС): при извлечении метазнаний эксперта, описании картины мира и составлении конфигуратора сельскохозяйственного процесса. Возможности их исследования, постановки, решения и внедрения учитываются при заполнении таблиц шаблонов знаний (раздел 2.2) и концептуализации ПО в части методов (разделы 3.2, 3.3) и последовательностей.

В разделе 1.4 мы кратко описали методологические основы создания БЗ, а также привели примеры некоторых БЗ, относящиеся в той или иной степени к сельскому хозяйству. Эти БЗ функционируют в основном в виде информационных ресурсов.

Поскольку накопление данных в сельском хозяйстве и оценка их достоверности в настоящее время представляет собой весьма трудоемкий процесс (например, получение комплексной агроэкологической информации о рабочем участке в М 1 : 1000 для достаточно обширной территории), предлагается в качестве выхода из сложившей-

ся ситуации использование аксиоматического подхода формирования и формализации сельскохозяйственных знаний.

В разделе 3.4 перечислены восемь направлений исследований логики, проводимых в основном с помощью математики. В математике доказательством называется цепочка логических умозаключений, показывающая, что при каком-то наборе аксиом и правил вывода верно некоторое утверждение. В зависимости от контекста может иметься в виду формальное доказательство (построенная по специальным правилам последовательность утверждений, записанная на формальном языке) или текст на естественном языке, по которому при желании можно восстановить формальное доказательство. Аксиомой принято считать исходное положение какой-либо теории, принимаемое в рамках данной теории истинным без требования доказательства и используемое при доказательстве других её положений, которые в свою очередь называются теоремами.

Доказательство – процесс (метод) установления истины, логическая операция обоснования истинности утверждения с помощью фактов и связанных с ним суждений. С помощью совокупности логических приёмов истинность какого-либо суждения обосновывается исходя из других истинных суждений. Есть и другое определение, более короткое: доказательство есть совокупность рассуждений, делающих данное утверждение очевидным.

Доказательство достигает своей цели, когда при помощи его обнаруживается, что данное утверждение есть необходимое следствие аксиом или какого-нибудь другого утверждения, уже доказанного. Всякое доказательство основано на том, что при правильном умозаключении из истинного утверждения нельзя вывести ложного заключения. Правила вывода – предписания или разрешения, позволяющие из суждений одной логической структуры как посылок вывести суждение некоторой логической структуры как заключение. Их особенность состоит в том, что признание истинности заключения производится на основании не содержания посылок, а их логической структуры.

В агрономии наряду с индуктивным (от частного к общему) развит и аксиоматико-дедуктивный способ мышления: многие научно обоснованные рекомендации выступают здесь как аксиомы. Агроно-

мические рекомендации как конкретные истины отражают сущность земледельческих явлений и те специфические условия, в которых эти явления развиваются. В то же время эти конкретные рекомендации имеют ограниченное применение в связи со специфичностью условий, в которых они разработаны.

В соответствии с задачами развития ИИ в области сельского хозяйства можно сформулировать некоторые аксиомы агрономии, которые имеют значение независимо от географии и условий их применения. Этот подход является необходимым в связи с представлениями знаний в какой-нибудь из возможных моделей, например продукционной, или для создания соответствующих онтологий.

Ниже приведены несколько примеров таких аксиом.

Сельскохозяйственные культуры подразделяются на зерновые, зернобобовые, кормовые, масличные, эфиромасличные, технические, овощные, лекарственные, цветочные, плодовые, ягодные, картофель, сахарную свёклу, виноград.

Зерновые культуры входят в состав семейства Мятликовые (Poaceae), или Злаковые (Gramineae), кроме гречихи, которая принадлежит к семейству Гречишные (Polygonaceae). К зерновым культурам относятся 9 ботанических родов: пшеницу, рожь, тритикале, ячмень, овес, кукурузу, просо, сорго, в каждый из них входит обычно несколько видов. В их составе могут быть несколько разновидностей, которые в свою очередь делятся на сорта.

Зерновые культуры делятся на две группы. К первой группе относятся культуры северного происхождения, где летом длинный день: пшеница, рожь, тритикале, ячмень и овес. Во вторую группу входят культуры южного происхождения, где летом короткий день: кукуруза, просо и сорго. Культуры первой группы имеют озимые и яровые формы, второй – только яровые.

С продвижением короткодневных культур на север увеличиваются продолжительность их вегетационного периода и накопление вегетативной массы.

Сельскохозяйственные культуры классифицируются:

– по требованию к плодородию почвы: а) наиболее требовательные культуры к плодородию почвы: пшеница, сахарная свекла, ячмень, горох, лен, фасоль; б) менее требовательные к плодородию культуры,

отличающиеся хорошо развитой корневой системой или повышенной усвояющей способностью корней: озимая рожь, сорго, овес, нут, чина;

– по отношению к эродированности почвы: на эродированных почвах наиболее сильно снижается урожайность у следующих культур: сахарной свеклы, картофеля, подсолнечника, пшеницы, проса. На склоновых почвах допустимо возделывать ячмень, гречиху, зернобобовые, однолетние травы, овёс, озимую рожь, многолетние травы;

– по отношению к кислотности почвы: на щелочных почвах целесообразно возделывать люцерну, сахарную свеклу, нут, капусту, на нейтральных и слабокислых – пшеницу, ячмень, кукурузу, зернобобовые, подсолнечник, на кислых – люпин, картофель, рожь, овес, гречиху, просо, на засоленных – люцерну желтую, донник, житняк, нут, ячмень, рапс, горчицу, сахарную свеклу;

– по отношению к гранулометрическому составу: на легких почвах (супесчаные) целесообразно возделывать озимую рожь, овес, сорго, картофель, эспарцет, люцерну желтую, житняк, на среднесуглинистых – овес, просо, гречиху, ячмень, сорго, подсолнечник, сою, горох, картофель, на тяжелосуглинистых и глинистых – пшеницу, ячмень, кукурузу, подсолнечник, нут, сахарную свеклу, донник, люцерну синюю;

– по отношению к климатическим условиям: в засушливых и теплообеспеченных районах лучше сеять короткодневные засухоустойчивые (с глубокоразвитой корневой системой или экономно расходующие влагу и имеющие транспирационный коэффициент 250–300) культуры: сорго, просо, кукурузу, нут, чину, люцерну, сахарную свеклу, подсолнечник, житняк. Во влагообеспеченных районах – длиннодневные: зерновые, картофель, рапс, кормовые бобы, вику, у которых транспирационный коэффициент составляет 450–500 и более.

Почвы зоны формирования короткодневных культур, как правило, средние и тяжелые по гранулометрическому составу, имеют нейтральную или щелочную реакцию среды, богаты одновалентными и двухвалентными катионами, поэтому культуры короткого дня требуют нейтральных или слабокислых почв с высокой емкостью почвенного поглощающего комплекса.

В северных широтах, где сформировались виды длиннодневного фотопериодизма, почвы чаще легкого гранулометрического состава, слабокислые и кислые, с низким содержанием основных элементов

минерального питания. Эти культуры лучше выдерживают кислые почвы, не богатые питательными веществами.

Для прохождения каждого межфазного периода онтогенеза растению необходима определенная сумма активных температур. Активной температурой считается нижний порог температуры, при которой все физиологические процессы в растении проходят нормально. Условно за такой порог, например для зерновых культур умеренных широт, принята температура 10 °С.

Для прохождения онтогенеза каждому виду и сорту требуется своя сумма активных температур, обусловленная генотипом. Зная сумму активных температур сорта, можно определить ареал устойчивого вызревания его семян, зная сумму активных температур за каждый межфазный период, можно с большой степенью надежности прогнозировать наступление каждой фазы развития.

Для длиннодневных культур с увеличением длины дня сокращаются межфазные периоды, следовательно, и время на накопление массы вегетативных органов; сокращается период вегетации, но при этом снижается масса растений.

Чем в более жестких условиях сформировался вид, тем меньшие требования он предъявляет к условиям выращивания. Чем дальше возделывают вид от зоны его происхождения, тем большее число основных факторов среды приходится корректировать агротехническими приемами, тем больше затрачивать средств на единицу продукции этого вида.

Почва (почвенный покров). В соответствии с законом аналогичных топографических рядов в любой зоне распределение почв на элементах рельефа имеет аналогичный характер. Закон описан в главе 3 данной книги [Захаров, 1927].

Различают три группы факторов, обуславливающих плодородие почвы:

– биологические факторы – количество и состав органического вещества, почвенная биота, биологическая активность почвы, чистота посевов от сорняков, вредителей и болезней;

– агрохимические факторы – содержание и режим питательных веществ, щелочно-кислотные свойства почвы;

– агрофизические факторы – гранулометрический состав почвы, режим влаги, структура, сложение и строение пахотного и подпахотного слоя.

Плодородная почва отвечает следующим требованиям:

- содержит достаточное количество элементов питания и воды;
- обладает способностью аккумулировать воду и питательные вещества;
- обеспечивает оптимальные условия воздушного и теплового режимов;
- пригодна для использования современных машин и орудий;
- устойчива к различным факторам разрушения;
- обладает сильным фитосанитарным эффектом (быстро устраняет почвоутомление).

Воспроизводство плодородия почвы – устранение отрицательных последствий, вызванных в почве возделыванием культурных растений.

Наряду с понятием «плодородие почвы» в агрономии широко используют термин «окультуривание почвы».

Факторы, по которым оценивается окультуренность почв:

- агрофизические показатели плодородия: а) плотность почвы (1,1–1,2 г/см³ – орт изменяется в зависимости от культуры, порозность 50–55%, воздухоемкость 25–30%); б) структура почвы – мелкокомковатая, водопрочность более 40%; в) мощность пахотного слоя – 25–30 см;
- биологические показатели плодородия: а) содержание в пахотном слое гумуса более 3%; б) высокая активность почвенной биоты; в) хорошее фитосанитарное состояние – количество сорняков, вредителей и болезней меньше экономического порога вредоносности.

Для автоморфных зональных почв, не осложненных негативными для сельскохозяйственных растений признаками (солонцеватость, засоленность, уплотненность, слитость, эродированность и др.), а также при однородном гранулометрическом составе, плодородие почв и продуктивность агроценозов определяются комплексом физико-географических условий (тепло- и влагообеспеченностью).

Различные почвы не могут быть одинаково плодородны для всех растений. Плодородие всегда конкретно. Одна и та же почва для одних растений может быть плодородной, для других – малоплодородной. Эта особенность почвенного плодородия является основой рационального размещения сельскохозяйственных растений, направленного на оптимальную специализацию сельскохозяйственного производства.

Растения сохранили в своем генетическом фонде адаптивную реакцию к первоначальной природной среде (Н.И. Вавилов), первоначальную

чальным почвам, с которыми когда-то эти растения составляли оптимальное почвенно-экологическое единство. Данное единство определяло максимальную биологическую продуктивность.

Приемами улучшения режима питания растений являются:

- известкование кислых и гипсование солонцовых почв, а также фитомелеорация;
- выращивание многолетних бобовых трав;
- мероприятия по предотвращению ветровой и водной эрозии;
- внедрение в севообороты высокоурожайных сортов и гибридов культурных растений.

Органическое вещество почвы оптимизирует для растений многие её физические характеристики. Чем выше содержание в почвах органических веществ, тем шире диапазон физической спелости: почвы могут обрабатываться в более широком интервале влажности.

Многогумусные почвы легко обрабатываются и менее подвержены уплотнению. Гумусовые вещества обладают высокой обменной поглотительной способностью. За счет этого биофильные элементы удерживаются коллоидами от вымывания за пределы почвы, но легко доступны корневым системам растений. Высокая поглотительная способность гумусовых веществ также обеспечивает почве способность противостоять вредному воздействию кислотных и щелочных растворов и поддерживать на определенном уровне реакцию среды.

Более богатые в прошлом почвы теряют гумус намного больше, чем малогумусные. По мере снижения содержания в почвах органического вещества темпы дегумификации снижаются. Дегумификация не носит прямолинейный характер, и максимальное снижение содержания гумуса отмечается в первые годы после распашки целинных почв. В дальнейшем процесс дегумификации замедляется и стабилизируется на качественно ином уровне.

Фитосанитарное состояние почвы оценивается по наличию в ней генеративных и вегетативных органов размножения сорных растений, вредителей и возбудителей болезней культурных растений, токсичных веществ, выделяемых корнями растений и микроорганизмами, а также продуктов техногенного загрязнения.

Для улучшения фитосанитарного состояния почвы проводятся следующие мероприятия:

- освоение севооборотов, посевы промежуточных культур;
- выращивание сортов и гибридов культурных растений, устойчивых к вредителям и болезням;
- применение рациональных приемов обработки почвы;
- разработка и осуществление интегрированной системы защиты растений.

Примеры можно продолжать, однако ограничимся приведенными выше. Создание аксиоматики сельского хозяйства – отдельная тема для обсуждения, в том числе в рамках развития онтологий по направлениям исследований (агрономия, животноводство, механизация, экономика и др.).

4.4. Репозиторий Знаний

Репозиторий Знаний – название для БЗ, принятое в Лаборатории и программно работающего на ПО MediaWiki, структурированной в виде семантической сети. То, что существует в виде того, что мы называем Репозиторий Знаний, есть в точности институциональный репозиторий согласно определению в Википедии.

Модель цифрового («институционального») репозитория описана в статье А.М. Федотова и др. (2015), в которой дан сравнительный анализ наиболее популярных систем поддержки институциональных репозиториях и обоснован выбор системы DSpace для организации подсистемы долговременного хранения информационных ресурсов [Федотов и др., 2015].

Авторы статьи для организации системы долговременного хранения информационных ресурсов использовали стандарт ISO-14721:2012 1 (OAIS – Open Archive Information System). На основе этого стандарта создана концепция «институционального репозитория» как системы долговременного хранения, накопления информации и обеспечения надежного доступа к цифровым объектам, представляющим собой результат интеллектуальной деятельности научного коллектива. К основным особенностям институционального репозитория относятся [Федотов и др., 2015]:

- обеспечение разграниченного доступа к разнородным цифровым объектам (публикациям, изображениям и т.д.);

- организация доступа к информационным ресурсам отечественного и мирового сообщества (в том числе с помощью полнотекстового индексирования мировыми поисковыми системами);
- унифицированный доступ к метаданным по стандартным протоколам (поддержка интероперабельности);
- возможность организации единой точки доступа к информационным ресурсам;
- сохранение других информационных ресурсов, в том числе неопубликованных.

Авторы статьи отмечают, что DSpace является самым популярным в академической среде программным обеспечением для создания архива электронных ресурсов (цифрового репозитория). DSpace обеспечивает платформу для долгосрочного хранения цифровых материалов, используемых в научных исследованиях. Платформа DSpace разрабатывалась совместно компанией Hewlett-Packard и библиотеками MIT (Massachusetts Institute of Technology) и является открытым (лицензия BSD) кроссплатформенным Java приложением. Для хранения метаданных используется СУБД Oracle или PostgreSQL.

В Лаборатории в 2019 г. для создания базы знаний и семантических сетей применено серверно-сетевое программное обеспечение MediaWiki, которое предназначено для создания онлайн-энциклопедий и ведёт своё происхождение из программного кода известного онлайн-проект Wikipedia. Созданный в Лаборатории MediaWiki-проект получил наименование «Репозиторий Знаний».

Система программного обеспечения MediaWiki выстроена так же, как в онлайн-проекте Wikipedia: есть страницы и категории. На страницах размещается непосредственная информация, относящаяся к знаниям в соответствии со структурой хранилища данных, при этом каждая страница относится к одной или нескольким категориям. Категории также содержат небольшое описание и перечень входящих в них страниц и категорий. В категории должна находиться хотя бы одна страница, глубина вложения категорий не ограничена. В системе существует поиск по ключевым словам и заглавиям страниц.

Вся цепочка управления знаниями стекается в журналы через метаданные. Метаданные заносятся в компоненты ХД в соответствии со структурой АИС. Структура АИС может выступать в роли корневой

структуры организации знаний в виде иерархического дерева, структура компонентов ХД – дополнять её до частично упорядоченного множества. При размещении элементов знаний в такую структуру автоматически будут формироваться семантические сети (раздел 2.2). Кроме двух упомянутых структур, обеспечивать все полезные свойства семантических сетей будут привязанные к элементам знаний метаданные.

Помимо стандартных функций внесения знаний и поиска по ключевым словам в Репозитории Знаний заложены и стандартные функции систем управления базами данных. Вот некоторые функции для страниц: вывести категории, в которых находится данная страница; показать дерево всех предшествующих категорий; показать дерево связанных через перекрёстные ссылки или метаданные страниц; показать дерево, связывающее две или более различных страниц. Из функций для категорий можно отметить: по списку категорий определить страницы, входящие в каждую из категорий списка; показать дерево вложенных страниц и категорий.

Репозиторий Знаний – удобный инструмент постепенного внесения знаний о предметных областях. В любой момент можно получить связный граф, содержащий страницы, категории и связи между ними и перевести этот граф в формат UML со всеми вытекающими последствиями (раздел 3.5). Таким образом, у эксперта есть выбор: изучать формат UML или разобраться, как качественно вносить знания в Репозиторий Знаний.

Другое применение Репозитория Знаний – комплектование ХД. Например, один из его компонентов – реестр моделей. Модели вносятся в Репозиторий Знаний с использованием таблицы специального формата – Карточки модели. Наименование полей (и тэги в этих полях) таблицы могут сигнализировать об использовании математической формулы. Внесение алгоритма применения модели в Карточку модели обычно без них не обходится. Зафиксировав наличие формулы, записанной на языке LaTeX, можно перевести её в формат языка программирования информационной интеллектуальной системы, внести в перечень методов (как абстрактных объектов UML) для одной или нескольких ПО. Далекое не всё сразу удаётся автоматизировать, тогда это выполняется вручную силами сотрудников научного коллектива.

Семантические сети, автоматически образованные связями между категориями и страницами Репозитория Знаний, могут быть приведены в любом из стандартов, принятых в англоязычном сообществе [Drury et al., 2019]. Для этого необходимо в ключевые слова или мета-данные страниц добавить переводы рассмотренных на этих страницах понятий. Это процесс может быть оптимизирован с помощью специального расширения MediaWiki для работы с семантическими сетями – Semantic MediaWiki.

В заключение отметим, что отечественная и зарубежная наука достаточно далеко продвинулась в теоретических и практических разработках отдельных элементов АИС, пусть и в несколько других конфигурациях. Для примера можно взять «Адаптивно-ландшафтные системы земледелия Новосибирской области» [Адаптивно-ландшафтные..., 2002]. Эта книга содержит огромное количество справочной и структурной информации в явном или неявном виде по всем концептуальным элементам системы (1–16; см. раздел 4.1). Книгу можно рассматривать как отдельный внешний источник данных вида (2). Данные из книги могут быть активно задействованы при концептуальном и онтологическом описании пространств (1, 5, 6), а также при встраивании в АИС конкретных элементов этих пространств – (7, 11, 12). Книга содержит существенную информацию о принятых в научном сообществе форматах данных, поэтому на её основе можно задавать структуру баз оцифрованных данных (3), ориентироваться на книгу при разработке этапов работы с журналами (4, 13, 14). В принципе, на одной этой книге можно построить всю систему (1–16), кроме моделей (5, 9, 10, 12).

Развитие существующих интеллектуальных систем идёт в направлении накопления данных обо всём и выдачи рекомендаций, содержащих огромное количество агропоказателей. В главе 1 мы упоминали о том, что зарубежные исследователи пришли к убеждению о необходимости создания нового поколения сельскохозяйственных систем, моделей и знаний под общим названием NextGen. Одной из причин создания нового поколения интеллектуальных систем является то, что программное обеспечение (frameworks), используемое в этих системах, датируются 1970–1990 гг., до появления современных передовых технологий сбора данных, вычислений, хранения, доступа,

обработки, языков программного обеспечения и стандартов кодирования. К таким системам относятся, например Decision Support System for Agrotechnology Transfer (DSSAT), Agricultural Production Systems Simulator (APSIM) и др.

Изменить парадигму IT-технологий в России в ближайшие годы невозможно, поэтому нужно встраиваться. Чтобы выйти на передний край этого процесса в ближайшее время необходимо практическое решение. Путей (в некотором роде – дорожной карты) быстрой практической реализации аграрной интеллектуальной системы существует шесть:

1) взять одно, несколько или все решения по моделированию сельскохозяйственных объектов, созданные в СФНЦА РАН и других НИИ сельскохозяйственного профиля, объединиться с ними путём дозачленения и актуализации используемых там баз данных за счёт обработки некоторых источников внешних данных вида (2), выделить эти источники, научиться их обрабатывать и отправлять обработанные данные в базу данных программы, получать в программах результаты, научиться эти результаты отображать и интерпретировать;

2) создать сетевую систему сбора данных, информации и знаний в СФНЦА РАН, воспроизвести АИС (1–16) и после закрепления и отработки всех функций системы стараться частично оптимизировать её работу с помощью автоматизации. Для этого нужно профессиональное управление, сильный административный ресурс, так как будет задействовано большое количество исследователей, инструкций и регламентов работ. Убеждены, что весь необходимый в системе (1–16) функционал существует у многих сотрудников СФНЦА РАН;

3) взять хорошо развитую программу имитационного моделирования и в ней динамически формировать все возможные процессы в сельском хозяйстве, обчислять их и представлять графическую и анимированную визуализацию внешних и внутренних данных. Для этого имеются полностью удовлетворяющие этим требованиям бесплатно распространяемые среды имитационного моделирования, например true-world. Для освоения в работе этой программы нужно разобратся в формате входных данных (это база данных из 35 таблиц), сконструировать вход, запустить программу и постепенно регулировать входные данные;

4) создать рабочую группу из двух человек – инженера по знаниям и IT-специалиста – и выработать концепцию реализации распределённой системы, основной блок которой будет лежать на сервере СФНЦА РАН, но она будет обращаться в различные интернет-сервисы для выполнения отдельных операций. Для этого нужна работа над описаниями свойств и функций системы с расстановкой их приоритетов, на основе которых формировать списки необходимых для реализации инструментов, найти эти инструменты и прописать способ их соединения с центральным блоком;

5) создать программный продукт, обладающий всеми необходимыми функциями с нуля. Для этого нужно найти талантливую и продуктивную программиста, которому не нужно техническое задание и который быстро поймёт весь необходимый функционал системы по описанию (1–16), обеспечить программисту приемлемые условия работы, быстро предоставлять ему ответы на возникающие по ходу вопросы, активно участвовать в тестировании и комментировании текущих версий программы, первая из которых уже появится через 3 недели после начала работы;

6) создать в современной среде разработки типа WinDEV минимальный программный продукт и постепенно его разворачивать: для этого нужна работа над описаниями свойств и функций системы с расстановкой их приоритетов, на основе этого сформировать техническое задание, далее разобраться в методах разработки WinDEV и осуществить разработку.

Любое из этих решений может привести к быстрым результатам в течение месяцев, но между собой они мало совместимы или несовместимы вовсе.

ГЛАВА 5

МАШИННОЕ ОБУЧЕНИЕ И ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ

5.1. Постановка задачи применительно к предметной области, термины и определения

В последние годы для решения ряда задач, которые считаются весьма сложными с точки зрения формального логического описания и построения четких алгоритмических моделей, используются методы машинного обучения (machine learning) как один из подходов в рамках более общего термина ИИ. Данная возможность появилась в связи с огромным прогрессом в развитии вычислительной техники в последние десятилетия. С самого начала следует уточнить, что ИИ вовсе не означает, что у компьютера появился реальный интеллект и ему напрямую можно поручать творческие задачи, которые ранее можно было поручить опытному и высокоинтеллектуальному специалисту. Скорее, данный термин следует понимать как сложившийся в исторической перспективе, так как подходы к решению подобных задач были созданы в попытке скопировать работу нейронных сетей головного мозга.

В инженерных науках подобные подходы широко известны в качестве так называемого биодизайна. Например, летая на современных самолетах, мы наблюдаем изгибы на кончиках крыльев – винглеты. Данные элементы конструкции крыла служат для срыва турбулентных потоков с краев несущих плоскостей крыльев, что приводит к тому, что практически вся плоскость крыла обтекается ламинарным потоком, таким образом достигается максимальная подъемная сила. Этот элемент конструкции был подсмотрен в природе. Хищные птицы, парящие в восходящих потоках воздуха, используют этот же прием, оттопыривая перья на кончиках крыльев. Данный факт стал подсказкой для инженеров в области аэродинамики, однако крыло самолета во всем остальном – это совсем не то, что крыло птицы.

Данная аналогия могла быть самостоятельно обнаружена машиной при автоматизации заполнения таблицы шаблонов знаний «совместимость объектов» (см. главу 2, рис. 18); объекты «винглет» и «крыло птицы» имеют сходство и тем самым конфигурации объектов «полёт самолёта» и «полёт птицы» имеют в своём составе похожие объекты.

Существует множество попыток дать разумное определение термина машинное обучение, чтобы корректно описать возможности и ограничения данного метода [Alpaydin, 2020]. С одной стороны, методы машинного обучения позволяют сегодня отлично справляться с задачами, еще недавно относившимся к творческим и практически неразрешимым, например распознаванием образов, переводом текстов, игрой в шахматы или го (кит.-вэйци). С другой стороны, специалисты в данной области прекрасно знают, что результат решения подобной задачи напрямую зависит от корректности формулировки задачи человеком-экспертом, качества и количества обучающих данных и др. Сформулированное Томом Митчелом (2006) определение машинного обучения является, на наш взгляд, одним из наиболее корректных: «Мы считаем, что машина обучается с учетом конкретной задачи T , системы оценки эффективности P для конкретной задачи, основываясь на опыте E . В зависимости от того, как мы определяем T , P и E , задачу обучения можно назвать добычей данных, автономными исследованиями, обновлением базы данных, программированием на основе примеров и т.д.» [Mitchell, 2006].

От себя можно добавить, чтобы еще больше конкретизировать данное определение, что обучается вовсе не машина или некие электронные мозги, а конкретная программа, например виртуальная нейронная сеть. Таким образом, программа на каждой из итераций процесса становится точнее в решении поставленной задачи или создании ожидаемого «образа» в соответствии с метрикой, определенной человеком.

То, что ИИ – вовсе не человекоподобный разум, способный решать абстрактные задачи, можно понять, обнаружив наличие вполне определенных базовых подходов, применяемых в данной области знаний:

– «дерево решений» (decision tree). Наиболее распространенными алгоритмами обучения в этой категории являются деревья классификации и регрессии (classification and regression trees) [Breiman et al.,

1984], детектор автоматического взаимодействия хи-квадрат (chi-square automatic interaction detector) [Kass, 1980] и итерационный дихотомайзер (iterative dichotomiser) [Quinlan, 1993];

– «случайный лес» (random forest). Обычно используют два известных метода – повышения (boosting) [Shapire et al., 1998] и сбор Бреймана (bagging Breiman) [Breiman, 1996];

– «искусственная нейронная сеть» (artificial neural network). Алгоритмы обучения, обычно используемые в искусственной нейронной сети, включают сети радиальных базисных функций (radial basis function networks) [Broomhead, Lowe, 1988], алгоритмы перцептрона (perceptron algorithms) [Rosenblatt, 1958], обратное распространение (back-propagation) [Linnainmaa, 1976] и упругое обратное распространение (resilient back-propagation) [Riedmiller, Braun, 1993] и др.;

– «наивный байесовский классификатор» (naive Bayes). Кроме этого, применяется байесовская сеть (bayesian network) [Pearl, 1988], смесь гауссианов (mixture of gaussians) [Duda et al., 1973] и байесовская сеть убеждений (bayesian belief network) [Neapolitan, 1987];

– «глубокое обучение» (deep learning – DL). DL обычно называют глубокие нейронные сети (deep neural networks (DNNs) [LeCun et al., 2015].

К методам машинного обучения также относят «тренировку» Байесовских сетей. Данные сети по своей структуре аналогом нейронной сети не являются, а представляют собой ориентированные ациклические графы. Данный факт лишний раз подтверждает, что понятие «машинное обучение» – это сборное определение подходов со схожей идеологией для решения широкого круга аналитических задач. Конкретные постановки задач, связанные с применением пяти базовых подходов машинного обучения входят в группу понятий «Операции и алгоритмы обработки данных» (раздел 2.1) и первоначально формируются в таблицах шаблонов знаний (прежде всего «Источники данных» и Последовательности» – см. главу 2, рис. 13, 14.), с помощью языка UML (раздел 3.2) изображаются на диаграммах классов (в виде методов) и диаграммах последовательностей. В таблицах методов (см. главу 3, табл. 11.) алгоритмы машинного обучения встречаются в ячейках голубого цвета.

Рассмотрим особенности перечисленных выше алгоритмов.

Дерево решений (decision tree).

Дерево решений – метод, являющийся одним из наиболее эффективных инструментов интеллектуального анализа данных и предсказательной аналитики, которые позволяют решать задачи классификации и регрессии.

Дерево решений представляет собой иерархическую структуру, состоящую из решающих правил вида «Если ..., то ...». Правила автоматически генерируются в процессе обучения на обучающем множестве и формулируются практически на естественном языке. Например: «Если длительное время была дождливая погода, то возможен вынос элементов из почвы и необходимо внесение удобрений», или наоборот: «Если в течение длительного времени не было осадков, а запасов влаги в почве в начале периода недостаточно, то необходим полив или ожидается неурожай». Деревья решений как аналитические модели являются более наглядными для понимания, чем, например, многослойные нейронные сети.

Поскольку правила в деревьях решений получаются путём обобщения множества отдельных наблюдений (обучающих примеров), описывающих ПО, то по аналогии с соответствующим методом логического вывода, их называют индуктивными правилами, а сам процесс обучения – индукцией деревьев решений.

Развитие деревьев решений как самообучающихся моделей для анализа данных связано с именами Джона Куинлена [Quinlan, 1986; Quinlan, Ross, 1993], который разработал алгоритм ID3 и его усовершенствованные модификации C4.5 и C5.0, а также Лео Бреймана [Breiman, 2001], который предложил алгоритм CART и метод случайного леса.

Общими достоинствами алгоритмов, основанных на подходе «дерева решений» являются:

- простота в понимании и интерпретации;
- нетребовательны к подготовке данных. Прочие техники требуют нормализации данных, добавления фиктивных переменных, а также удаления пропущенных данных;
- способны работать как с категориальными, так и с интервальными переменными. Прочие методы работают лишь с теми данными,

где присутствует только один тип переменных. Например, метод отношений может быть применён только на номинальных переменных, метод нейронных сетей – только на переменных, измеренных по интервальной шкале;

- используют модель «белого ящика». Если определённая ситуация наблюдается в модели, то её можно объяснить при помощи булевой логики. Примером «черного ящика» может быть искусственная нейронная сеть, так как результаты вычислений данной модели поддаются объяснению с трудом;

- позволяют оценить модель при помощи статистических тестов. Это даёт возможность оценить надёжность модели;

- являются надёжным методом. Метод хорошо работает даже в том случае, если были нарушены первоначальные предположения, включённые в модель;

- позволяют работать с большим объёмом информации без специальных подготовительных процедур. Данный метод не требует специального оборудования для работы с большими базами данных.

Недостатки:

- проблема получения оптимального дерева решений является NP-полной с точки зрения некоторых аспектов оптимальности даже для простых задач. Вычислительная задача называется NP-полной (non-deterministic polynomial – «недетерминированные с полиномиальным временем»), если для неё не существует эффективных алгоритмов решения. Таким образом, практическое применение алгоритма деревьев решений основано на эвристических алгоритмах, таких как алгоритм «жадности», где единственно оптимальное решение выбирается локально в каждом узле. Такие алгоритмы не могут обеспечить оптимальность всего дерева в целом;

- в процессе построения дерева решений могут создаваться слишком сложные конструкции, которые недостаточно полно представляют данные. Данная проблема называется переобучением. Для того чтобы её избежать, необходимо использовать метод «регулирования глубины дерева»;

- существуют концепты, которые сложно понять из модели, так как модель описывает их сложным путём. В этом случае мы имеем дело с непомерно большими деревьями. Существует несколько подходов

решения данной проблемы, например попытка изменить репрезентацию концепта в модели (составление новых суждений) или использование алгоритмов, которые более полно описывают и репрезентируют концепт (например, метод статистических отношений, индуктивная логика программирования);

– для данных, которые включают категориальные переменные с большим набором уровней (закрытий), большой информационный вес присваивается тем атрибутам, которые имеют большее количество уровней.

Случайный лес (random forest).

Случайный лес – алгоритм машинного обучения, предложенный Лео Брейманом и Адель Катлер, заключающийся в использовании комитета (ансамбля) решающих деревьев [Breiman, 2001]. Алгоритм применяется для задач классификации, регрессии и кластеризации. Основная идея заключается в использовании большого ансамбля решающих деревьев, каждое из которых само по себе даёт очень невысокое качество классификации, но при этом хороший результат может быть получен за счёт большого количества используемых деревьев.

Достоинства:

– способность эффективно обрабатывать данные с большим числом признаков и классов;

– нечувствительность к масштабированию (и вообще к любым монотонным преобразованиям) значений признаков;

– одинаково хорошо обрабатываются как непрерывные, так и дискретные признаки. Существуют методы построения деревьев по данным с пропущенными значениями признаков;

– существуют методы оценивания значимости отдельных признаков в модели;

– внутренняя оценка способности модели к обобщению (тест по неотобраным образцам out-of-bag);

– высокая параллелизуемость и масштабируемость.

Недостатки:

– большой размер получающихся моделей. Требуется большое количество используемой памяти для хранения модели в зависимости от числа деревьев.

Искусственная нейронная сеть (artificial neural network).

Искусственная нейронная сеть (ИНС) – это совокупность искусственных нейронов, организованная аппаратными или программными средствами [Hopfield, 1982; Hassoun, 1995]. Сеть строится по принципу организации и функционирования биологических нейронных сетей – сетей нервных клеток живого организма. Сети организуются по принципу конструктора, собирая определенное количество нейронов послойно. Входные данные подаются на «входной слой», итоговый результат генерируется «выходным слоем». Слои, которые находятся посередине, называются «скрытыми».

Принцип работы каждого из элементов сети (искусственного нейрона) подсмотрен в живых системах и является природоподобным. В природе каждый нейрон снабжен несколькими дендритами и одним синапсом. Дендриты – это короткие и сильно разветвлённые отростки нейрона, служащие главным местом для образования влияющих на нейрон возбуждающих и тормозных сигналов (синапсов). После преодоления нейроном порога возбуждения от нейрона формируется исходящий сигнал, передаваемый по синапсу.

Похожий принцип реализован и на примере искусственного нейрона. У каждого из таких нейронов есть несколько входных сигналов, которые могут быть как «возбуждающими» (со знаком «+»), так и тормозящими (со знаком «-»). Эти сигналы с определенными коэффициентами (весами) подаются на нейрон. Как только суммарное управляющее воздействие на нейрон превышает некий порог (называемый порогом активации), нейрон генерирует сигнал, передающийся на нейрон следующего слоя искусственной нейронной сети. В следующем слое этап повторяется. При этом по-прежнему на каждый из нейронов слоя поступает несколько сигналов, сгенерированных в предыдущем слое, а выдается единственный сигнал.

В процессе «обучения» или «тренировки» ИНС происходит подбор «весов» входящих сигналов таким образом, чтобы на наборе обучающих данных получить максимальное соответствие данных соответствующему обучающему результату. В дальнейшем такая «обученная» сеть способна сама выдавать результаты по новым исходным данным с высокой степенью прогнозной достоверности.

Сильной стороной ИНС является их способность решать довольно сложные абстрактные задачи, которые ранее считались в чистом виде творческими, например распознавать сложные графические образы или тексты. При этом подобные сети могут потреблять относительно мало вычислительных ресурсов по сравнению, например, с методом «дерева решений». Однако процессы, происходящие внутри сетей в момент «обучения» или «тренировки», очень слабо поддаются пониманию и строгому описанию. В некотором смысле, собранные под задачу ИНС в дальнейшем они превращаются в «черный ящик», и следовательно приходится просто доверять генерируемым результатам.

В последние годы данный факт вызывает серьезную озабоченность специалистов по интерпретации данных. Существует, как минимум, два последствия того факта, что мы вынуждены воспринимать тренируемые ИНС в качестве «черного ящика» и доверять полученному результату, не совсем понимая, как данный результат был получен.

Во-первых, результаты тренировки всецело зависят от количества и качества используемых данных. Если в данных присутствует системная ошибка, она приведет к соответствующей настройке «весов» входящих сигналов где-то в глубине слоев нейронной сети. Внести изначальные ограничители, чтобы отследить критические ошибки, не представляется возможным. Таким образом, вся ответственность за результаты расчетов ложится на тех (того), кто представил некорректные или неполные данные. Если речь идет о расчётах, которые могут привести к чувствительным потерям для окружающих в случае ошибочных суждений, мера ответственности за данные и их качество возрастает многократно.

Во-вторых, если не до конца понятен алгоритм, которым получен итоговый результат, насколько вправе специалист использовать данные результаты в чувствительных областях, например при постановке диагноза пациентам, даже если все результаты расчетов приводили к верным результатам? В конце концов, обучение таких нейронных сетей – это процесс, бесконечно растянутый во времени. С точки зрения ИНС, ошибочно поставленный диагноз в данном конкретном случае – просто еще один обучающий пример и повод «перевзвесить» входящие сигналы в слоях нейронов, чтобы следующий диагноз был более точным. Однако с точки зрения пациента – это вопрос жизни и смерти.

Таким образом, общепринятым в настоящий момент подходом является то, что ИНС – прекрасный инструмент, однако на 100% полагаться на результат расчета нейронной сети в чувствительных областях не представляется возможным. Здесь результат такого расчета используется в качестве консультативного, а окончательное решение остается за специалистом-экспертом.

Наивный байесовский классификатор (naive Bayes).

Наивный байесовский классификатор – вероятностный классификатор, основанный на применении теоремы Байеса со строгими (наивными) предположениями о независимости [Murphy, 2006; Николенко, Тулупьев, 2009].

Наивный байесовский классификатор – один из самых простых из алгоритмов классификации, однако очень часто он работает не хуже, а то и лучше более сложных алгоритмов. Достоинством наивного байесовского классификатора является малое количество данных, необходимых для обучения, оценки параметров и классификации.

Глубокое обучение (deep learning).

Глубокое обучение характеризуется как класс алгоритмов машинного обучения, который [Deng, Yu, 2014]:

- использует многослойную систему нелинейных фильтров для извлечения признаков с преобразованиями. Каждый последующий слой получает на входе выходные данные предыдущего слоя. В данном случае речь может идти о многослойных (глубоких) нейронных сетях. Система глубокого обучения может сочетать алгоритмы обучения с учителем и без учителя, при этом анализ образца представляет собой обучение без учителя, а классификация – обучение с учителем;

- обладает несколькими слоями выявления признаков или параметров представления данных (обучение без учителя). Признаки организованы иерархически, признаки более высокого уровня являются производными от признаков более низкого уровня;

- является частью более широкой области машинного обучения и изучения представлений данных;

- формирует в процессе обучения слои на нескольких уровнях представлений, которые соответствуют различным уровням абстракции;

- слои образуют иерархию понятий.

Все определения констатируют наличие нелинейной обработки признаков каждого слоя, формируя иерархию от низкого до высокого уровня.

Состав конкретных нелинейных слоёв зависит от решаемой проблемы. Используются как скрытые слои нейронной сети, так и слои сложных логических преобразований. Система может включать скрытые переменные, организованные послойно в глубоких генеративных моделях, таких как узлы в глубокой сети доверия и глубокой ограниченной машине Больцмана.

Алгоритмы глубокого обучения противопоставлены алгоритмам неглубокого обучения по количеству параметризованных преобразований, с которыми сталкивается сигнал, распространяющийся от входного слоя к выходному слою, где параметризованным преобразованием считается такой блок обработки данных, у которого есть обучаемые параметры, такие как веса или пороги [Shmidhuber, 2015].

Системы глубокого обучения применяются в таких областях, как компьютерное зрение, распознавание речи, обработка естественного языка, аудиораспознавание, биоинформатика, где для ряда задач были продемонстрированы существенно лучшие результаты, чем ранее.

В англоязычной научной литературе активно обсуждается применение машинного обучения в системах сельскохозяйственного производства. В статье Лиакоса с соавт. (2018) осуществлен обзор публикаций и проанализированы работы по машинному обучению, применяемого для управления продуктивностью посевов, включая приложения по прогнозированию урожайности, выявлению болезней, качеству сельскохозяйственных культур и распознаванию видов; управления животноводством, включая сведения о состоянии животных; управления водными ресурсами; управления плодородием почвы [Liakos et al., 2018]. Авторы обзора подчеркивают, что машинное обучение с использованием сенсорных данных и систем управления фермой способствует развитию программ с поддержкой ИИ в режиме реального времени, которые в свою очередь предоставляют обширные рекомендации для поддержки принятия решений и действий фермеров.

Однако осталось и скептическое представление, что глубокое обучение – это не что иное, как модное слово для иного обозначения ИНС. По крайней мере, из внимательного прочтения определений,

относящихся к «слоям» обучения, вполне можно сделать такой вывод.

Пять подходов к машинному обучению могут выполняться с учителем или без учителя.

Обучение с учителем (supervised learning)

Обучение с учителем предполагает наличие полного набора размеченных данных для тренировки модели на всех этапах ее построения. Наличие полностью размеченного массива данных означает, что каждому примеру в обучающем наборе соответствует ответ, который алгоритм должен получить. Таким образом, размеченный массив данных, например из фотографий растений, обучит нейронную сеть, где изображены розы, ромашки или пшеница. Когда сеть получит новое фото, она сравнит его с примерами из обучающего массива данных, чтобы предсказать ответ.

Обучение с учителем больше всего подходит для задач, когда имеется внушительный набор достоверных данных для обучения алгоритма. Но так бывает далеко не всегда. Недостаток данных (неполнота или неточность) – наиболее часто встречающаяся проблема в машинном обучении. В рассматриваемой области применения (прогноз сельскохозяйственной деятельности) – это наиболее актуальная из существующих на сегодня проблем.

Обучение без учителя (unsupervised learning)

В обучении без учителя у модели есть набор данных и нет явных указаний, что с ним делать. ИНС пытается самостоятельно найти ассоциации и соотношения в данных, извлекая полезные признаки и анализируя их.

В зависимости от задачи модель систематизирует данные по-разному:

- кластеризация. Можно посмотреть на коллекцию фотографий птиц и разделить их на группы по видам, опираясь на цвет пера, размер или форму клюва. Именно в этом заключается кластеризация – наиболее распространенная задача для обучения без учителя. Алгоритм подбирает похожие данные, находя общие признаки, и группирует их вместе;

- обнаружение аномалий. Банки могут обнаружить мошеннические операции, выявляя необычные действия в покупательском поведении клиентов. Например, подозрительно, если одна кредитная карта используется в Москве и Владивостоке в один и тот же день.

Похожим образом обучение без учителя используют для нахождения некорректных данных;

– ассоциации. Рассматривая пару ключевых признаков объекта, модель может соотнести с ними другие, с которыми существует связь. Наиболее наглядно данные технологии работают в контекстной рекламе, с которой сталкивается любой пользователь интернета. После того, как пользователь набрал в поисковике отель в городе N, в браузере начинают всплывать подсказки об отелях, туристических обзорах и наиболее дешевых билетах по данному маршруту. Это пример ассоциаций: некоторые характеристики объекта коррелируют с другими признаками. В медицинской генетике, анализируя огромные массивы данных, находят ассоциации изменений (точечных мутаций) в определенных генах с частотой проявления определенных заболеваний у пациентов. Таким образом выявляют наследственные заболевания;

– автоэнкодеры принимают входные данные, кодируют их, а затем пытаются воссоздать начальные данные из полученного кода. Не так много реальных ситуаций, когда используют простой автоэнкодер. Но стоит добавить слои и возможности расширятся: используя зашумленные и исходные версии изображений для обучения, автоэнкодеры могут удалять шум из видеоданных, изображений или медицинских сканов, чтобы повысить качество данных. Также автоэнкодеры используются для сжатия данных и определения минимально возможного объема информации, требуемого для восстановления исходных образов или данных. Как это работает, будет показано ниже.

В обучении без учителя сложно вычислить точность алгоритма, так как в данных отсутствуют «правильные ответы» или метки. Но размеченные данные часто ненадежные или их слишком дорого получить. К тому же систематическая ошибка в начальных данных приводит к систематически неправильно настроенному алгоритму решения. В таких случаях, предоставляя модели свободу действий для поиска зависимостей, можно получить хорошие результаты [osnovy-data-science].

Для выбора подходящего алгоритма специалисты (эксперты) должны изначально проанализировать задачу в поисках наиболее подходящих решений, исходя из объема и структуры имеющихся данных и характера решаемой задачи.

По нашего рассмотрения – возможность получения прогнозов возможных результатов сельскохозяйственной деятельности на предстоящий сезон с определенной степенью достоверности. Данные прогнозы могут быть рассмотрены в качестве альтернативных сценариев при выборе оптимальной стратегии планирования посевов культур с целью оптимизации деятельности агропредприятия или группы агропредприятий на локальном, региональном и глобальном уровне. Данный вид прогнозирования, по определению, является попыткой прогнозирования сложных многофакторных систем в непрерывно меняющихся условиях, многие из которых определяются лишь с некоторой степенью достоверности. Во многом данная задача сходна с так называемыми творческими задачами и сильно отличается от классических задач прогнозирования на основе устоявшихся моделей. Именно в силу этих причин логично применить подходы ИИ к данному виду прогнозной деятельности.

5.2. Достоверность и полнота исходных данных и релевантность прогноза. Понятие «образа» события

Задача предсказания результатов сельскохозяйственной деятельности лежит в области достаточно абстрактных предположений при наличии неполных данных и отсутствии алгоритмов со 100%-й достоверностью. В любом случае предсказание будущих событий очень похоже на задачу создания «образа» по неполным данным. События в будущем периоде, которое мы должны описать здесь и сейчас, – это лишь картинка, которую необходимо нарисовать, опираясь на имеющиеся данные о предполагаемом событии, а потом сравнить с той картинкой, которую будем наблюдать в момент времени совершения предсказанного события. Отличие нарисованной картинке от реальной (осуществленной) – это и есть «функция ошибки», которую необходимо будет минимизировать в процессе поиска оптимального решения задачи.

В дальнейшем мы будем стараться как можно чаще использовать именно понятие «образа» и избегать таких понятий, как «модель» или «прогноз». На наш взгляд, именно данный термин наиболее четко отражает суть решаемой задачи. Условия, в которых происходит сельскохозяйственная деятельность, прежде всего погодные условия, яв-

ляются вариативными и во многом недоопределенными. Невозможно иметь точный прогноз погоды на следующий сельскохозяйственный год. Погодные условия нынешнего периода никогда больше в точности не повторятся. Таким образом, не существует объективной реальности, в которой можно было бы проверить на работоспособность численную «модель», так как невозможно точно повторить эксперимент. Если же ввести понятие «образ», это позволит разрешить обозначенные парадоксы. Образ создается по многим факторам и может в разной степени соответствовать действительности. Образ можно рассматривать так же, как комплект массивов данных, сформированный в отношении конкретного сельскохозяйственного объекта в некоторый момент времени и отражающий группу понятий «Цифровой объект», а любая информация об образе, связанная с его изменениями во времени – группу понятий «Комплекс изменений цифрового объекта» (раздел 2.2). Подходы к формированию образа для сельскохозяйственного объекта предложены в разделе 2.4. С точки зрения UML (раздел 3.2), образ – это раздел фрагментов диаграмм UML различных видов. С точки зрения общей структуры сельскохозяйственной ИИС (разделы 4.1–4.2), образ – это комплект баз данных или их фрагментов как компонентов ХД, собираемых активными контурами и узлами АИС для решаемой сельскохозяйственной задачи. Главная логическая функция рассматриваемых в этой главе алгоритмов машинного обучения – определить допустимые отличия образа от наблюдаемого результата в правильно выбранной метрике.

Следует проанализировать, какими методами в настоящее время наиболее успешно решаются «идеологически» сходные задачи. Очевиден огромный прогресс, достигнутый ИИ в решении задач распознавания визуальных образов в последние годы. Действительно, постановка задачи выглядит схожим образом. Имеются некие исходные данные об интересующем нас образе. Данных может быть весьма много, при этом их качество и полнота оставляют желать лучшего. По этим данным создается «образ» (это довольно широкое понятие, необязательно картинка с портретом, хотя на данном примере происходит достаточно понятная обывателю визуализация), который в дальнейшем сравнивается с оригиналом. Отличие образа и оригина-

ла – и есть та ошибка, которую необходимо минимизировать в процессе настройки алгоритма (обучения).

Из опыта последнего десятилетия следует, что для решения подобных задач наилучшим образом подходит глубокое или машинное обучение. В общем виде процесс выглядит следующим образом. Существует некий обучающий массив данных, в котором присутствуют данные, влияющие неочевидным образом на конечный результат (событие, образ). Существует также набор результатов (событий, образов), которые могут быть поставлены в соответствие определенным комбинациям исходных данных. Комбинация задается экспертом, т.е. работа эксперта заключается в первую очередь в назначении алгоритма выбора подмножества из множества данных, каждое из которых (подмножеств) будет поставлено в соответствие определенному результату (событию, образу).

На следующем этапе выбирается подходящий машинный алгоритм анализа данных или ИНС из имеющихся в доступе, или нейронная сеть собирается из имеющихся в доступе вычислительных модулей. Сеть выбирается таким образом, чтобы на вход можно было подать в качестве исходных значений выбранное подмножество данных, а на выходе иметь ожидаемый результат (образ). В тренировочном массиве данных присутствуют как данные, так и фактические результаты. Функция ошибки в этом случае – разница между результатом (образом), генерируемом ИНС или иным машинным алгоритмом, и фактическим результатом, имеющимся в распоряжении. Далее происходит итерационная «тренировка» сети или алгоритма с использованием тренировочных данных на минимизацию данной функции ошибки.

Для прогноза будущих событий выбирается массив данных, распределенных по времени события, уже наблюдавшихся в течение длительного времени. Таким образом нейронная сеть или машинный алгоритм тренируется сопоставлять подмножества данных из периода, предшествующему событию, с результатом события. После того, как сеть на имеющемся массиве данных соотносит прогнозные значения событий с реальными значениями с приемлемой точностью, можно приступить к следующему этапу.

На следующем этапе используют данные, полученные уже в следующем временном промежутке, предшествующем прогнозируемо-

му. Уже натренированная сеть создает «образ» будущего в соответствии со своим алгоритмом, показавшим эффективность на обучающем массиве данных. Осталось дождаться результата и сравнить с предсказанным значением. После получения результата получим функцию ошибки, которую также можно использовать для дальнейшего обучения. Таким образом, мы имеем моделирование процесса «накопления опыта» системой.

Проиллюстрируем применимость данного подхода к конкретной задаче сельскохозяйственной деятельности. Очевидно, что весьма значительный вклад в потенциальную урожайность сельскохозяйственных культур вносят природные факторы. При этом именно природно-климатические факторы являются весьма переменными и трудно прогнозируемыми. Для каждой территории возможно отнесение к той или иной природно-климатической зоне. Климат изменяется очень медленно по мере смены климатических эпох. Однако в пределах одной климатической зоны возможны весьма резкие колебания погодных условий как от суток к суткам, так от сезона к сезону. Так же от сезона к сезону изменяются динамические (развернутые по времени) погодные параметры. Почему это может быть важно? Объясним на простом примере. Сельскохозяйственная деятельность зависит от жизненных циклов живых объектов: растений и животных. Пусть в исследуемой климатической зоне возобновление вегетации озимых культур ожидается в середине апреля. Исследуемый апрель выдался соответствующим средним климатическим нормам по средним температурам. Ожидается, что возобновление вегетации культур, как и положено, произойдет с 10 по 15 апреля. При этом в данном месяце случился однократный ночной заморозок. Скорее всего, данное событие не повлияет на средние месячные или декадные величины температуры, но будет очень большая разница по результату воздействия на культуру, случится данный заморозок в I или III декаде месяца. Во втором случае, если заморозок произошел после возобновления вегетации культуры, урожайность ее в сильной степени уменьшится.

Прогнозы погоды являются достоверными лишь в перспективе нескольких суток. Очевидно, что в текущем периоде практически невозможно сделать предсказание погодных условий на следующий сельскохозяйственный год с точностью, достаточной для планиро-

вания корректирующих мероприятий. Однако есть понимание, что средние значения климатических параметров для данной территории за длительный временной промежуток являются весьма устойчивыми величинами. Можно предположить, что наблюдение за достаточно большим количеством погодных факторов на выбранной территории может быть ассоциировано с условиями для сельскохозяйственной деятельности (для каждой конкретной культуры) на предстоящий сельскохозяйственный сезон.

В дальнейшем будет приведен пример использования данных по рядам урожайности и рядам погодных данных, основанных на среднесуточных температурах. Из этих рядов будут сформированы массивы обучающих примеров. Например, результату по определенной культуре, представленном в виде хорошая урожайность/плохая урожайность (удачный/неудачный год), будет сопоставлена погодная картина предыдущего года. Будет показано, что реальные данные не всегда полные и не всегда корректные. Будет показано, что для адекватной тренировки методом машинного обучения необходимо достаточное количество качественных обучающих примеров. Однако даже наличие неполного массива обучающих данных позволяет убедиться, что система способна к обучению и можно надеяться на получение более существенного результата в будущем по мере получения достаточно-го числа данных и по мере оптимизации применяемых подходов. Данный факт будет показан на реальных примерах в конце данной главы.

5.3. Влияние корректности данных на точность результата по опыту решения некорректных обратных задач.

Влияние аппаратной функции на детализацию образа

Для визуализации (для лучшего понимания) процесса получения знаний из данных, а также для осознания ограничений на качество и количество используемых данных можно привести аналогию с получением информации в процессе физического эксперимента. В теории физического эксперимента существует понятие «аппаратной функции», т.е. функции отклика измеряющего устройства в ответ на воздействие изучаемого физического свойства на данное устройство. В каком-то смысле физический эксперимент моделирует процесс познания окружающего мира, только на более простом и интуитивно по-

нятном уровне. Разберем понятие аппаратной функции и её влияние на требования к точности производимых измерений, в нашем случае, на полноту используемых данных.

На вход измерительного устройства подается измеряемое физическое воздействие, например электромагнитное излучение (свет), интенсивность и длину волны которого нам предстоит измерить. На выходе прибора имеем данные в виде численных значений (например, мВ), которые связаны с искомыми (измеряемыми) параметрами через аппаратную функцию. Пусть искомое физическое воздействие имеет вид непрерывной функции распределения (например, спектр излучения исследуемого источника). На выходе измеряющего устройства имеем набор данных (разовых измерений) в зависимости от использованных на входе световых фильтров:

$$D(m) = \int K(m, x)f(x)dx, \quad (1)$$

где, $D(m)$ – данные, т.е. измеренная величина датчика (ток или напряжение), в зависимости от фильтра m ; f – искомая функция распределения в зависимости (в данном случае) от длины волны x ; K – аппаратная функция фильтра m .

Изначально фильтры калибруются с помощью стандартных источников излучения. Калибровка в данном случае – аналог «тренировки» с использованием тестовых данных. Набор значений измерений на выходе из прибора – наши данные, неизвестный спектр излучения – та искомая функция $f(x)$, которую необходимо восстановить, используя массив экспериментальных точек $D(m)$. В некотором смысле, восстанавливается «образ» или «прогноз», т.е. некая функция, которую в дальнейшем можно будет использовать в расчетных (предсказательных) целях с приемлемой точностью. Таким образом, мы приходим к формулировке классической обратной задачи: *имея набор экспериментальных точек и зная аппаратную функцию прибора, необходимо восстановить функцию воздействия на прибор.*

При этом данные измерений получены с некоторой погрешностью, могут быть неполными (частично утерянными). Данные условия приводят к вопросу о чувствительности алгоритма обратной задачи к ошибкам (неточностям определения) данных. Во многом это зависит от аппаратной функции используемых устройств. Например, если

аппаратная функция имеет форму Δ -функций или ступеньки, то точность (неопределенность) данных будет соответствовать точности решений. Если аппаратная функция является непрерывной и достаточно гладкой функцией, то ошибки или неопределенности измеренных значений будут драматически влиять на получаемый «образ» искомой физической величины, что будет показано на примере ниже.

Те же проблемы возникают при получении «образа» или «прогноза» методом ИИ. В машинном обучении есть понятие «переобученной модели». Поясним на примере. Например, есть задача идентификации зрительных образов – опознавание объекта по фотографии. В случае, если аппаратные функции имеют форму Δ -функций, это означает 100%-е соответствие определенной фотографии определенному персонажу. Если размерность пространства решений принимается за ∞ , т.е. ничем не ограниченную точность, то в итоге это означает, что попиксельно оцифрованная картинка будет поставлена в соответствие персонажу, изображенному на данной картинке. Такой алгоритм безошибочно выберёт в *любом* множестве картинок именно *ту самую* картинку, на которой проходило обучение, если она присутствует в тестовом массиве. Однако полезность данного алгоритма для дальнейшего использования стремится к нулю, так как персонаж будет опознан только *на одной* из всего множества возможных картинок и только на той, которая до пикселя совпадет с обучающим примером. Такой алгоритм будет «переобученным».

Аналогичным образом получается с точным алгоритмом решения обратной задачи. В случае наличия 100% точных и полных данных можно получить 100%-е соответствие функции исходного физического воздействия. Однако даже незначительное отклонение (ошибка) в данных измерений приводит к абсолютно иному результату восстановленной физической функции. Алгоритм становится «неустойчивым», 100%-й устойчивый алгоритм приводит *всегда* к одному и тому же результату вне зависимости от данных, но ценность такого алгоритма равна нулю. Соответственно задача заключается в поиске алгоритма достаточно устойчивого, исходя из качества имеющихся в распоряжении данных и восстанавливающего исходную функцию с максимально возможной (но не 100%-й) точностью в данных условиях (компромисс, как и в ИИ). Как правило, при решении обратных

задач это достигается введением изначальных ограничений на пространство решений, в котором осуществляется поиск данных решений, в первую очередь на размерность данного пространства.

Примером подобной обратной задачи может служить восстановление спектра размеров аэрозольных частиц по значению процента частиц, прошедших через n сеточек (данная величина имеет термин «проскок»). В науке об аэрозолях (атмосферных аэрозолях) данная методика широко применяется в так называемом методе «диффузионных батарей», так как осаждение аэрозольных частиц на сеточках осуществляется в результате диффузии частиц на поверхность сеточек. Измерительный прибор, используемый для разделения частиц на фракции, называется сетчатой диффузионной батареей, сокращенно СДБ. Автором применялась данная методика при изучении атмосферных процессов [Bashurova et al., 1992]. Также в ряде работ показано влияние точности (определенности) исходных данных (величины проскоков) на количество независимых параметров в решении (искомой функции распределения) [Koutsenogii, 1993]. Сам математический аппарат довольно подробно описан в приведенных работах, однако ниже он будет кратко рассмотрен, так как придется определять функцию ошибки или достоверности алгоритма, а это трудно сделать без освещения используемых формул.

Величина концентрации атмосферных аэрозолей – одна из вариабельных характеристик атмосферы, по своему поведению схожая с погодными характеристиками, которые мы будем изучать как фактор, влияющий на сельскохозяйственную деятельность. В описанной выше задаче определялась характеристика натурального объекта (распределение аэрозольных частиц по размерам) *in vivo* с помощью неких инструментов, которые в процессе измерений подвергают изменениям сам объект измерения. Естественно, что в данном случае сразу возникает понятие об аппаратной функции измерения. Таким образом, как уже отмечено выше, *следствием любого измерения является результат взвешивания аппаратной функции используемого измерительного устройства с исходной искомой физической характеристикой объекта, подлежащей измерению в результате данного эксперимента.*

Из названного выше утверждения однозначно следует следующий вывод: *восстановление исходной величины из результатов эксперимента всегда сводится к решению той или иной обратной задачи.*

Формально в рамках предложенного подхода могут быть использованы аппаратные функции в виде ступенек, прямоугольников и др. Последние далеко не физические функции с первой производной в виде δ -функций. Во многих случаях для простоты интерпретации это допустимо, хотя и не всегда корректно. О чем идет речь? Возьмем задачу определения числа объектов в определенном диапазоне размеров. Например, имеем некий прибор с размером отсечения x . В идеале, воспользовавшись этим прибором, мы знаем число объектов с размером больше/меньше x и общее число объектов. Соответственно сразу можно определить процент объектов в заданном интервале размеров. Однако на практике это неосуществимо, так как всегда есть вероятность преодоления барьера отсечения частицами иного размера, и наоборот. Фактическая (физическая) аппаратная функция будет выглядеть, как некоторая кривая. Чем ближе данная кривая по форме к «ступеньке», тем меньше проблем с интерпретацией результатов. Чем более «размазанной» является аппаратная функция, тем сложнее будет интерпретировать результат, и тем выше будет зависимость результата от ошибок измерений.

На рис. 32 показано графическое изображение представления аппаратных функций в зависимости от параметра x , по которому происходит аппаратное отсечение (например, размер, или длина волны). Цифрой 1 обозначена «идеальная» аппаратная функция, которая позволяет решить поставленную задачу без сложного математического инструментария, но которой (аппаратной функции) на самом деле не бывает в природе. Цифрой 2 обозначена реальная аппаратная функция, похожая на ту, что бывает в действительности.

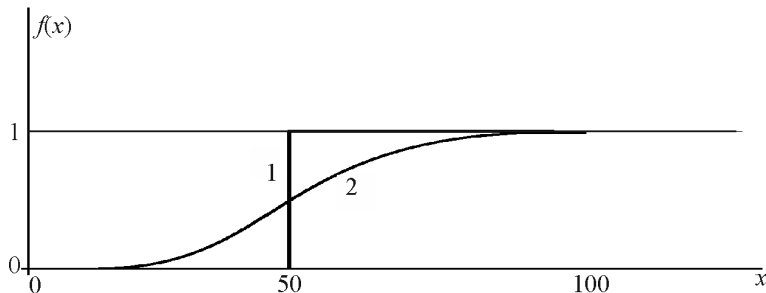


Рис. 32. Возможные аппаратные функции

Опираясь на сказанное выше, можно привести еще один факт из практического опыта решения обратных задач эксперимента, а именно: такие задачи при ближайшем рассмотрении оказываются некорректными, т.е. имеют неоднозначные (множественные) решения. Что это значит на практике? Часто целый набор значений исходных искомым параметров может давать результаты, близкие к наблюдаемым в эксперименте, по крайней мере в интервалах точности определения экспериментальных данных. Отсюда следует, что задача тем более определена, чем выше определенность измеряемых значений и чем круче аппаратная функция измерительных устройств (чем ближе производная этой функции к δ -функции).

Если задача является недоопределенной, то должен существовать целый класс решений исходной искомой величины, правильно описывающий наблюдаемые экспериментальные данные. Как будет показано ниже, в большинстве случаев определение спектра размеров атмосферного аэрозоля (что может также относиться к так называемому «генератору погоды») сводится к некорректной обратной задаче. Автор в работе [Bashurova et al., 1992] призывает использовать вместо выражения *спектр размеров* выражение *функция спектра размеров* (или понимать это по умолчанию). Данное определение близко к введенному нами понятию «образ», так как в отсутствие реальных данных (не можем потрогать, а восстанавливаем по косвенным признакам реакции прибора) нужна та функция, которая позволит использовать ее для определенных целей. Функция спектра размеров, получаемая из экспериментальных данных, обладает правом иметь мало общего с некой гипотетической действительностью, зато она должна иметь достаточную прогностическую силу, чтобы быть используемой в модельных вычислениях там, где это необходимо для определения неких реальных физических параметров, которые впоследствии так же могут контролироваться экспериментом.

Метод так называемых сетчатых диффузионных багарей (СДБ) часто используется при определении спектра размеров субмикронных частиц. При прохождении нормального потока аэрозоля через сетки аэрозольные частицы (АЧ) вследствие броуновской диффузии могут достигать поверхности волокон сетки и удаляться из потока, оставаясь на поверхности сеток. Величина эффективности захвата зависит

от размера частиц и используется в СДБ для разделения аэрозольных частиц для получения спектра размеров. Малые частицы вследствие большего коэффициента диффузии более эффективно удаляются из потока, чем большие.

После разделения число частиц, прошедших через известное число сеток СДБ, определяется счетчиком частиц и является численной безразмерной величиной (штуки частиц). Данные СДБ представляются в форме так называемых проскоков. Проскоком называется отношение числа частиц, прошедших через заданное число сеток (один из каналов СДБ), к полному числу частиц в начальном потоке (0-ой канал). Математическая проблема заключается в восстановлении спектра размеров АЧ из проскоков, получаемых с помощью СДБ.

В зарубежной литературе эта проблема получила широкое освещение. Здесь приводим названия в английской транскрипции некоторых из известных алгоритмов:

- graphical stripping [Sinclair, 1972; Sinclair et al., 1979];
- regularization [Тихонов и Арсенин, 1979; Crump, Seinfeld, 1981];
- nonlinear iterations [Cheng, Yeh, 1984];
- nonlinear programming using physical constraints [Cooper, Spielman, 1976];
- fitting [Helsper et al., 1982; Dzubay, Hasan, 1990].

Данные методы используются как при обработке данных СДБ, так и при восстановлении массового спектра размеров аэрозольных частиц при использовании каскадных импакторов [Dzubay, Hasan, 1990; Winklmayr et al., 1990] и при определении счетного спектра размеров аэрозольных частиц при использовании электростатического классификатора аэрозоля [Helsper et al., 1982]. Те же методы могут быть использованы для восстановления спектра длин волн по интенсивности электромагнитного излучения, прошедшего через оптические фильтры, а также для любой обратной задачи восстановления исходной физической функции по данным воздействия на эту функцию ограничителей с известной аппаратной функцией.

Большинство из перечисленных выше методов дают осциллирующие решения. Это означает, что при поиске решения изначально наложено ограничение на класс функций, среди множества которых осуществляется поиск таких решений. С точки зрения логики, дан-

ный подход вовсе не лишен смысла, так как физики всегда пытаются разложить реальные физические функции по «модам» или «гармоникам», т.е. пытаться выразить в виде ряда гармонических (осциллирующих) функций. Однако метод регуляризации лишен этого недостатка, поскольку решение изначально ищется в классе достаточно гладких функций. Метод подгонки дает хорошие результаты, однако часто возникают трудности с обоснованием выбора решения в форме той или иной аналитической функции. Кроме того, различные нулевые приближения приводят к разным конечным результатам вычислений.

Рассмотрим алгоритм, основанный на методе регуляризации Тихонова [Тихонов и Арсенин, 1979; Bashurova et al., 1992]. Выражение, описывающее проскок монодисперсного аэрозоля радиусом менее 0,1 микрон (когда диффузионный член является определяющим) через m сеток, выглядит следующим образом [Cheng, Yeh, 1984]:

$$K(m, x) = \exp(-C \cdot m \cdot D(r)^{2/3}), \quad (2)$$

где $x = \ln(r)$, r – радиус частиц; C – константа, зависящая от скорости потока аэрозоля и параметров сеток; $D(r)$ – коэффициент диффузии.

Кривые проскоков для различных значений m показаны на рис. 33. Это и есть аппаратные функции для данной задачи.

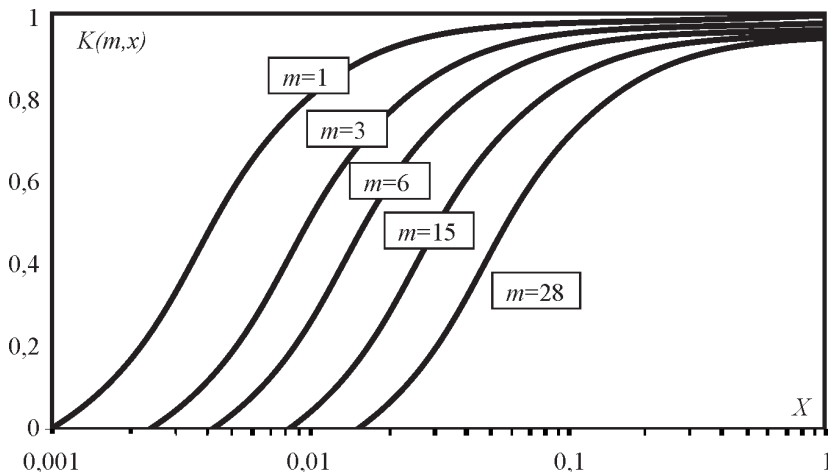


Рис. 33. Аппаратные функции для каждого из каналов m в зависимости от размера x

Для частиц радиусом более 0,1 микрон не только диффузия, но и импакция на волокнах сетки начинает играть значительную роль. Однако число таких частиц по счету в реальных атмосферных условиях невелико, и они не оказывают существенного влияния на измеренные значения проскоков, поэтому не следует распространять решение обратной задачи для СДБ на область размеров более 0,1 микрон. Таким образом, если принимать во внимание «физику» процесса, область применимости предлагаемого алгоритма для решения обратной задачи ограничивается. Данный пример показывает область критической важности экспертного вмешательства в чисто математический процесс решения обратной задачи. Например, именно базирываясь на экспертном мнении, допустимо вносить ограничения в область определения или пространство, на котором будут получены допустимые решения.

Для полидисперсного аэрозоля значение проскока $P(m)$ может быть выражено через нормированный спектр размеров аэрозоля $f(x)$, что является искомой физической величиной по аналогии со спектром электромагнитного излучения. В этом случае проскок полидисперсных аэрозольных частиц через m сеток записывается так же, как и для оптических характеристик через «взвешивание» искомой функции с аппаратной функцией так, как это было предложено выше по формуле (1):

$$P(m) = \int K(m, x)f(x)dx.$$

«Проскок» в данном случае – исходная и искомая непрерывная физическая величина (функция), взвешенная с аппаратной функцией измерительного прибора (или измерительной методики), в результате получается выражение для описания результатов эксперимента. При этом результаты экспериментов могут быть получены экспериментальным путем (увидены наблюдателем), а в результате математической обработки необходимо получить исходную функцию, которую на самом деле никто и никогда не видит впрямую. Чтобы получить $f(x)$ из проскоков $P(m)$, необходимо минимизировать функционал:

$$\|P(m) = \int K(m, x)f(x)dx\| \rightarrow \min. \quad (3)$$

Данное выражение означает, что функция $f(x)$, минимизирующая норму, рассматривается как решение уравнения (3). Выражение (3) необходимо заставить [Тихонов и Арсенин, 1979]:

$$\| P(m) = \int K(m, x)f(x)dx \| + \delta S[f] \rightarrow \min, \quad (4)$$

используя стабилизатор $S[f]$ и положительный параметр регуляризации δ . Как будет показано ниже, с математической точки зрения мы использовали манипуляцию с диагональю «решающей матрицы». Данный подход применяется довольно часто в различных алгоритмах, приводящих к использованию матричных операторов, например при использовании алгоритма, известного как нейронная сеть Хопфилда [Hopfield, 1982]. Отличие процедуры регуляризации в усилении диагонального элемента, чтобы избежать в ходе реализации математического алгоритма необходимости инвертирования потенциально вырожденной матрицы.

В методе регуляризации Тихонова часто используется стабилизатор второго порядка, обеспечивающий гладкость получаемых решений:

$$S[f] = \int \frac{d^2 f}{dx^2} dx. \quad (5)$$

Норма в уравнении (4) была определена как

$$\| P(m_i) - \int K(m_i, x)f(x)dx \|^2 = \sum_{m=1}^M [P(m_i, x) - \int_a^b K(m_i, x)f(x)dx]^2, \quad (6)$$

где $[a, b]$ – интервал, на котором ищется решение; $P(m_i)$ – набор измеренных проскоков; m_i – число сеток в i -м канале СДБ (ключевой параметр каждого из измерений); M – полное число точек измерения.

Если решение ищется в точках,

$$x_j = a + d \cdot (j-1/2),$$

где $d = (b-a)/J; j = 1, \dots, J$,

то решение уравнения (4) эквивалентно решению системы линейных уравнений [Тихонов и Арсенин, 1979]:

$$(A^T A + (\delta/d^2)Q^2)f = A^T p, \quad (5)$$

где A^T – транспонированная матрица A с элементами:

$$A_{ij} = K(m_p x_j) * d, \quad (6)$$

где f – вектор с компонентами:

$$f_i = f(x_i);$$

p – вектор с компонентами:

$$p_j = P(m_j) \text{ и матрица};$$

Q – конечноразностный оператор второго порядка с 0-ми элементами, кроме:

$$Q_{jj} = -2; Q_{jj+1} = Q_{jj-1} = 1.$$

Подгонка наилучших значений параметра регуляризации d и интервала решения $[a, b]$ производилась из условия минимизации параметра $Discr$:

$$Discr = \frac{|p - Af|}{\sqrt{M}}, \quad (7)$$

где M – число проскоков (точек измерения). Данный параметр служит для определения качества получаемых решений. Этот параметр показывает, насколько хорошо полученное решение удовлетворяет исходным данным.

В работе [Bashurova et al., 1992] использовалась аналитическая функция для аппроксимации коэффициента диффузии $D(r)$, применяемая в качестве основной функции, описывающей физический процесс. Для других примеров обратных задач может быть использована иная функция, достаточно точно описывающая процесс в заданном интервале решений. Если в подобного рода функции существуют коэффициенты, требующие дополнительного уточнения, проводятся калибровочные эксперименты или статистически корректно интерпретируются результаты натурных опытов.

Чтобы проверить возможность применения описанного выше алгоритма к решению обратной задачи СДБ, автором проводились численные эксперименты с использованием пробных функций. Спектр размеров был задан в форме суммы логнормальных распределений в соответствии с формулой. Логнормальные распределения характеризуется суммой мод, для каждой из которых задается характерный размер, амплитуда и ширина, – три независимых параметра.

Для данных целей была составлена программа, симулирующая измерения. Пользователю необходимо задать параметры спектра размеров: число аэрозольных мод и параметры каждой моды для получения симулированных «результатов измерений». Одиннадцать значений проскоков СДБ вычислялись по формуле (2), затем 2%-я случайная ошибка добавлялась к вычисленным значениям. Из полученных таким путем значений затем восстанавливался обратно исходный спектр размеров методом регуляризации Тихонова.

Были проведены эксперименты по восстановлению одно-, двух-, и трёхмодальных спектров размеров. Примеры для результатов вычислений представлены на рис. 34. Исходные спектры размеров, использованные для симуляции экспериментальных результатов, показаны сплошной линией, восстановленные спектры размеров – обозначены кружочками.

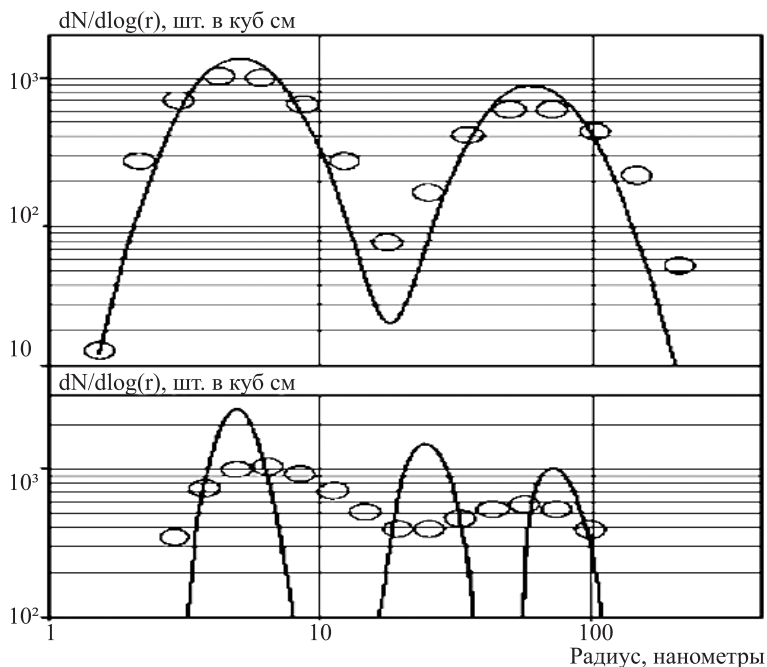


Рис. 34. Восстановление двух- и трехмодальных распределений по размерам. Сплошные линии – исходные данные. Кружки – восстановленные функции после численной имитации процесса измерений

Для одномодальных спектров размеров (не показаны на рисунке) восстановленные спектры размеров были очень близки к исходным. В этом случае проблем с восстановлением данных в результате решения обратной задачи не возникало.

В случае бимодальных распределений (см. рис. 34, верхний график) восстановленная функция достаточно хорошо воспроизводила исходную форму. Модальные радиусы воспроизводились с достаточно высокой точностью, однако ширины мод были несколько увеличены по сравнению с исходными. Это объясняется тем, что решение системы с введенной регуляризацией всегда стремится к более гладкой функции: решения ищутся в пространстве наиболее гладких функций из удовлетворяющих условию прямой задачи с допустимой точностью.

Успешное восстановление одно- и двухмодальных распределений с помощью описанной процедуры уже представлено в литературе [Bashurova и др., 1992]. Как и в более ранних работах, авторам не удалось разрешить трехмодальный спектр (см. рис. 34, нижний график). В данном случае при восстановлении исходное распределение заменялось широким бимодальным.

Трехмодовая логнормальная функция определяется девятью параметрами: N_i (концентрация, или амплитуда моды) σ_i (ширина моды), R_i (характерный радиус моды), где $i = 1, \dots, 3$. Причина того, что трехмодальное распределение не разрешается при решении обратной задачи СДБ, – набор измеренных проскоков СДБ на самом деле обеспечивает менее девяти независимых данных, хотя число самих проскоков может быть больше девяти. Данный пример – частный случай довольно общей проблемы использования гладких аппаратных функций.

Поиск минимального числа независимых исходных данных в общем массиве данных является довольно актуальной задачей, используемой в алгоритмах ИИ и машинного обучения. Данная задача решается при создании программ, сжимающих большие данные, например изображения. С точки зрения прогнозирования результатов сельскохозяйственной деятельности, имея в распоряжении большой массив данных погодных наблюдений в данной местности, необходимо учитывать насколько независимыми являются данные в массиве. Например, имеется две величины среднесуточных температур, при этом на календарной шкале сутки или одни после, или в коротком временном

интервале нескольких дней. Известно, что в один из дней наблюдалась температура -20°C . Наблюдатель вправе сделать вывод, что данные сутки относятся к зимнему периоду времени (если наблюдения проводятся в зоне континентального климата средних широт Северного полушария). Хотя полный интервал наблюдаемых температур в данной местности задан от -30 до 30°C , наблюдение положительных температур в близком временном интервале от -20°C является маловероятным. Таким образом, две величины связаны некими ограничениями и не являются 100%-ми независимыми данными.

Возвращаемся к описываемому примеру решения обратной задачи: чтобы восстановить спектр размеров аэрозоля из данных СДБ в общем виде, необходимо решить систему интегральных уравнений Фредгольма первого порядка (выражение (3)). Находя решение в точках x ($x = \ln(r)$) и заменяя интеграл на сумму, система (3) может быть сведена к системе линейных уравнений:

$$P(m) = \sum_j K(m, r(x_j)) f(x_j) \cdot d, \quad (8)$$

где $d = x_{j+1} - x_j$;

или в матричной форме: $p = A \cdot f$, (9)

где p – вектор с компонентами $p_i = P(m_i)$; f – вектор с компонентами $f_j = f(x_j)$; A – матрица с элементами:

$$A_{ij} = K(m_i, r(x_j)) \cdot d.$$

Чтобы получить решение обратной задачи в точках x , общее число которых отличается от числа измеренных проскоков, можно использовать следующее уравнение:

$$f = (A^T \cdot A)^{-1} \cdot A^T \cdot p, \quad (10)$$

где A^T – транспонированная матрица A (свелось к уравнению (5) без стабилизирующего члена).

К сожалению, решение $f(x)$ интегральных уравнений (3) чрезвычайно чувствительно к точности определения левой части $P(m)$ [Bashurova и et al., 1992] вследствие пологости аппаратных функций СДБ (смотри рис. 2). На практике это означает, что существует беско-

нечное число функций $f(x)$, которые при подстановке в уравнение (3) приводят к левой части $P(m)$, отличающейся от экспериментальных значений на величину меньшую, чем точность проведения измерений.

Данное утверждение по большому счету верно для любой обратной задачи, а также для задачи прогнозирования. Для задачи прогнозирования постулат можно перефразировать следующим образом: существует бесконечное множество вариантов исходных параметров, приводящих к одинаковому итоговому результату. Например, пусть имеется ряд идеальных погодных условий для получения оптимального урожая определенной культуры на конкретном рабочем участке. Возникает вопрос: при каких условиях урожай данной культуры может быть не более 75% от максимального? Вариантов условий, приводящих к тому же результату великое множество, если не бесконечно. Сыграть роль может как недостаточные температуры, так и недостаточная влажность, а может быть и экстремальное климатическое событие, например однократный поздний заморозок.

Обращение становится нестабильным, если пытаться получить решение $f(x)$, определяемое большим числом независимых параметров, чем в действительности обеспечивает левая часть $P(m)$ [Cooper, Wu, 1990]. При использовании метода регуляризации Тихонова решение может быть определено в точках, число которых неограниченно велико, однако значения функции в данных точках не будут являться независимыми. Стабильность обращения в данном случае определяется дополнительным регуляризационным членом $(\delta/d) \cdot Q$ в уравнении (4) по сравнению с уравнением (3).

Описанная нестабильность при обращении задачи показывает, что число физически независимых строк или столбцов матрицы A (или квадратной матрицы $A^T A$) меньше числа точек в решении x . В данном случае физическая линейная зависимость означает, что строка (столбец) матрицы может быть приближена линейной комбинацией строк (столбцов) с точностью, превышающей точность, обеспечиваемую исходными данными. В дальнейшем будет дан ответ, сколько независимых значений функции распределения частиц по размерам можно определить с помощью одного набора проскоков СДБ, или в общем виде, как влияет «гладкость» аппаратной функции на возможное число независимых параметров решения.

Число независимых значений в одном наборе проскоков СДБ было определено как эмпирически (в описанных выше имитационных экспериментах), изучая стабильность процедуры обращения данных СДБ без регуляризации для различного числа точек в решении, так и теоретически.

Для эмпирического на основе результатов наблюдений физических параметров (сходных с параметрами показателей погоды) составлена программа для прямого (без регуляризационного члена) обращения данных на базе уравнения (3). Решение искалось в J точках. Отрицательные значения $f(x)$ заменялись на 0. Качество полученного решения определялось параметром $Discr$ (выражение (7)). Для восстановления спектров использовались наборы проскоков, реально полученные во время реальных натуральных измерений.

На рис. 35 показана зависимость минимального параметра $Discr$ от числа точек в решении J для некоторых из обработанных спектров, измеренных в реальных натуральных экспериментах. Если $J > 6$, решения подвержены сильной осцилляции, плохо удовлетворяя исходным экспериментальным данным ($Discr > 100\%$). При уменьшении J до 6 или

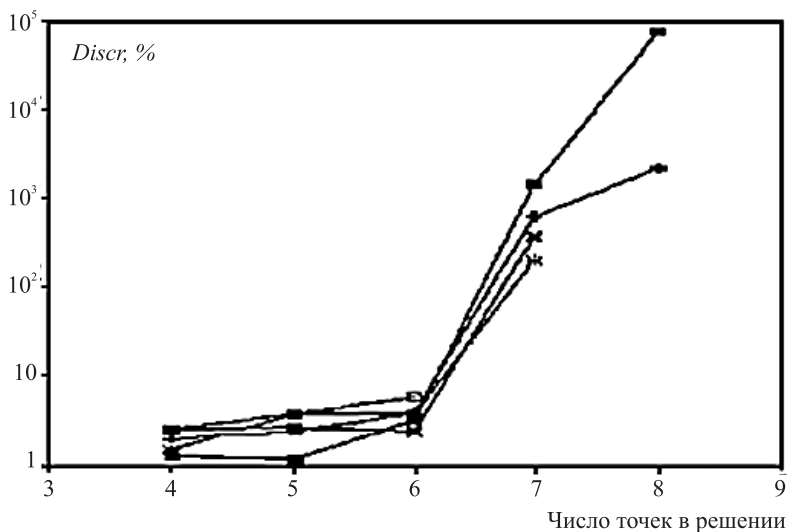


Рис. 35. Зависимость устойчивости параметра $Discr$ от числа точек (независимых параметров) в решении обратной задачи восстановления распределения размеров. Число входных параметров (каналов $m = 11$)

менее 6 можно получить вполне приемлемые решения. Это означает, что, используя применяемую СДБ (марка TSI 3040), можно получить лишь 6 или менее независимых данных или значений проскоков. При этом в данном конкретном изделии число каналов m равнялось 11. Непосвященному наблюдателю должно казаться, что на входе задачи мы имеем 11 независимых значений, значит, ожидаемое стабильное решение может содержать 11 параметров. Однако по факту это не так. Из 11 проскоков в реальности можно извлечь не более 6 независимых параметров.

Решения, полученные прямым обращением, сравнивались с решениями для того же набора проскоков, полученными методом регуляризации Тихонова. Решение с регуляризацией ($Discr = 2,8\%$) имело сходную форму с решением без регуляризации, но с меньшим числом точек.

5.4. Понятие «достоверности и независимости данных». **Применимость образа в зависимости** **от специфики решаемой задачи**

Очень трудно определить, какое из бесконечного числа возможных решений – «истинный» спектр размеров аэрозоля. Любой спектр размеров, когда-либо опубликованный в литературе, является как функцией измеряемого аэрозоля, так и функцией измерительного устройства и процедуры обработки данных. Следовательно, имеет смысл сравнивать результаты, полученные одними и теми же или сходными измерительными системами. Это относится не только к аэрозолям или атмосферным параметрам, но и к любым измерениям, требующим решения обратной задачи и сложной интерпретации.

С данной точки зрения, результаты восстановления трехмодового распределения, приведенные на рис. 34 (нижний график), не могут быть рассмотрены как провалившийся эксперимент. Данный численный эксперимент показывает ограничения на разрешающую способность СДБ, связанную с формой ее аппаратной функции, с неточностью в определении счетной концентрации аэрозоля (неопределенность исходных данных) и на то, как следует интерпретировать результаты измерений. Например, если восстановленный спектр размеров одномодален, можно утверждать с достаточной определенно-

стью, что истинное распределение по размерам является также одномодальным. Если получен бимодальный спектр, то можно лишь утверждать, что исходное распределение мультимодально с неизвестным числом мод. Например, во многих работах используется в основном термин бимодальное распределение, а не мультимодальное (хотя именно это и подразумевается), поскольку рассматриваются лишь решения конкретной обратной задачи для СДБ.

Неопределенность исходных данных применительно к погодным данным означает не столько неточность измерений или данных, полученных с приборов измерений, сколько вариативность в определенных интервалах. Это связано с естественными флуктуациями наблюдаемых значений, обусловленных постоянными изменениями погодных условий. Большинство используемых климатических параметров в моделях – результат усреднения или по временной шкале, или в пределах определенной территории (площади). Любая средняя или «усредненная» величина характеризуется, как минимум, двумя параметрами: средним значением и, например, «стандартным отклонением» в зависимости от используемой процедуры усреднения. Последняя величина характеризует диапазон флуктуаций значения на области определения (например, временном интервале), на которой ищется среднее.

Единственная возможность определить истинность решения обратной задачи – посмотреть, насколько данное решение удовлетворяет исходным экспериментальным данным. Параметр *Discr* (выражение (7)) может быть использован для данной цели для описанной задачи восстановления исходных параметров атмосферного аэрозоля. Все решения, которые удовлетворяют экспериментальным данным в пределах коридора ошибок и неопределенностей, имеют равное право рассматриваться в качестве истинных решений. Число таких решений может быть неограниченно.

Выбор нужного решения в каждом случае определяется конкретной необходимостью. Если качество исходных данных допускает существование решения в широком диапазоне классов функций, то предпочтительней будет использовать решения из класса наиболее простых функций, определяемых минимальным числом параметров, достаточных для определения исходных характеристик объекта, что-

бы выбранные решения при использовании в прямой задаче приводили к данным с удовлетворительной точностью в допустимых интервалах неопределенности. Однако применительно к описанной задаче с атмосферными аэрозолями более наглядным может быть более сложное бимодальное распределение, так как позволяет «визуализовать» физический процесс перекачки аэрозольной массы из мелкой аэрозольной фракции в более крупную вследствие коагуляции. Следовательно, выбор класса функций решения зависит еще и от исходной цели поставленной задачи.

Рассмотрим, как, анализируя вид аппаратной функции и качество (особенности) используемых данных, можно получить представление о минимальном числе независимых параметров, которое может быть получено в результате решения обратной задачи.

Конкретно для задачи СДБ достоверность или интервалы неопределенности для данных определяются достоверностью определения счетной концентрации аэрозольных частиц после разделения аэрозоля на фракции с помощью СДБ. Приборная ошибка в измерениях составляет несколько процентов. Однако в природе счетная концентрация атмосферного аэрозоля претерпевает значительные флуктуации во времени. Данные флуктуации приводят к значительным неопределенностям в определении аэрозольной концентрации, которые и дают основной вклад в неопределенность значения концентраций и, следовательно, проскоков. Из-за нестабильности процедуры решения обратной задачи погрешность в определении проскоков трудно пересчитать в погрешность определения спектра размеров. Например, используя данные атмосферных измерений, для которых 10%-я точность считается очень высокой, можно получать одно- и двухмодальные спектры размеров из одного и того же набора данных, что было показано численным моделированием.

Для того чтобы наглядно продемонстрировать, сколько независимых параметров может содержать решение в зависимости от аппаратной функции и погрешности в определении данных, рассмотрим графическое представление аппаратных функций СДБ. Относительная величина проскоков, которая определяет спектр размеров, изменяется в интервале от 0 до 1. Аппаратные функции СДБ (выражение 2) простираются через весь диапазон размеров и являются почти плоскими

(см. рис. 33, следует учитывать логарифмический масштаб представления). Для получения величин проскоков эти функции интегрируются с множителем $f(x)$ – функцией спектра размеров частиц. Поскольку в лучшем случае неопределенность в определении $f(x)$ составляет 15–30% (что определяется вариабельностью концентрации АЧ во времени), то физически неразличимых значений проскоков может быть только от 3 до 6. Это число возможных неперекрывающихся интервалов 0,15–0,30 в интервале 0–1. Это интервал определения функции по оси Y (в рассматриваемой задаче).

Применительно к моделям ИИ также необходимо определять размерности искомых решений, отличные от размерности массивов исходных данных. В настоящее время эти подходы используются для создания «кодеров», сжимающих данные, например для целей архивирования. Одной из задач, решаемых в этом случае, является определение минимального числа независимых переменных в образе, необходимых для полноценного восстановления из образа начальных данных.

Приведем пример того, как это работает. Сначала имеется массив обучающих данных, где набор определенных исходных данных поставлен в соответствие результату. Следует отметить, что исходные данные – это данные измерений и натуральных экспериментов, и никто изначально не знает, насколько они независимы друг от друга. Они могут не быть «независимыми», что было подробно показано на примере анализа данных по таким атмосферным явлениям, как атмосферный аэрозоль или температура воздуха.

Тот факт, что при анализе зависимости биологических процессов, таких как выращивание растений, в зависимости от физических параметров, таких как погодные условия, приходится опираться на данные, которые физически являются не независимыми друг от друга, обуславливает логичность использования таких подходов, как Байесовская сеть. Еще раз подчеркнем, что данные формально независимых экспериментов (то есть разных измерений) могут оказаться в зависимости друг от друга в физическом смысле, как описано в примере со среднесуточными температурами. По факту мы имеем дело с ситуацией, когда необходимо определить вероятность какого-либо события при условии, что произошло другое статистически взаимосвя-

висимое с ним событие. Это напрямую случай, описываемый теоремой Байеса. Данный вывод указывает на потенциальное направление развития подходов к решению задач сельскохозяйственного прогнозирования.

Пример оптимизации пространства размерности данных, т.е. минимального числа независимых переменных, необходимых для полного описания задачи, – это то, как в IT работают кодировщики и архиваторы, используемые для сжатия данных с целью экономии места их хранения в вычислительном пространстве. Один из подходов применения метода машинного обучения к созданию кодировщика заключается в следующем. Берется достаточно крупный обучающий массив примеров. Например, альбом очень большого числа фотографий реальных объектов. Все данные оцифровываются. Каждой фотографии соответствует цифровой массив в определенном формате. Затем происходит кодировка, ставящая каждой фотографии в соответствие некий образ в N -мерном пространстве (вектор). Аналогом этого процесса является восстановление образа методом фоторобота. Исходную фотографию можно оцифровывать по пикселям (точкам), где каждой точке ставится в соответствие яркость, цвет и др. Однако это далеко не исчерпывающий набор признаков, позволяющий довольно точно восстановить образ (например, портрет человека), запечатленный на фотографии. Если подходить к задаче описания портрета, можно сформулировать абсолютно иной набор признаков и их числовых значений, приводящий к точному распознаванию образа. Например, лицо круглое или овальное, волосы темные или светлые, прямые или кудрявые, нос прямой, курносый или с горбинкой и др. Число используемых признаков – это размерность пространства. Набор значений каждого из признаков, соответствующего конкретной картинке, – образ или «вектор» в заданном пространстве образов, кодирующий искомую картинку.

Следующий этап создания кодера методом машинного обучения – тренировка модели, которая ставит каждому объекту в соответствие образ в N -мерном пространстве, а затем восстанавливает данный образ в некую картинку, которая должна максимально соответствовать исходной картинке в исходной метрике. Разница между восстановленной картинкой и исходной в оригинальной метрике, в которой

происходила оцифровка, – та функция ошибки, которая минимизируется на большом числе обучающих примеров, настраивая веса внутри программы кодера. Когда достигается требуемая точность процесса сжатия – восстановления данных, можно использовать полученный результат по-разному. Можно отрезать часть программы, которая занималась кодированием, и использовать ее в дальнейшем для сжатия данных. Однако нас интересует другая возможность. Можно начать уменьшать размерность N и смотреть, начиная с какой величины не удастся восстановить искомый образ с требуемой точностью. Таким образом можно установить минимальное количество независимых переменных, требуемых для описания любого объекта из пространства образов, или по-другому: для описания реально изучаемого объекта.

В приведенном выше примере решения обратной задачи для восстановления распределения по размерам получалось, что функция, заданная 4–6 параметрами хорошо справлялась с аппроксимацией реального физического явления. Для функций с меньшим числом степеней свободы получать точное решение не удавалось, для функций с большим числом степеней свободы решение становилось нестабильным.

Данные численные эксперименты с реальными физическими данными (см. рис. 34 и 35) является иллюстрацией того, что при попытке получения решения или «образа» с детализацией большей, чем позволяют исходные данные, алгоритм теряет стабильность. Единственный способ стабилизировать ситуацию – внесение дополнительных ограничений на пространство возможных решений, что происходит при использовании алгоритма «регуляризации». Фактически решение ищется среди достаточно гладких функций. В переводе с математического на общеполитический язык данный подход можно характеризовать, как вариант «принципа бритвы Оккама». В качестве решения обратной задачи может быть использована самая простая функция из возможных, которая может быть подставлена в условия прямой задачи и даст расхождения с исходными данными на уровне допустимой ошибки или неопределенности.

Этот же принцип легко применить к понятию образа. Образ – это некое модельное представление будущего. Однако если данный образ максимально соответствует (не противоречит) сегодняшним данным

и данным, накопленным за предыдущие периоды, можно принять его на сегодняшний момент и с его помощью пытаться оценить интересующие нас параметры (значения) в будущих периодах. После наступления «будущего» возможно сравнить предсказанные значения с фактом. Если отклонения находятся в пределах ожидаемых интервалов, это будет подтверждением того, что используемый образ может быть использован и в дальнейшем. Если же расхождения окажутся более существенными, то выводов может быть, как минимум, два: мы имеем завышенные ожидания, и опора на предыдущие знания не позволяет создать ожидаемый образ будущих периодов, поэтому необходимо снизить ожидания от точности предсказаний (расширить ожидаемые интервалы); необходимы изменения в самом образе (дополнительная тренировка с использованием большего числа данных).

Еще одной визуализацией применения подхода, связанного с наличием виртуальной аппаратной функции применительно к сельскохозяйственной деятельности и связанной с простыми примерами из жизни, является интерпретация так называемых примет. В упрощенном виде приметы выглядят следующим образом: если большой урожай рябины – ожидается суровая зима, если черемуха зацвела в определенный период – ожидается дождливое лето и др. Все это достаточно в общем виде сформулированные утверждения на основе многолетних наблюдений. На самом деле, значимых примет может быть всего несколько, но иногда с их помощью получают неожиданно точные прогнозы. Как это может работать? Большой или маленький урожай рябины – это результат воздействия ряда природно-климатических факторов в течение сельскохозяйственного сезона на рябиновое дерево. В итоге имеем дискретный (бинарный) результат – плохой/хороший урожай. В данном случае рябина – это измерительный прибор, на который действуют многочисленные природно-климатические факторы в течение длительного периода времени. Результат измерения прибора – урожай, а биология рябинового дерева и есть та самая аппаратная функция, которая будучи проинтегрированной (подверженной воздействию) со всей совокупностью внешних факторов выдает урожай как интегральный результат. В данном случае функция является сильно закругляющей (сглаживающей), так как бесконечное количество вариантов исходных данных будут приводить к одному и тому

же результату. Тем не менее, функция позволяет выделить важную комбинацию климатических факторов, непосредственно связанных именно с продуктивностью биологического объекта в данном сельскохозяйственном сезоне, и дает определенный прогноз на будущий сезон. Собрав необходимое и возможное количество таких «индикаторов», возможно генерировать прогнозы с допустимой степенью достоверности.

5.5. Перспективы применения описанных подходов к задаче прогноза урожайности сельскохозяйственных культур

Для наглядности предлагается рассмотреть в общем виде нашу задачу прогнозирования потенциальной урожайности (обусловленной климатическими ограничениями) на заданной территории исходя из данных погодных наблюдений. Погодные наблюдения – это ряды цифр, собранных с измерительных приборов и отражающих понятные параметры: температуру воздуха, влажность (относительную), осадки (в мм), скорость ветра, облачность и др. Для стороннего наблюдателя очевидно, что данный набор параметров включает в себя далеко не независимые параметры. Так, при прочих равных условиях в случае меньшей облачности следует ожидать более высокую температуру воздуха вследствие нагрева почвы солнечной радиацией, так же облачность коррелирует с осадками и т.д. К тому же данные в течение суток A коррелируют с данными в течение суток $A + 1$. Если в день A выпало большое количество осадков при положительных температурах, логично ожидать повышенную влажность воздуха в течение следующих нескольких суток. Сложно ожидать положительные температуры в январе и отрицательные температуры в июле, если измерения проводятся на юге Западной Сибири. Из этого примера также видно, что данные, собранные на географически близко расположенных метеостанциях, в строгом физическом смысле не являются независимыми.

Так сколько же независимых данных может содержать массив данных по X метеостанциям, если мы имеем измерения Y среднесуточных параметров в течение года? Формально в нашем распоряжении массив из $N = X \cdot Y \cdot 365$ чисел.

Каждое из чисел получено в результате независимого измерительного эксперимента. С логической точки зрения, мы имеем массив из N числа независимых значений. Однако, как уже было показано выше, с точки зрения физического смысла данные значения не являются независимыми. Поскольку не существует формульной связи значений друг с другом, невозможно аналитически получить количество независимых параметров задачи или, иными словами, «степеней свободы» в изначальных данных. Такого рода задачи могут быть решены в рамках подходов, обозначенных как «ENCODING» – «DECODING» в процессе машинного обучения.

Проанализируем, как может быть использован данный подход применительно к нашей задаче. Пусть, имея массив погодных данных из N значений (или, как часто обозначают, «векторов»), мы хотим предсказать какое-то событие. Это событие можно классифицировать как «образ», процесс предсказания – это реконструкция образа. Нас интересуют важные параметры, влияющие на урожайность сельскохозяйственных культур в предстоящем сезоне. Набору исходных данных – погодные данные на предстоящий сезон – можно поставить в соответствие урожайность в текущем сезоне. Если удастся создать тренируемую модель ИИ, которая будет ставить в соответствие определенную картину погодных явлений соответствующим прогнозным ожиданиям урожайности, то следующим этапом может стать задача по компрессии исходных данных с сохранением достоверности прогноза.

Рассмотрим пример реализации данной стратегии. Генератор погоды – это аналог (в нашем рассмотрении) искомой функции. Понимая, что вместо искомой функции, описывающей физическое явление, нас устроит *любая* функция, которая будучи взвешенной с аппаратной функцией, даст вычислительный результат, совпадающий с результатом фактических измерений с допустимой точностью.

Как постулируют авторы генератора погоды, данный подход предполагает замену концепции «погода – урожай» на концепцию «климат – урожай». О чем идет речь? Генератор погоды выдает ряды погодных данных, имитирующих погодные явления на заданной территории и в заданный период времени, соответствующие климатической норме со статистически достоверными отклонениями от средних величин.

Погодная картинка, полученная с помощью генератора погоды, ни в коем случае не должна восприниматься как прогноз погоды, так как описание погодных явлений дается на промежуток времени, находящийся за горизонтом любых возможных достоверных временных интервалов предсказания погоды. Картину посуточных погодных явлений следует воспринимать как некий образ или функцию образа, соответствующую климату на описываемой территории.

Тем не менее, прогнозы результатов сельскохозяйственной деятельности с использованием генератора погоды могут быть достаточно точными. С чем это связано? Генератор погоды выдает погодные явления, интегральные характеристики которых соответствуют климатической норме заданной территории, или климату. Возникает вопрос: почему не брать просто средние значения за определенный период, соответствующие климатической норме? Ответ неочевиден. На биологические процессы влияет не только суммарные показатели климата, но и флуктуации (например, суточные) погодных характеристик. Именно данную функцию (суточные колебания характеристик погоды) позволяет смоделировать генератор погоды.

Используя генератор погоды в качестве источника данных о погоде, чтобы впоследствии тренировать модель нейронной сети на предсказание климатически обусловленной урожайности на исследуемой территории, можно получить следующие преимущества:

1. Непрерывность данных на *любом* заданном временном интервале.
2. Полноту и достоверность (не требующую проверки) данных также на любом заданном временном интервале.
3. Возможен более полный учёт данных по урожайности с учётом применявшихся агротехнологий в хозяйствах, так как выбор хозяйств не привязан к обязательному наличию для данного хозяйства релевантных погодных данных. Можно выбрать хозяйство или группу хозяйств из определенной климатической зоны, наиболее аккуратно ведущих свой внутривладельческий учет, и сгенерировать требуемые для обучения модели ряды погоды, соответствующие климатической зоне выбранных хозяйств.

Каков основной недостаток данного подхода? Модель *не будет* каким-либо образом учитывать возникновение экстремальных климатических явлений (поздние заморозки, засуха и др.).

Таким образом, существует возможность достаточного точного и полного (машинного) обучения модели (например, нейронной сети), которая после обучения будет адекватно рассчитывать возможную, обусловленную климатической зоной урожайность заданной культуры и давать адекватные рекомендации по выбору агротехнологий на предстоящий сельскохозяйственный период.

5.6. Пример вычислений с применением описанного подхода

Исходя из сказанного выше, авторами была предложена следующая модель для машинного обучения ИИ. В качестве исходных данных используется массив климатических (погодных) данных со станций метеонаблюдения на исследуемой территории. Для начала использовались ряды среднесуточных температур. Сеть метеостанций должна более или менее перекрывать географически исследуемый район. Так как изначально предполагается, что связь между наблюдаемыми погодными значениями за предыдущий период и благоприятность текущего периода не является установленной (иначе уже имелся бы эффективный прогнозный инструмент), то мы не можем предположить требуемую детализацию данных и необходимый период наблюдения для получения достаточных данных. В связи с этим в начальном приближении будем использовать весь массив погодных данных с посуточным разрешением за 1–5 лет перед прогнозным периодом.

Далее надо определиться с тем, какой результат должна предсказывать нейронная сеть. В пределах территории, покрытой сетью метеостанций, существуют данные статистики по урожайности основных культур по районам и годам. Наша задача – научиться предсказывать, урожайный или неурожайный год ожидается в том или ином районе для той или иной сельскохозяйственной культуры, исходя из данных метеонаблюдений на территории, в которую входит данный район, за предыдущие несколько лет. Можно брать среднюю урожайность в районе культуры за несколько лет (3–5 в применяемой тестовой задаче), предыдущих наблюдению (E_1), и сравнивать с урожайностью в год наблюдений (E_2). Отношение E_2/E_1 и будет искомой величиной прогноза. Пусть E величина прогноза, даваемая машинным алгоритмом. Таким образом, минимизируемый функционал будет ($E-E_2$), или

в безразмерном варианте $(E-E_2)/E_1$. В случае, если прогноз удался на 100% ($E = E_2$), функция будет равна 0. После того, как сеть научится давать прогноз с приемлемой точностью на массиве данных E_2 , можно начать получать значения E в качестве прогнозных, базируясь на погодных наблюдениях за предыдущий период. В результате будет прогноз ожидаемой урожайности культуры в районе в процентах от средней урожайности за предыдущие годы. Таким образом, можно предположить, насколько урожайным является ожидаемый сельскохозяйственный период по той или иной культуре, чтобы вовремя внести корректирующие мероприятия в агротехнику.

К сожалению, существует проблема наличия или получения доступа к полным и структурированным базам данных по погодным рядам и урожайностям за много лет в заданной географической области. Пример попытки использования имеющихся неполных рядов данных будет приведен несколько позже. Сейчас отметим, что наличие недостаточного количества обучающих данных приводит к необходимости использования для решения прогнозных задач гибридных моделей, использующих как данные, так и экспертные знания агрономической науки.

Вернёмся к примеру использования имеющего массива данных для целей машинного обучения прогноза урожайности культур по погодным данным предыдущего периода. Для тестирования метода машинного обучения применительно к рядам урожайности по статистическим данным были использованы временные ряды по погоде (1971–2019 гг.) и урожайности пшеницы, ячменя, овса и ржи (1971–2000 гг.) на территории Новосибирской области по имеющимся статистическим данным. Данные «Росгидромета» с 19 метеостанций, расположенных в Новосибирской области, взяты с сайта «Погода и климат» (<http://www.pogodaiklimat.ru/>). Данные официальной статистики по сборам основных зерновых культур по районам Новосибирской области предоставлены структурным подразделением СФНЦА РАН – СибНИИЭСХ. Для тренировки на имеющихся массивах данных использованы следующие численные модели машинного обучения:

- XGBClassifier;
- Random Forest Classifier.

Была высказана гипотеза возможности тренировки моделей на обучающих примерах, сформированных на базе имеющихся массивов данных, с целью в дальнейшем создать ансамбль этих двух моделей для наиболее точного прогнозирования. Имеющиеся данные показали недостаточную полноту для формирования необходимого количества обучающих примеров для корректной тренировки модели. Целью тренировки было научить модель предсказывать благоприятный или неблагоприятный год по ожидаемой урожайности для каждой из культур. Для этого вводился показатель «благоприятности». Определялась средняя урожайность культуры за несколько лет (от трех до пяти), предшествующих году наблюдений. Если урожайность в исследуемый год превышала среднюю за предыдущие годы, год относили к благоприятным, если была меньше средней, к неблагоприятным. Урожайность по году соотносилась с температурными профилями предыдущего года для поиска соответствия или ассоциации.

Для оценки эффективности тренировки модели введен показатель «ассигасу». Данный показатель определялся следующим образом. Выбраны 19 районов Новосибирской области с минимально необходимым числом комбинаций погода – урожайность по имеющимся культурам (выбраны данные по овсу и пшенице). В качестве основного погодного показателя использовались температурные ряды. Из части рядов сформированы обучающие примеры, часть рядов использовалась для проверочных примеров. Число обучающих и проверочных примеров старались задавать в соотношении 9 к 1. Для каждого из районов после обучения системы на обучающих примерах, сформированных из части данных, проводилась проверка на проверочном массиве. Параметр «ассигасу» определялся как число лет, для которых удалось правильно предсказать показатель благоприятности, к общему числу проверочных примеров. В дальнейшем показатель усреднялся по районам, таким образом, получался средний показатель за год.

При усреднении температурных показателей за предыдущие годы средний показатель точности «ассигасу» по тем районам, по которым удалось набрать обучающие примеры по годам, составил 0,67. Данная величина не отображает реальное положение дел, так как разброс показателя «ассигасу» по годам был достаточно велик, например, от

0,0 за 1984 г. до 0,89 за 1985 г. Так как погодные данные давались на вход модели в весьма усредненном виде, можно сказать, что данный вывод относится непосредственно к «году», т.е. есть годы урожайные, а есть неурожайные. Зависимость от объективных характеристик года составляет в среднем около 70%. Данный результат в значительной степени является очевидным и не противоречащим существующим представлениям, что косвенно подтверждает правильность выбранного подхода. При этом в части случаев предвидеть урожайность года можно заранее, в части – не получается, что тоже более или менее очевидный результат.

Факт неустойчивости результатов тренировки на проверочных данных может быть показателем того, что алгоритм не устойчив к изменению данных вследствие недостаточного числа обучающих примеров для тренировки. По результатам проведенного анализа данных методами машинного обучения высказана вторая рабочая гипотеза, что существуют различия по годам, когда зависимость урожайности от предшествующей погоды существенно увеличивается или уменьшается. Данный факт согласуется с фундаментальными представлениями в области биологии растений. Точный прогноз периодов такой зависимости может стать важным результатом для принятия управленческих решений специалистами агропредприятий. Однако дальнейшие исследования в данном направлении требуют кратного увеличения объема релевантной информации в распоряжении исследователей.

Был предложен и протестирован алгоритм извлечения дополнительной информации из существующих рядов данных с пропусками через увеличение числа обучающих примеров. Недостаток числа обучающих примеров вытекает из двух основных проблем: пропусках в данных по урожайности и данных по погодным рядам. Заполнять данные по урожайности на основе известной выборки не представляется возможным, так как это противоречит принципам работы с временными рядами, поэтому было решено заполнить пропуски по погодным данным (рис. 36). Для этого предложен численный алгоритм, аналогичный принципу «генератора погоды». Данный алгоритм позволяет создавать имитационную картину колебаний погоды в рамках климатической нормы без изменений средних параметров за длитель-

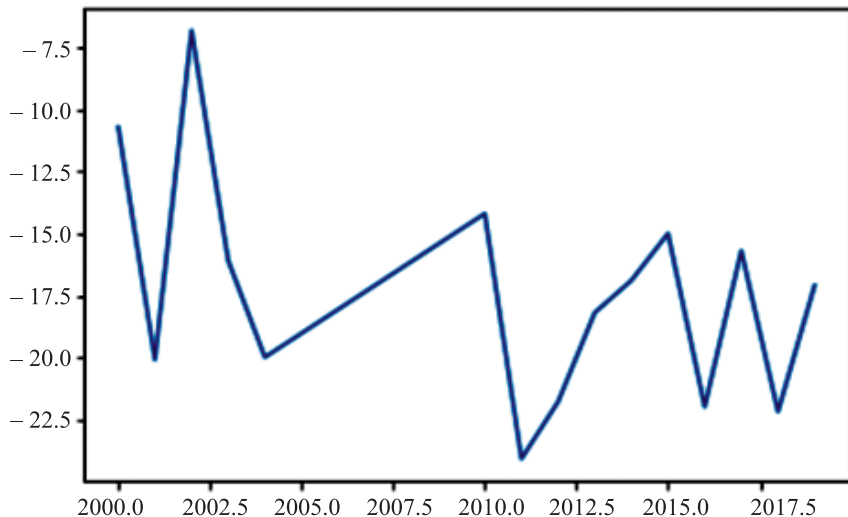


Рис. 36. Пример «восстановленного» годового профиля среднемесячной температуры января для Баганского района Новосибирской области

ный период, характеризующих климат. При заполнении пропущенных месяцев данными из построенного распределения для погоды на определенный месяц за счет существующих замеров в среднегодовой температуре окажется информация из реального мира, следовательно, больше действительных данных будет задействовано в процессе тренировки алгоритма, что может улучшить показатели используемых алгоритмов.

Выдвинута гипотеза о том, что среднемесячным температурам соответствует некоторое нормальное распределение. Для восстановления параметров нормального распределения использован метод бутстреп, построены доверительные интервалы для оценок этих параметров и взяты средние значения для «предсказания» погоды в пропущенных месяцах.

Подобные операции проведены для каждого района. Среднегодовая температура в таком случае – 2,05 °С, без учета новых данных – 2,1 °С. Используя модификации данных по погоде, а также уменьшив требования к семплу (в связи с тем, что связь урожайности находили по ассоциации с погодой на глубину не более двух предыдущих лет), получили 260 тренировочных и тестовых примеров, чтократно больш

ше по сравнению с условиями первых тренировок. Итоговая средняя точность прогноза по годам на модифицированных данных составила 0,58. Разброс от 0,3125 до 0,871.

Из полученных результатов сделано два вывода:

- при увеличении числа обучающих примеров исчезают годы с коэффициентом угадывания 0, что позволяет надеяться на дальнейший прогресс по мере появления возможности тренировки модели на большем объеме значимых данных;

- для дальнейшей тренировки модели необходимо найти способ получить более полные данные по урожайности, не прибегая к аппроксимациям, так как урожайность является объектом прогноза.

Впоследствии была высказана гипотеза, что дальнейшая детализация температурных профилей периодов, на которых строится прогноз, может привести к более точному образу, на его основании будет возможно осуществлять более стабильные предсказания. При этом имелось в виду, что существует предельная детализация образа, дальше которой ситуация может или стабилизироваться, или даже ухудшиться. Поиск оптимальной размерности входного образа в дальнейшем превратится в отдельную задачу. После эксперимента по изменению входного вектора для тренировки сети (вместо показателя по году на тренировку подавался вектор среднемесячных температур) точность предсказания тренируемой нами модели выросла до 0,61. В связи с этим предложено сделать еще большую дискретизацию по времени и подавать вектор, каждым элементом которого является усредненное значение по 10 подряд идущим дням: т.е. размерность у входного вектора выросла от 12 (число месяцев) до 354 (скользящее окно с шагом 1 по 10 дням). После экспериментов с новым входным вектором точность модели снизилась до 0,55, что, вероятно, связано со следующими причинами:

- при дозаполнении данных «синтетические» показатели учитываются не 1 раз, а до 10, что не увеличивает качество модели;

- модель стала работать с избыточным числом переменных, значительно превышающим реальное число независимых исходных данных, что отрицательно сказалось на стабильности алгоритма.

С учетом того, что тренируемой моделью является Random ForestClassifier, можно предположить, что из такого небольшого ко-

личества примеров классификатор не смог выделить сколько-нибудь значимые правила, так как пространство входных признаков выросло почти в 30 раз, а количество данных уменьшилось. Следствием этого является то, что теперь классификатору проще переучиться на тренировочные данные, а не выделять сколько-нибудь значимые признаки для классификации. Еще одно объяснение данного результата экспериментов: возможно, мы «проскочили» оптимальную размерность вектора образа для данной конкретной модели и данного массива данных.

Для проверки высказанной гипотезы предложено еще раз изменить дискретизацию по времени и подавать вектор, каждым элементом которого является усредненное значение по 10 подряд идущим дням без пересечений, т.е. размерность у входного вектора стала равной 36 (35 значений – усредненные по 10 дням, последнее значение усреднение по 15 дням). После экспериментов с новым входным вектором точность модели практически сравнялась с результатами, когда на вход подавался вектор из 12 значений (среднемесячные значения температур). Точность экспериментов с 12-значным входным вектором и 36-значным была одинаковой (0,61) вплоть до четвертого знака после запятой. Разброс значений «точность» изменился незначительно и составил от 0,38 до 0,77, что больше похоже на статистически допустимое различие.

Таким образом, серия экспериментов по тренировке сети на температурных данных показала, что число значимых независимых погодных параметров (по крайней мере, в использованном массиве), достоверно влияющих на урожайность следующего года, вряд ли превышает 12. При увеличении числа используемых значений результат модели не изменяется. Важным фактом является то, что на основании температурного профиля предыдущего года можно выделить, будет ли последующий год погодозависимым (с вероятностью 0,61–0,67) или нет. Для того чтобы научиться тренировать модель на выявление ожиданий от следующего года для принятия осознанного решения, следует ли прибегать к корректирующим мероприятиям или нет, необходима проверка на большем массиве более корректных данных со сменой искомого параметра ошибки для тренировки.

Важным результатом проведенной серии тренировок стало понимание того, что размерность погодного образа для использования в будущих моделях на самом деле не настолько велика. Поиск баланса между излишней и недостаточной гибкостью модели является ключевой концепцией машинного обучения и называется компромиссом между вариативностью и погрешностью (bias-variance tradeoff). Ключевой вопрос для любой применяемой модели – корректность исходных данных. Качество исходных данных существенно ограничивает круг подходов и моделей, которые могут быть использованы для решения задачи.

Итог рассмотрения проблемы интеллектуального анализа данных с помощью машинного обучения коротко можно сформулировать в следующем виде.

Задача получения прогнозов урожайностей различных культур в целях планирования оптимизации сельскохозяйственной деятельности – многопараметрическая и сложная задача, решаемая в условиях неполноты данных и недоопределенности граничных условий. Во многом решение подобных задач требует творческого подхода. Наиболее оптимальным подходом к их решению являются методы машинного или глубокого обучения.

Решение описанных задач сводится к созданию «образа» будущих событий в правильно выбранной метрике. Задача данного образа – не точное соответствие «картинке» будущего, а возможность предсказания нужных параметров с допустимой (возможной) степенью точности при использовании соответствующих алгоритмов, наложенных на образ. Для адекватной тренировки (обучения) алгоритмов глубокого машинного обучения необходимо достаточное количество обучающих примеров. Это требует скрупулёзного подхода к сбору статистических данных о результатах сельскохозяйственной деятельности.

Решение задачи создания адекватных и релевантных прогнозов возможной урожайности основных сельскохозяйственных культур для будущих вегетационных периодов создаст базу для совершенно нового подхода к планированию и оптимизации ведения сельского хозяйства.

ЗАКЛЮЧЕНИЕ

В соответствии с принятой Стратегией научно-технологического развития РФ современное аграрное производство ориентируется на все более широкое использование интеллектуальных информационных систем, обеспечивающих переход к новой модели развития, основанной на применении автоматизированных систем контроля и принятия решений, комплексной автоматизации и роботизации технологических процессов в растениеводстве и животноводстве, а также технологиях проектирования и моделирования сельскохозяйственного производства.

В условиях вирусной пандемии 2020 г. развернулась мировая тенденция на создание более коротких и устойчивых цепочек производства (в противовес «глобальным» цепочкам производства) критически важной продукции для нужд населения. К числу такой продукции, кроме медикаментов и медицинских изделий, относятся продукты питания. Сложился консенсус в отношении того, что необходимо развивать модель устойчивого и более продуктивного сельского хозяйства прежде всего на локальном и региональном уровне. Только такой подход способен решить как местные и региональные, так и глобальные проблемы продовольственной безопасности XXI в. Логический переход от региональной продовольственной безопасности к глобальной стабильности стал очевиден еще в период «кризиса беженцев» в 2015, 2016 гг. Если не решить вопрос с продовольственной безопасностью в каждом конкретном регионе, следующей проблемой станет массовое перемещение людей из регионов с недостаточной продовольственной обеспеченностью в более благополучные.

Решение проблемы продовольственной безопасности невозможно без эффективного планирования, которое должно строиться на адекватном прогнозировании. Сельское хозяйство является отраслью продуктивной деятельности человека, тесно связанной с жизненным циклом живых организмов (культурных растений и домашних животных), подверженной влиянию многочисленных внешних факторов с

огромной изменчивостью и не поддающихся прогнозированию, прежде всего погоды.

Увеличение эффективности сельскохозяйственного производства на основе существующих технологий в любом случае должно достигнуть предела, так как современный подход строится на применении «корректирующих» мероприятий по факту уже случившейся проблемы и не опирается на сколько-нибудь достоверный прогноз. Второе направление развития технологий – снижение зависимости сельского хозяйства от внешних факторов: создание культур, устойчивых к засухе, вредителям, болезням, морозостойких и др. Однако у данного направления также имеются лимиты, связанные с фундаментальными физическими и биологическими ограничениями. Например, жизнь невозможна при замерзании физиологических жидкостей или их полном испарении и др. К тому же приобретение сельскохозяйственными культурами одних качеств происходит за счет утраты или модификации других, иногда важных потребительских свойств. Так, клубника, предназначенная для длительного хранения и перевозки, становится гораздо более жесткой, теряя такое важное потребительское свойство, как нежность и сочность.

В условиях невозможности спрогнозировать длительность предстоящего благоприятного сельскохозяйственного периода в районах рискованного земледелия сельхозпроизводителям приходится делать выбор в пользу культур и сортов с более коротким вегетационным периодом, принимая во внимание, что ожидаемая урожайность будет заведомо ниже. Если удалось бы создать систему адекватного (в пределах ожидаемой погрешности) прогнозирования в комбинации с иными способами интенсификации сельского хозяйства, это позволило быкратно увеличить производительность сельского хозяйства и снизить риски.

Становится понятно, что только развитие существующих технологий сельского хозяйства не является решением всех проблем. Иначе это превратится в набор подходов, применяемых ситуационно без учета многофакторности проблем. В последние десятилетия огромный прогресс наметился в науках по обработке больших объемов информации (Big Data Science) с целью получения новых знаний, до сих пор считавшихся недостижимых при использовании классических

подходов. Применение этих методов в области сельскохозяйственного производства, прежде всего в части прогнозирования и поиска скрытых закономерностей и ассоциаций, открывает новые возможности для поиска оптимальных решений при планировании сельскохозяйственной деятельности для достижения оптимального результата при использовании имеющихся, часто ограниченных, ресурсов.

Так как в области науки о больших данных невозможно добиться прогресса, не обладая доступом к тем самым «большим данным», особое внимание следует уделить системе сбора и хранения данных о сельскохозяйственном производстве. Сбор данных необходимо поставить на регулярную основу с осуществлением контроля за качеством и полнотой получаемой информации.

Систематический анализ данных предполагает также решение важных фундаментальных проблем, связанных с интерпретацией данных. Необходимо разработать универсальный автоматизированный подход к оценке достоверности данных. Наличие системных ошибок или сознательное искажение данных приводят к системным ошибкам при обработке алгоритмами машинного обучения. Также необходимо иметь подход к оценке объема фактической информации (количества реально независимых данных или переменных), которые содержатся в имеющемся в распоряжении массиве экспериментальных данных или результатах наблюдений.

Прогнозирование благоприятной или неблагоприятной ситуации с ожидаемой урожайностью сельскохозяйственной культуры в следующем сезоне возможно на основе создания виртуального «образа» ожидаемого года. Создаваемый «образ» отражает интересующие нас показатели таким образом, чтобы прогноз, основанный на учете данных показателей, укладывался в известный интервал достоверности. Если по объективным причинам невозможно уложиться в требуемый интервал достоверности, то в результате вычислений необходимо иметь данные по фактическому интервалу достоверности.

Страна, которой удастся раньше других внедрить данную систему в народное хозяйство, сможет резко снизить себестоимость сельскохозяйственного производства, решить внутренние проблемы, обеспечить продовольственную безопасность и получить существенные конкурентные преимущества на глобальном рынке продуктов питания.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

Абрамчук И.Н., Пястолов С.М., Сафронова А.А. Модели управления знаниями в инновационных организациях // Инновационная экономика: информация, аналитика, прогнозы. – 2011. – № 3-4. – С. 63–68.

Агроэкологическая оценка земель, проектирование адаптивно-ландшафтных систем земледелия и агротехнологий: метод. руководство. – М.: Росинформагротех, 2005. – 784 с.

Адаптивно-ландшафтные системы земледелия Новосибирской области / РАСХН. Сиб. отд-ние. СибНИИЗХим. – Новосибирск, 2002 – 388 с.

Азуле О. Послание Генерального директора ЮНЕСКО г-жи Одрэ Азуле по случаю Всемирного дня логики. – ЮНЕСКО, 2020. – 2 с.

Акоф Р., Сасиени М. Основы исследования операций. – М.: Мир, 1971. – 536 с.

Алетдинова А., Ленский А., Цыбина Я. Имитационное моделирование как современный инструментарий для формирования машинно-тракторного парка // Аграрная экономика. – 2019. – Т. 3. – № 286. – С. 48–54.

Алексеева И.Ю. Человеческое знание и его компьютерный образ. – М.: ИФ РАН, 1993. – 218 с.

Анищенко А.Н. «Умное» сельское хозяйство как перспективный вектор роста аграрного сектора экономики России // Продовольственная политика и безопасность. – 2019. – Т. 6. – №. 2. – С. 97–106.

Арутюнов А.Л. Производство и применение биотоплива в сельском хозяйстве России // Научные труды: Институт народнохозяйственного прогнозирования РАН. – 2009. – № 7. – С. 722–734.

Аршинский Л.В. Теоретические основы искусственного интеллекта. – Иркутск: ИрГУПС, 2016. – 142 с.

Баженов Н.А. В Математическом центре в Академгородке отпраздновали День логики // Наука в Сибири (от 20 февраля 2020 г.).

Башмаков А.И., Башмаков И.А. Интеллектуальные информационные технологии. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2005. – 304 с.

Барсегян А.А., Куприянов М.С., Степаненко А.В., Холод И.И. Методы и модели анализа данных: OLAP и Data Mining. – СПб.: БХВ-Петербург, 2004. – 336 с.

Белоногов Г.Г., Кузнецов Б.А. Языковые средства автоматизированных информационных систем. – М.: Наука, 1983. – 288 с.

Бендик Н.В., Борхошкин О.В. Разработка базы знаний для сельскохозяйственных товаропроизводителей Иркутской области // Цифровые технологии и системы в сельском хозяйстве: материалы междунар. науч.-практич. конф. (8–10 октября 2019 г., пос. Молодежный, Иркутская область). – Иркутск: Изд-во Иркутский ГАУ, 2019. – С. 30–37.

Бибило П., Романов В. Логическое проектирование дискретных устройств с использованием продукционно-фреймовой модели представления знаний. – Минск: Беларус. навука, 2011. – 279 с.

Бова В.В. Концептуальная модель представления знаний при построении интеллектуальных информационных систем // Известия Южного федерального университета. Технические науки. – 2014. – № 7 (156). – С. 109–117.

Бова В.В., Курейчик В.В., Нужнов Е.В. Проблемы представления знаний в интегрированных системах поддержки управленческих решений // Известия Южного федерального университета. Технические науки. – 2010. – Т. 108. – № 7. – С. 107–113.

Бова В.В., Заруба Д.В., Курейчик В.В. Эволюционный подход к решению задачи интеграции онтологий // Известия Южного федерального университета. Технические науки. – 2015. – №. 6 (167). – С. 41–56.

Богатырев М.Ю., Вакурин В.С. Концептуальное моделирование в исследовании биомедицинских данных // Математическая биология и биоинформатика. – 2013. – Т. 8. – № 1. – С. 340–349.

Болотова Л.С. Системы искусственного интеллекта: модели и технологии, основанные на знаниях. – М.: Изд-во «XXXXXXXX», 2012. – 664 с.

Болотова Л.С., Мороз Ю.В., Смирнов С.С. и др. Метод ситуационного анализа и проектирования модели предметной области произвольной природы // Теоретические вопросы вычислительной тех-

ники и программного обеспечения: межвуз. сб. науч. тр. – М.: МГТУ МИРЭА, 2011. – С. 5–26.

Бондарев В.Н., Аде Ф.Г. Искусственный интеллект. – Севастополь: Изд-во СевНТУ, 2002. – 615 с.

Братко И. Язык PROLOG (Пролог): алгоритмы искусственного интеллекта. – М.: Издательский дом «Вильямс», 2007. – 640 с.

Бриллюэн Л. Наука и теория информации. – М.: Книга по Требованию, 2012. – 390 с.

Бронфельд Г.Б. Некоторые возможности формального представления структуры знаний // Труды НГТУ им. Р.Е. Алексеева. – 2012. – № 4 (97). – С. 91–100.

Бронфельд Г.Б. Понятия «информация», «знание» и современная четырехуровневая структура знаний // Математическое моделирование развивающейся экономики, экологии и технологий: сб. материалов IX Всерос. науч. конф. «ЭКОМОД-2016». – Киров, 2016. – С. 229–237.

Булаткин Г.А. Производство биотоплива второго поколения из растительного сырья // Вестник Российской академии наук. – 2010. – Т. 80. – № 5-6. – С. 522–527.

Булинский А.В., Ширяев А.Н. Теория случайных процессов. – М.: Физматлит, 2005. – 408 с.

Буч Г., Рамбо Д., Якобсон И. Язык UML. Руководство пользователя. – М.: ДМК Пресс, 2006. – 496 с.

Бычков И.В., Дородных Н.О., Юрин А.Ю. Подход к разработке программных компонентов для формирования баз знаний на основе концептуальных моделей // Вычислительные технологии. – 2016. – Т. 21. – № 4. – С. 16–36.

Вагин В.Н. Знание в интеллектуальных системах // Новости искусственного интеллекта. – 2002. – № 6 (54). – С. 8–18.

Варганова М.Л. Цифровая трансформация российского АПК в современных условиях // Экономика, предпринимательство и право. – 2019. – Т. 9. – № 4. – С. 301–310.

Васильев Д.Н., Чернов В.Г. Интеллектуальные информационные системы: основы теории построения. – Владимир: Изд-во Владимирского гос. ун-та, 2008. – 120 с.

Василькова В.В. Порядок и хаос в развитии социальных систем: синергетика и теория социальной самоорганизации. – СПб.: Изд-во Лань, 1999. – 480 с.

Власов М.В., Шишкина А.Ю. К вопросу о моделировании генерации знаний на предприятиях ОПК // Вестник Удмуртского университета. Серия: Экономика и право. – 2015. – Т. 25. – № 6-2. – С. 122–130.

Вовченко А.Е., Захаров В.Н., Калиниченко Л.А. и др. От спецификаций требований к концептуальной схеме // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: тр. 12-й Всерос. науч. конф. «RCDL-2010». – Казань, 2010. – С. 375–381.

Воевода А.А., Зимаев И.В. Соотнесение структурных и временных масштабов UML-диаграмм // Сб. науч. тр. НГТУ. – 2008. – № 4 (54). – С. 59–62.

Воевода А.А., Марков А.В. Методика автоматизированного проектирования программного обеспечения функционирования сложных систем на основе совместного использования UML-диаграмм и сетей Петри // Современные технологии. Системный анализ. Моделирование. – 2014. – № 2 (42). – С. 110–115.

Гаврилов А.В. Гибридные интеллектуальные системы. – Новосибирск: Изд-во НГТУ, 2002. – 142 с.

Гаврилова Т.А., Хорошевский В.Ф. Базы знаний интеллектуальных систем. – СПб.: Питер, 2000. – 384 с.

Гаврилова Т.А., Кудрявцев Д.В., Горовой В.А. Модели и методы формирования онтологий // Научно-технические ведомости СПб-ГПУ. – 2006. № 4. – С. 21–28.

Гаврилова Т.А., Алсуфьев А.И., Кокоулина Л.О. Управление знаниями с российским акцентом: победы и поражения // Инновации. – 2017. – № 1 (219). – С. 59–69.

Ганичева А.В., Редкова У.В., Узун Е.В. Интеллектуальные системы в сельском хозяйстве // Инновационные подходы к развитию науки и производства регионов: сб. науч. тр. по материалам Национальной науч.-практ. конф. – Тверь: Тверская ГСХА, 2019. – С. 461–464.

Гатаулин А.М., Гавриков А.В., Сорокина Т.М. и др. Математическое моделирование экономических процессов в сельском хозяйстве / под общ. ред. А.М. Гатаулина. – М.: Агропромиздат, 1990. – 432 с.

Гейтинг А. Интуиционизм. – М.: Мир, 1965. – 202 с.

Геоинформатика: в 2 кн. Кн. 1. / Е.Г. Капралов, А.В. Кошкарев, В.С. Тикунов и др.; под ред. В.С. Тикунова. – 3-е изд., перераб. и доп. – М.: Издательский центр «Академия», 2010. – 400 с.

Гиг Дж. Прикладная общая теория систем. Т. 1. – М.: Мир, 1981. – 336 с.

Гладун В.П. Растущие пирамидальные сети // Новости искусственного интеллекта. – 2004. – № 1. – С. 30–40.

Гладкий А.В. Формальные грамматики и языки. – М.: Наука, 1973. – 368 с.

Гладков Л.А., Курейчик В.В., Курейчик В.М. и др. Биоинспирированные методы в оптимизации. – М.: Физматлит, 2009. – 384 с.

Гозман О., Жаворонкова А., Рубальская А., Рягузов А. Путеводитель по MBA в России и за рубежом. – М.: Альпина Бизнес Букс, 2008. – 408 с.

Гома Х. UML. Проектирование систем реального времени, распределенных и параллельных приложений. – М.: ДМК Пресс, 2016. – 700 с.

Гриф М.Г., Гениатулина Е.В. Методы представления знаний в интеллектуальных системах // Сб. науч. тр. НГТУ. – 2009. – № 2 (56). – С. 87–92.

Гуськова Н.Д., Неретина Е.А. Предпосылки формирования и ключевые положения концепции «Индустрия 4.0» // Цифровая экономика и «Индустрия 4.0»: проблемы и перспективы: тр. науч.-практ. конф. с междунар. участием. – СПб.: Изд-во Политехн. ун-та, 2017. – С. 51–57.

Дадян Э.Г. Проектирование современных баз данных. – М.: ИНФРА-М, 2017. – 120 с.

Джарратано Дж., Райли Г. Экспертные системы: принципы разработки и программирование. – М.: Издательский дом «Вильямс», 2007. – 1152 с.

Долинина О.Н. Классификация ошибок в базах знаний экспертных систем // Вестник СГТУ. – 2010. – № 4 (50). – Вып. 2. – С. 25–130.

Дубровин А.Д. Интеллектуальные информационные системы. Ч. 1. – М.: МГУКИ, 2008. – 197 с.

Ерешко Ф.И., Меденников В.И. Формирование цифровой платформы АПК // Проектирование будущего. Проблемы цифровой реальности: тр. 1-й междунар. конф. (8–9 февраля 2018 г., Москва). – М.: ИПМ им. М.В. Келдыша, 2018. – С. 65–73.

Журавлёв Ю.И., Флёров Ю.А., Вялый М.Н. Дискретный анализ. Формальные системы и алгоритмы. – М.: ООО Контакт Плюс, 2010. – 336 с.

Загорулько Ю.А. Семантические сети и системы продукций: методическое пособие. – Новосибирск: Изд-во НГУ, 1996. – 46 с.

Загорулько Ю.А. О концепции интегрированной модели представления знаний // Известия Томского политехнического университета. Инжиниринг георесурсов. – 2013. – Т. 322. – № 5. – С. 98–103.

Загорулько Ю.А. Семантическая технология разработки интеллектуальных систем, ориентированная на экспертов // Онтология проектирования. – 2015. – Т. 5. – № 1 (15). – С. 30–46.

Загорулько Ю.А., Загорулько Г.Б. Онтологии и их практическое применение в системах, основанных на знаниях // Знания – Онтологии – Теории: материалы Всерос. конф. с международным участием «ЗОНТ-2011». – Новосибирск: Институт математики им. С.Л. Соболева СО РАН. – 2011. – Т. 1. – С. 1–10.

Заде Л. Понятие лингвистической переменной и его применение к принятию приближенных решений. – М.: Мир, 1976. – Т. 61. – 165 с.

Заславский М.Н. Эрозиоведение. Основы противоэрозионного земледелия. – М.: Высшая школа, 1987. – 376 с.

Захаров С.А. Курс почвоведения. – М.; Л.: Госиздат, 1927. – 455 с.

Зозуля Д.М. Цифровизация российской экономики и Индустрия 4.0: вызовы и перспективы // Вопросы инновационной экономики. – 2018. – Т. 8. – № 1. – С. 1–14.

Иванов В.Н., Никонова Ю.С. Экономико-математическое моделирование оптимальных параметров в сельском хозяйстве // Территория науки. – 2016. – № 2. – С. 83–89.

Ивлев Ю.В. Предмет и перспективы развития логики // Логические исследования – 2018. – Т. 24. – № 1. – С. 15–128.

Иноземцев В.А. Недедуктивная логика и проблема компьютерной репрезентации знания // Известия МГТУ «МАМИ». – 2014. – Т. 5 – № 2 (20). – С. 104–109.

Искусственный интеллект: в 3 кн. Кн. 1. Системы общения и экспертные системы: справочник / науч. ред. Э.В. Попов. – М.: Радио и связь, 1990. – 464 с.

Искусственный интеллект: в 3 кн. Кн. 2. Модели и методы: справочник / под ред. Д.А. Пospelова. – М.: Радио и связь, 1990. – 304 с.

Искусственный интеллект: в 3 кн. Кн. 3. Программные и аппаратные средства: справочник / науч. ред. В.Н. Захаров, В.Ф. Хорошевский. – М.: Радио и связь, 1990а. – 368 с.

Каличкин В.К., Павлова А.И. Агрономические геоинформационные системы. – Новосибирск: СФНЦА РАН, 2018. – 347 с.

Каличкин В.К., Куценогий П.К. Структура сельскохозяйственных данных для интеллектуального анализа // Развитие сельского хозяйства на основе современных научных достижений и интеллектуальных цифровых технологий «Сибирь – агробιοтехнологии»: сб. док. междунар. науч.-практ. конф. «Сибит – 2019»: – Новосибирск: СФНЦА РАН, 2019. – С. 167–173.

Каличкин В.К., Корякин Р.А., Куценогий П.К. Архитектура и принципы работы аграрной интеллектуальной системы // Сибирский вестник сельскохозяйственной науки. – 2019. – Т. 49. – № 4. – С. 65–75.

Каличкин В.К., Корякин Р.А., Лужных Т.А., Риксен В.С. Выбор и размещение сельскохозяйственной культуры с использованием искусственного интеллекта // Достижения науки и техники АПК. – 2019а. – Т. 33. – № 10. – С. 67–70.

Каличкин В.К., Корякин Р.А., Максимович К.Ю. и др. Концептуальная модель агроэкологических свойств земель // Сибирский вестник сельскохозяйственной науки. – 2020. – Т. 50. – № 1. – С. 73–80.

Карпов А.Е., Гик Е.Я. Шахматный калейдоскоп. – М.: Наука, 1982. – 208 с.

Карпенко А.С. Современные исследования в философской логике // Вопросы философии. – 2003. – Т. 9. – С. 54–75.

Каширин Д.И., Каширин И.Ю. Модели представления знаний в системах искусственного интеллекта // Вестник РГРТУ. – 2010. – Т. 1. – № 31. – С. 36–44.

Клайн М. Математика. Утрата определённости. – М.: Мир, 1984. – 434 с.

Клини С.К. Введение в метаматематику. – М.: Иностранная литература, 1957. – 526 с.

Колмогоров А.Н. О принципе tertium non datur // Математический сборник. – 1925. – Т. 32. – № 4. – С. 646–667.

Колмогоров А.Н. Избранные труды. Математика и механика. – М.: Наука, 1985. С. 393–404.

Кордонский С.Г. Система понятий, используемых для описания поведения. – Томск: Биолого-почвенный факультет ТГУ (дипломная работа), 1977.

Кордонский С.Г., Бардин В.В. Прикладная герменевтика информационного пространства: картины мира, теоретические онтологии и верные матрицы // Философско-литературный журнал «Логос». – 2015. – Т. 5. – № 3. – С. 19–45.

Кордонский С.Г., Бардин В.В. Картины миры и информационная герменевтика // Телескоп: журнал социологических и маркетинговых исследований. – 2018. – № 3. – С. 19–23.

Корниенко А.А., Корниенко А.А., Корниенко А.В. К вопросу о философских предпосылках, состоянии и перспективах исследований по проблеме искусственного интеллекта // Известия Томского политехнического университета. Инжиниринг георесурсов. – 2013. – Т. 323. – № 6. – С. 210–215.

Кощева А.К. Новые константы в предтабличных суперинтуиционистских логиках: подход П.С. Новикова // Известия Института математики и информатики УдГУ. – 2016. – № 1. – С. 3–33.

Красильников О.Ю. Знания и информация в концепциях «новой» экономики // Известия Саратовского университета. Серия: Экономика. Управление. Право. – 2012. – Т. 12. – Вып. 2. – С. 12–16.

Крэг Л. Применение UML 2.0 и шаблонов проектирования = Applying UML and Patterns: An Introduction to Object-Oriented Analysis and Design and Iterative Development. 3-е изд. – М.: Издательский дом «Вильямс», 2006. – 736 с.

Кузнецов П.П. Семантические представления. – М.: Наука, 1986. – 290 с.

Курейчик В.М., Писаренко В.И., Кравченко Ю.А. Инновационные образовательные технологии в построении систем поддержки принятия групповых решений // Известия Южного федерального университета. Технические науки. – 2008. – № 4 (81). – С. 216–221.

Кухарев О.Н., Гнусарев И.В., Ясавкин Д.А. Применение экономико-математических моделей для создания оптимальной структуры машинно-тракторного парка в сельском хозяйстве // Нива Поволжья. – 2012. – № 4. – С. 121–124.

Лавров И.А., Максимова Л.Л. Задачи по теории множеств, математической логике и теории алгоритмов. – М.: Физматлит, 2002. – 256 с.

Лазько О.В., Семченкова С.В., Лукашева О.Л. Стратегические ориентиры управления процессами цифровой трансформации в агропромышленном комплексе // Московский экономический журнал. – 2018. – № 4. – С. 507–518.

Леньков И.И. Экономико-математическое моделирование экономических систем и процессов в сельском хозяйстве. – Минск: Дизайн ПРО, 1997. – 304 с.

Либих Ю. Химия в приложении к земледелию и физиологии. – М.: Гос. изд-во колхозной и совхозной литературы, 1936. – 408 с.

Липкович Э.И. Моделирование сверхкрупных агросистем // Экономика сельского хозяйства России. – 2008. – № 5. – С. 23–30.

Ллойд С. Программируя вселенную: Квантовый компьютер и будущее науки. – М.: Альпина нон-фикшн, 2014. – 256 с.

Лозовский В.С. Семантические сети // Представление знаний в человеко-машинных и робототехнических системах. – М.: ВИНТИ, 1984. – Т. А. – С. 84–121.

Лозовский В.С. Сетевые модели // Искусственный интеллект. В 3-х кн. Кн. 2. – М.: Радио и связь, 1990. – С. 28–49.

Лукашевич Н.В. Тезаурусы в задачах информационного поиска. – М.: Изд-во Московского университета, 2011. – 512 с.

Лупян Е.А., Прошин А.А., Бурцев М.А. и др. Центр коллективного пользования системами архивации, обработки и анализа спутниковых данных ИКИ РАН для решения задач изучения и мониторинга окружающей среды // Современные проблемы дистанционного зондирования Земли из космоса. – 2015. – Т. 12. – № 5. – С. 263–284.

Лысак И.В. Проблема сущности знания в эпистемологии и в исследованиях в области искусственного интеллекта // Манускрипт. – 2017. – № 5 (79). – С. 100–104.

Люггер Д.Ф. Искусственный интеллект: стратегии и методы решения сложных проблем. – М.: Издательский дом «Вильямс», 2004. – 864 с.

Лясников Н.В. Цифровой аграрный сектор России: обзор прорывных технологий четвертого технологического уклада // Продовольственная политика и безопасность. – 2018. – Т. 5. – № 4. – С. 169–181.

Макаренко С.И. Интеллектуальные информационные системы. – Ставрополь: СФ МГТУ им. М.А. Шолохова, 2009. – 206 с.

Максимова Л.Л., Юн В.Ф. Узнаваемость в предгейтинговых и стройных логиках // Сибирские электронные математические известия. – 2019. – Т. 16. – С. 427–434.

Мальцев А.И. Алгебраические системы. – М.: Наука, 1970. – 392 с.

Мансуров А.В. Нейросетевой метод определения температуры и влажности неоднородно увлажненных почв по данным СВЧ-радиометрии // ОНВ. – 2006. – № 4 (38). – С. 121–125.

Мардаев С.И. Суперинтуиционистские логики с условиями финитности: дис. ... канд. физ.-мат. наук: 01.01.06. – Новосибирск, 1987. – 109 с.

Маринко Г.И. Современные модели и школы в управлении знанием // Вестник Московского университета. Серия 21: Управление (государство и общество). – 2004. – № . – С. 1–23.

Марков А.В. Разработка программного обеспечения при совместном использовании UML-диаграмм и сетей Петри (обзор) // Сб. науч. тр. НГТУ. – 2013. – № 1. – С. 96–131.

Мартынов В.В. Основы семантического кодирования. – М.: ЕГУ, 2001. – 140 с.

Микешина Л.А. Философия науки. – М.: Прогресс-Традиция; Флинта, 2005. – 464 с.

Мильнер Б.З. Концепция управления знаниями в современных организациях // Российский журнал менеджмента. – 2003. – № 1. – С. 57–76.

Минский М. Фреймы для представления знаний / пер. с англ. О.Н. Гринбаума; под ред. Ф.М. Кулакова. – М.: Энергия, 1979. – 151 с.

Мозговой М.В. Классика программирования: алгоритмы, языки, автоматы, компиляторы. Практический подход. – СПб.: Наука и техника, 2006. – 320 с.

Москвин В.В. Типизация моделей агроэкосистем // Вестник ВГУ. Серия: Химия. Биология. Фармация. – 2011. – № 1. – С. 85–90.

Мошелла Д. Путеводитель по цифровому будущему: отрасли, организации и профессии. – М.: Альпина Паблишер, 2020. – 215 с.

Нариньяни А.С. Недоопределенность в системах представления и обработки знаний // Известия АН СССР. Техн. кибернетика. – 1986. – № 5. – С. 3–28.

Нариньяни А.С., Телерман В.В., Ушаков Д.М. и др. Программирование в ограничениях и недоопределённые модели // Информационные технологии. – 1998. – № 7. – С. 13–22.

Нариньяни А.С. Инженерия знаний и НЕ-факторы: краткий обзор-08. Вопросы искусственного интеллекта // Вестник НСМИИ РАН. – 2008. – № 1. – С. 61–77.

Нейлор К. Как построить свою экспертную систему. – М.: Энергоатомиздат, 1991. – 286 с.

Нестик Т.А. Модели управления знаниями в российских организациях: социально психологический анализ // Социально-экономические и психологические проблемы управления: сб. науч. ст. по материалам I (IV) междунар. науч.-практ. конф. (Москва, 23–25 апреля 2008 г.). – М.: ИСЭП, 2008. – С. 278–286.

2013 г.) / под общ. ред. М.Г. Ковтунович. – М.: Изд-во Московского государственного психолого-педагогического университета. – 2013. – С. 245–259.

Нечаев В., Тюпаков К., Сайфетдинова Н. Эффективность экономического и технико-технологического потенциала растениеводства // Экономика сельского хозяйства России. – 2012. – № 6. – С. 70–84.

Никаноров С.П. Концептуализация предметных областей. Серия: «Концептуальный анализ и проектирование». Методология и технология. – М.: Концепт, 2009. – 268 с.

Николенко С.И., Тулупьев А.Л. Самообучающиеся системы. – М.: МЦМНО, 2009. – 288 с.

Никонова А.А. «Новая системность» стратегического планирования в Индустрии 4.0 // Journal of new economy. – 2019. – Т. 20. – № 2. – С. 145–165.

Нильсон Н. Принципы искусственного интеллекта. – М.: Радио и связь, 1985. – 372 с.

Нитежук М.С. Верификация и поиск противоречий в базах знаний интеллектуальных систем // Молодая наука Сибири: электрон. науч. журн. – 2018. – № 2. Режим доступа: <http://mnv.irkups.ru/toma/22-2018>

Носонов А.М. Математическое моделирование принятия решений в агрогеографических исследованиях // Региональные исследования. – 2017. – № 4. – С. 5–14.

Огнивцев С.Б. Цифровизация экономики и экономика цифровизации АПК // Международный сельскохозяйственный журнал. – 2019. – № 2. – С. 77–80.

Ожерельева Т.А. Модели отношений данных в информационном поле // Международный журнал прикладных и фундаментальных исследований. – 2015. – № 2-1. – С. 22–24.

Орлов А.Д. Эрозия и эрозионноопасные земли Западной Сибири. – Новосибирск: Наука, 1983. – 207 с.

Осенний В.В., Франциско О.Ю. Перспективы применения методов и средств имитационного моделирования в аграрной сфере // Труды Кубанского государственного аграрного университета. – 2019. – № 79. – С. 55–60.

Осипов Г.С. Методы искусственного интеллекта. – М.: Физматлит, 2011. – 296 с.

Осуга С. Обработка знаний. – М.: Мир, 1989. – 293 с.

Палагин А.В., Крывый С.Л., Петренко Н.Г. Онтологические методы и средства обработки предметных знаний. – Луганск: Изд-во ВНУ им. В. Даля, 2012. – 324 с.

Палицин В. К вопросу о соотношении понятий знания, информация, данные // Наука и инновации. – 2018. – № 2 (180). – С. 44–49.

Пахомов Е.В. Цифровизация умного города // Инженерный вестник Дона: электронный научный журнал. – 2017. – Т. 47. – № 4 (47).

Петрушин А.Ф. Комплекс программ формирования и обработки баз данных и знаний в агрономии: автореф. дис. ... канд. техн. наук. – СПб.: АФИ, 2005. – 28 с.

Попов Э.В. Экспертные системы: решение неформализованных задач в диалоге с ЭВМ. – М.: Наука, 1987. – 288 с.

Поспелов Д.А. Логико-лингвистические модели в системах управления. – М.: Энергоиздат, 1981. – 231 с.

Поспелов Д.А. Ситуационное управление. Теория и практика. – М.: Наука, 1986. – 288 с.

Поспелов Д.А. Моделирование рассуждений: опыт анализа мыслительных актов. – М.: Радио и связь, 1989. – 184 с.

Поспелов Д.А. Продукционные модели // Искусственный интеллект: в 3-х кн. Кн. 2. – М.: Радио и связь, 1990. – С. 49–56.

Поспелова Л.Я. Мера непротиворечивости нечеткой базы знаний // Инженерные системы. // Труды VI междунар. науч.-практ. конф., посвященной 100-летию юбилею первого ректора РУДН профессора С.В. Румянцева. – М.: Изд-во РУДН, 2013. – С. 289–295.

Поспелова Л.Я. Мера согласованности нечеткой базы знаний и методика ее оценки // XII всероссийское совещание по проблемам управления ВСПУ-2014. – М.: Изд-во: Институт проблем управления им. В.А. Трапезникова, 2014. – С. 4010–4017.

Представление и использование знаний / под ред. Х. Уэно, М. Исидзука. – М.: Мир, 1989. – 220 с.

Пястолов С.М., Кураду К., Бонтис Н. Параллели в циклах формирования знания // Социальные и гуманитарные науки. Отечественная и зарубежная литература. Серия 8: Науковедение: реферативный журнал. – 2011. – № 3. – С. 88–94.

Рей Э. Изучаем XML / пер. с англ. – СПб.: Символ-Плюс, 2001. – 408 с.

Родионов А.В., Макарова Е.И. Экономико-математическое моделирование оптимизации структуры посевных площадей аграрных предприятий // Вестник Южно-Российского ГТУ. Серия: Социально-экономические науки. – 2016. – № 1. – С. 88–92.

Ручкин В.Н., Фулин В.А. Универсальный искусственный интеллект и экспертные системы. – СПб.: БХВ-Петербург, 2009. – 240 с.

Рыбина Г.В. Основы построения интеллектуальных систем. – М.: Финансы и статистика. Инфра-М, 2010. – 432 с.

Седякин В.П. Информация и знания // Научные ведомости Белгородского государственного университета. Серия: Философия. Социология. Право. – 2009. – Т. 8. – № 8 (63). – С. 180–187.

Седякин В.П., Цветков В.Я. Философия информационного подхода. – М.: МАКС Пресс, 2007. – 217 с.

Семёнов А.Л. Состояние и перспективы математического образования в России // Федеральный справочник. Образование в России. – 2013. – Т. 9. – С. 139–146.

Сидоркина И.Г. Системы искусственного интеллекта. – М.: КНОРУС, 2015. – 248 с.

Скобелев П.О. Онтологии деятельности для ситуационного управления предприятиями в реальном времени // Онтология проектирования. – 2012. – № 1 (3). – С. 6–38.

Скворцов Н.А., Калиниченко Л.А., Ковалев Д.Ю. Концептуальное моделирование предметных областей с интенсивным использованием данных // Аналитика и управление данными в областях с интенсивным использованием данных: тр. XVIII междунар. конф. «DAMDID/RCDL-2016» (Ершово, 11–14 октября 2016 г.). – 2016. – С. 33–41.

Скоркин В.К., Ларкин Д.К., Аксенова В.П. и др. Математическая модель и алгоритм расчета потребности в кормах для молочно-товарных ферм // Вестник ВНИИМЖ. – 2013. – № 3 (11). – С. 131–137.

Смагин Б.И. Экономический анализ и статистическое моделирование аграрного производства. – Мичуринск: Изд-во Мичуринского ГАУ, 2007. – 153 с.

Сорокин А.Б. Концептуальное проектирование интеллектуальных систем поддержки принятия решений // Онтология проектирования. – 2017. – Т. 7. – № 3 (25). – С. 247–269.

Сурмин Ю.П. Теория систем и системный анализ. – Киев: МАУП, 2003. – 368 с.

Слэйгл Дж. Искусственный интеллект. – М.: Мир, 1973. – 319 с.

Тарасенко Ф.П. Прикладной системный анализ. – М.: Кронус, 2017. – 322 с.

Терновой М.Ю., Штогрин Е.С. Формальная спецификация свойств баз нечетких знаний Мамдани на основе метаграфа // Вестник Харьковского национального университета им. В.Н. Каразина. Серия: Математическое моделирование. Информационные технологии. Автоматизированные системы управления. – 2015. – Т. 27. – С. 157–171.

Теслинов А.Г. Развитие систем управления: методология и концептуальные структуры. – М.: Глобус, 1998. – 299 с.

Тихонов А.Н., Арсенин В.Я. Методы решения некорректных задач. – М.: Наука, 1979.

Токмаков Г.П. Автоматизированное проектирование информационных систем: учебное пособие. – Ульяновск: Ульяновский ГТУ, 2015. – 121 с.

Толковый словарь по искусственному интеллекту / авторы-сост. А.Н. Аверкин, М.Г. Гаазе-Рапопорт, Д.А. Поспелов. – М.: Радио и связь, 1992. – 256 с.

Тоффоли Т., Марголус Н. Машины клеточных автоматов. – М.: Мир, 1991. – 280 с.

Тугускина Г.Н., Рожкова Л.В., Сальникова О.В. Управление знаниями в современных организациях // Известия высших учебных заведений. Поволжский регион. – 2019. – № 2 (50). – С. 210–218.

Тузовский А.Ф. Объектно-ориентированное программирование. – М.: Юрайт, 2019. – 206 с.

Тузовский А.Ф., Чириков С.В., Ямпольский В.З. Системы управления знаниями (методы и технологии) / под общ. ред. В.З. Ямпольского. – Томск: Изд-во НТЛ, 2005. – 260 с.

Улезько А.В., Курносое А.П., Тютюников А.А. Имитационное моделирование как инструмент исследования агроэкономических систем // Экономика сельскохозяйственных и перерабатывающих предприятий. – 2012. – № 8. – С. 28–30.

Уотермен О. Руководство по экспертным системам. – М.: Мир, 1989. – 388 с.

Ушаков Д.М., Телерман В.В. Системы программирования в ограничениях (обзор) // Системная информатика. Проблемы теории и методологии создания параллельных и распределенных систем: сб. науч. тр., вып. 7. – Новосибирск: Наука, 2000. – С. 275–310.

Фаулер М. UML. Основы. 3-е изд. / пер. с англ. – СПб: Символ-Плюс, 2004. – 192 с.

Федеральный государственный образовательный стандарт высшего образования. Утвержден приказом Министерства образования и науки Российской Федерации № 195 от 11 марта 2015 г.

Федотов А.М., Байдавлетов А.Т., Жижимов О.Л. и др. Цифровой репозиторий в научно-образовательной информационной системе // Вестник Новосибирского государственного университета. Серия: Информационные технологии. – 2015. – Т. 13. – № 3. – С. 68–86.

Фролов А.А., Ежов А.А. Дискуссия о нейрокомпьютерах. – М.: МИФИ, 2000. – 224 с.

Харитоновна Т.В., Волкова Г.А. Использование экономико-математических методов в планировании деятельности сельскохозяйственных организаций // Нива Поволжья. – 2015. – № 4 (37). – С. 155–160.

Хопкрофт Дж.Э., Мотванн Р., Ульман Д.Д. Введение в теорию автоматов, языков и вычислений. – М.: Издательский дом «Вильямс», 2002. – 527 с.

Цветков В.Я. Социальные аспекты информатизации образования // Международный журнал экспериментального образования. – 2013. – № 4. – С. 108–111.

Частиков А.П., Гаврилова Т.А., Белов Д.Л. Разработка экспертных систем. Среда CLIPS. – СПб.: БХВ-Петербург, 2003. – 608 с.

Чебанов С.В. Об адекватности использования герменевтических техник // Вестник ТвГУ. Серия: Филология. – 2008. – № 14. – С. 273–280.

Черняк Л. Grid как будущее компьютеринга // Открытые системы. СУБД. – 2003. – № 1. <http://www.osp.ru/os/2003/01/182390/>

Шенк Р. Обработка концептуальной информации. – М.: Энергия, 1980. – 360 с.

Шеров А.Г. Оптимизация структуры севооборотов сельскохозяйственных культур // Актуальные проблемы современной науки. – 2005. – № 3. – С. 155–156.

Штовба С.Д. Проектирование нечетких систем средствами MATLAB. – М.: Горячая линия-Телеком, 2007. – Т. 288. – 403 с.

Штовба С.Д. Обеспечение точности и прозрачности нечеткой модели Мамдани при обучении по экспериментальным данным // Проблемы управления и информатики. – 2007а. – № 4. – С. 102–114.

Штовба С.Д., Штовба Е.В., Панкевич О.Д. Критерии точности и компактности для оценки качества нечетких баз знаний в задачах идентификации // Научные труды Винницкого национального технического университета. – 2012. – № 4. – С. 1–8.

Штовба С.Д., Мазуренко В.В., Тылец Р.О. Информационная технология нечеткой идентификации для синтеза точных, компактных и интерпретабельных баз знаний // Computer Sciences and Telecommunications. – 2016. – № 1. – С. 8–22.

Экспертные системы. Принципы работы и примеры / А. Брукинг, П. Джонс, Ф. Кокс и др.; под ред. Р. Форсайта; пер. с англ. – М.: Радио и связь, 1987. – 224 с.

Юрин А.Ю. Нотация для проектирования баз знаний продукционных экспертных систем // Объектные системы. – 2016. – № 12. – С. 48–54.

Якушев В.В. Точное земледелие: теория и практика. – СПб.: АФИ, 2016. – 364 с.

Якушев В.В., Ломакин В.С., Часовских С.В. и др. К вопросу построения модели представления знаний в СППР с элементами точного земледелия // Применение средств дистанционного зондирования Земли в сельском хозяйстве: материалы II Всерос. конф. с международным участием (Санкт-Петербург, 26–28 сентября 2018 г.). – СПб.: АФИ, 2018. – С. 350–353.

Якушев В.П., Петрушин А.Ф. Методология проектирования компьютерных систем поддержки решений в агрономии // Сб. док. междунар. конф. по мягким вычислениям и измерениям «SMC-2000». – СПб.: СЭТУ, 2000. – Т. 2. – С. 137–140.

Якушев В.П., Михайленко И.М., Петрушин А.Ф. и др. Программно-аппаратный комплекс поддержки принятия технологических решений в точном земледелии. // Материалы междунар. науч.-практ. конф. «АГРОИНФО-2003». – Новосибирск, 2003. – С. 16–22.

Ackoff R.L. From data to wisdom // Journal of applied systems analysis. – 1989. – Vol. 16. – N. 1. – P. 3–9.

Abdul-Ghafour S., Ghodous P., Shariat B. et al. A common design-features ontology for product data semantics interoperability // IEEE/WIC/ACM: International Conference on Web Intelligence (WI'07). IEEE. – 2007. – P. 443–446.

Alavi M., Kayworth T.R., Leidner D.E. An empirical examination of the influence of organizational culture on knowledge management practices // Journal of management information systems. – 2005. – Vol. 22. – N 3. – P. 191–224.

Alpaydin E. Introduction to machine learning. – London: MIT press, 2020. – 683 p.

Anand A., Singh M.D. Understanding knowledge management // International Journal of Engineering Science and Technology. – 2011. – Vol. 3. – N 2. – P. 926–939.

Antle J.M., Basso B., Conant R.T., et al. Towards a new generation of agricultural system data, models and knowledge products: Design and improvement // Agricultural Systems. – 2017. – Vol. 155. – P. 255–268.

Antle J.M., Jones J.W., Rosenzweig C. Next generation agricultural system models and knowledge products: synthesis and strategy // Agricultural Systems. – 2017a. – Vol. 155. – P. 179–185.

Apt K.R. Principles of Constraint Programming. – New York: Cambridge University Press, 2003. – 407 p.

Asrar-ul-Haq M., Anwar S., Nisar T. A systematic review of knowledge management and knowledge sharing: Trends, issues, and challenges // Cogent Business & Management. – 2016. – Vol. 3. – N 1. – P. 1–17.

Aven T. A conceptual framework for linking risk and the elements of the data-information-knowledge-wisdom (DIKW) hierarchy // Reliability Engineering & System Safety. – 2013. – Vol. 111. – P. 30–36.

Baader F., Calvanese D., McGuinness D.L. et al. The Description Logic Handbook: Theory, Implementation, Applications. – Cambridge, 2003. – 574 p.

Bashurova V.S., Dreiling V., Hodger T.V. et al. Measurements of condensation nuclei size distribution in remote continental area – Lake Baikal, Siberia // Journal of aerosol science. – 1992. – Vol. 23. – N 2. – P. 191–199.

Baskarada S., Koronios A. Data, information, knowledge, wisdom (DIKW): a semiotic theoretical and empirical exploration of the hierarchy and its quality dimension // Australasian Journal of Information Systems. – 2013. – Vol. 18. – N 1. – P. 5–24.

Baskerville R., Dulipovici A. The theoretical foundations of knowledge management // Knowledge Management Research & Practice. – 2006. – Vol. 4. – N 2. – P. 83–105.

Berners-Lee T., Hendler J., Lassila O. The Semantic Web // Scientific American. – 2001. – Vol. 284. – N 5. – P. 28–37.

Bernstein J.H. The data-information-knowledge-wisdom hierarchy and its antithesis // Proceedings North American Symposium on Knowledge Organization. – 2009. – Vol. 2. Syracuse, NY. – P. 68–75.

Boisot M.H. Information Space: A Framework for Learning in Organizations, Institutions and Culture, Routledge // London, UK. 1995.

Boisot M., Canals A., MacMillan I. Simulating I-Space (SIS): An agent-based approach to modeling knowledge flows // Working papers of the Sol C. Snider Entrepreneurial Research Center, Wharton School, University of Pennsylvania, 2004. – 35 p.

Boose J.H. A survey of knowledge acquisition techniques and tools // Knowledge acquisition. – 1989. – Vol. 1. – N 1. – P. 3–37.

Breuker J., Wielinga B. Models of expertise in knowledge acquisition // Studies in Computer Science and Artificial Intelligence. – North-Holland. – 1989. – Vol. 5. – P. 265–295.

Breiman L. Bagging predictors // Machine learning. – 1996. – Vol. 24. – N 2. – P. 123–140.

Breiman L. Random forests // Machine learning. – 2001. – Vol. 45. – N 1. – P. 5–32.

Breiman L., Friedman J.H., Olshen R.A. et al. Classification and Regression Trees. – Routledge, Abingdon, UK. – 1984. – Vol. 19.

Broomhead D.S., Lowe D. Multivariable Functional Interpolation and Adaptive Networks // Complex Syst. – 1988. – N 2. – P. 321–355.

Buchanan B.G., Barstow D., Bechtel R. et al. Constructing an expert system. In Building Expert Systems // Addison Wisley. – 1983. – P. 127–167.

Cheng Y.S., Yeh H.C. Analysis of screen diffusion battery data // American Industrial Hygiene Association Journal. – 1984. – Vol. 45. – N 8. – P. 556–561.

Computer-based medical consultations: MYCIN / E. Shortliffe (ed.). – Elsevier. – 2012. – Vol. 2. – 267 p.

Cooke N.J. Knowledge elicitation // Handbook of applied cognition. – 1999. – P. 479–509.

Cooper D.W., Spielman L.A. Data inversion using nonlinear programming with physical constraints: aerosol size distribution measurement by impactors // Atmospheric Environment (1967). – 1976. – Vol. 10. – N 9. – P. 723–729.

Cooper D.W., Wu J.J. The inversion matrix and error estimation in data inversion: application to diffusion battery measurements // Journal of aerosol science. – 1990. – Vol. 21. – N 2. – P. 217–226.

Cordon O. A historical review of evolutionary learning methods for Mamdani-type fuzzy rule-based systems: Designing interpretable genetic fuzzy systems // International Journal of Approximate Reasoning. – 2011. – Vol. 52. – P. 894–913.

Crump J.G., Seinfeld J.H. A new algorithm for inversion of aerosol size distribution data // *Aerosol Science and Technology*. – 1981. – Vol. 1. – N 1. – P. 15–34.

Curado C., Bontis N. Parallels in knowledge cycles // *Computers in Human Behavior*. – 2011. – Vol. 27. – N 4. – P. 1438–1444.

Davenport T.H., de Long D.W., Beers M.C. Successful Knowledge Management Project // *Sloan Management Review*. – 1998. Vol. 39. – N 2. – P. 43–57.

Deng L., Yu D. Deep Learning: Methods and Applications // *Foundations and Trends in Signal Processing*. – 2014. – Vol. 7. – N 3-4. – P. 197–387.

Dobing B., Parsons J. How UML is used // *Communications of the ACM*. – 2006. – Vol. 49. – N 5. – P. 109–113.

Drury B. Fernandes R., Moura M-F. et al. A survey of semantic web technology for agriculture // *Information Processing in Agriculture*. – 2019. – Vol. 6. – N 4. – P. 487–501.

Duda R.O., Hart P.E., Stork D.G. Pattern classification and scene analysis. – New York: Wiley, 1973. – Vol. 3. – P. 731–739.

Dzubay T.G., Hasan H. Fitting multimodal lognormal size distributions to cascade impactor data // *Aerosol science and technology*. – 1990. – Vol. 13. – N 2. P. 144–150.

Evans J., Foster J. Metaknowledge // *Science*. – 2011. – Vol. 331. – P. 721–725.

Fellbaum C. WordNet: An Electronic Database. – MIT Press, 1998. – 449 p.

Frické M. The knowledge pyramid: a critique of the DIKW hierarchy // *Journal of information science*. – 2009. – Vol. 35. – N 2. – P. 131–142.

Gerring J. Case Study Research: Principles and Practices. – New York: Cambridge University Press, 2007. – 278 p.

Groeneveld J., Müller B., Buchmann C.M. et al. Theoretical foundations of human decision-making in agent-based land use models – A review // *Environmental modeling & software*. – 2017. – Vol. 87. – P. 39–48.

Guarino N. Formal Ontology in Information Systems // *Formal Ontology in Information Systems. Proceedings of FOIS'98 (Trento, Italy, 6–8 June 1998)*. – Amsterdam: IOS Press, 1998. – P. 3–15.

Hand D.J., Adams N.M. Data Mining // Wiley StatsRef: Statistics Reference Online. – 2014. – P. 1–7.

Hassoun M.H. Fundamentals of artificial neural networks // MIT press, 1995. – 58 p.

Heinemann P.H. Decision support systems for food and agriculture // Systems analysis and modeling in food and agriculture. – 2010. – P. 121–132

Helsper C., Fissan H., Kapadia A. et al. Data inversion by simplex minimization for the electrical aerosol analyzer // Aerosol Science and Technology. – 1982. – Vol. 1. – N 2. – P. 135–146.

Hermans F., Klerkx L., Roep D. Structural conditions for collaboration and learning in innovation networks: using an innovation system performance lens to analyse agricultural knowledge systems // The Journal of Agricultural Education and Extension. – 2015. – Vol. 21. – N 1. – P. 35–54.

Hey T., Tansley S., Tolle K. The fourth paradigm: data-intensive scientific discovery. – Redmond, WA: Microsoft research, 2009. – Vol. 1. – 287 p.

Hopfield J.J. Neural networks and physical systems with emergent collective computational abilities // Proceedings of National Academy of Sciences. – 1982. – Vol. 79. – N 8. – P. 2554–2558.

Huang G.B., Zhu Q.Y., Siew C.K. Extreme learning machine: theory and applications // Neurocomputing. – 2006. – Vol. 70. – N 1-3. – P. 489–501.

Hwang J.S. The fourth industrial revolution (industry 4.0): intelligent manufacturing // SMT Magazine. – 2016. – Vol. 3. – P. 616–630.

Ingram J. Agricultural transition: Niche and regime knowledge systems' boundary dynamics // Environmental Innovation and Societal Transitions. – 2018. – Vol. 26. – P. 117–135.

Jackson P. Introduction to expert systems. – Addison-Wesley Longman Publishing Co., Inc., 1998. – 542 p.

Jakku E., Thorburn P.J. A conceptual framework for guiding the participatory development of agricultural decision support systems // Agricultural systems. 2010. – Vol. 10. – N 9. – P. 675–682.

Jakobson I., Spence I., Bittner K. Use Case 2.0. The Guide to Succeeding with Use Cases. – Ivar Jacobson International SA, 2011. – 55 p.

Janssen S.J.C., Porter C.H., Moore A.D. et al. Towards a new generation of agricultural system data, models and knowledge products: Information and communication technology // *Agricultural systems*. – 2017. – Vol. 155. – P. 200–212.

Jennex M.E., Smolnik S., Croasdell D.T. Towards a Consensus Knowledge Management Success Definition // *The Journal of Information and Knowledge Management Systems*. – 2009. – N 39 (2). – P. 174–188.

Jennex M.E., Smolnik S., Croasdell D. Knowledge management success in practice // *IEEE: 47th Hawaii International Conference on System Sciences*. – 2014. – P. 3615–3624.

Jones J.W., Antle J.M., Basso B. et al. Brief history of agricultural systems modeling // *Agricultural systems*. – 2017. – Vol. 155. – P. 240–254.

Kass G.V. An exploratory technique for investigating large quantities of categorical data // *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. – 1980. – Vol. 29. – N 2. – P. 119–127.

Klerkx L., Jansen J. Building knowledge systems for sustainable agriculture: supporting private advisors to adequately address sustainable farm management in regular service contacts // *International Journal of Agricultural Sustainability*. – 2010. – Vol. 8. – N 3. – P. 148–163.

Knight K. Unification: A Multidisciplinary Survey // *ACM Computing Surveys (CSUR)*. – 1989. – Vol. 21. – N 1. – P. 93–124.

Koutsenogii P.K. Condensation nuclei size distributions Siberia // *Journal of aerosol science*. – 1993. – Vol. 24. – N 1. – P. 31–32.

Lavrac N., Mozetic I. Methods for knowledge acquisition and refinement in second generation expert systems // *ACM SIGART Bulletin*. – 1989. – N 108. – P. 63–69.

Le Q.B., Seidl R., Scholz R.W. Feedback loops and types of adaptation in the modelling of land-use decisions in an agent-based simulation // *Environmental Modelling & Software*. – 2012. – Vol. 27. – P. 83–96.

LeCun Y., Bengio Y., Hinton G. Deep learning // *Nature*. – 2015. – Vol. 521. – N 7553. – P. 436–444.

Lee J., Hossein D.; Jaskaran S. et al. Industrial Artificial Intelligence for industry 4.0-based manufacturing systems // *Manufacturing letters*. – 2018. – Vol. 6. – N 18. – P. 20–23.

Lee H., Choi B. Knowledge management enablers, processes, and organizational performance: An integrative view and empirical examination // *Journal of management information systems*. – 2003. – Vol. 20. – N 1. – P. 179–228.

Liakos K., Busato P., Moshou D. et al. Machine learning in agriculture: A review // *Sensors*. – 2018. – Vol. 18. – N 8. Режим доступа: file:///C:/Users/Super%20User/Downloads/sensors-18-02674%20(1).pdf

Liew A. Understanding data, information, knowledge and their inter-relationships // *Journal of knowledge management practice*. – 2007. – Vol. 8. – N 2. – P. 1–16.

Liew A. DIKIW: Data, information, knowledge, intelligence, wisdom and their interrelationships // *Business Management Dynamics*. – 2013. – Vol. 2. – N 10. – P. 49–62.

Linnainmaa S. Taylor expansion of the accumulated rounding error // *BIT Numerical Mathematics*. – 1976. – Vol. 16. – N 2. – P. 146–160.

Liu S., Yang Y., Forrest J. *Grey Data Analysis. Methods, Models and Applications*. – Singapore: Springer, 2017. – 351 p.

Lokers R., Knapen R., Janssen S. et al. Analysis of Big Data technologies for use in agro-environmental science // *Environmental Modelling & Software*. – 2016. – Vol. 84. – P. 94–504.

March J. Exploration and exploitation in organizational learning // *Organization Science*. – 1991. – N 2. – P. 71–87.

Mitchell T.M. *The discipline of machine learning*. – Pittsburgh: Carnegie Mellon University, School of Computer Science, Machine Learning Department, 2006. – Vol. 9. – 12 p.

Morris C.W. *Foundations of the Theory of Signs* // *International encyclopedia of unified science*. – Chicago University Press, 1938. – Vol. 1. – N 2. – P. 1–59.

Murphy K.P. *Naive bayes classifiers* // *University of British Columbia*. – 2006. – Vol. 8. – P. 1–8.

Napoleone A., Macchi M., Pozzetti A. A review on the characteristics of cyber-physical systems for the future smart factories // *Journal of Manufacturing Systems*. – 2020. – Vol. 54. – P. 305–335.

Neapolitan R.E. Models for reasoning under uncertainty // Applied Artificial Intelligence an International Journal. – 1987. – Vol. 1. – N 4. – P. 337–366.

Nonaka I., Takeuchi H. The knowledge-creating company: How Japanese companies create the dynamics of innovation. – Oxford university press, 1995. – 299 p.

Oliveira M., Maçada A.C., Curado C. Knowledge management mechanisms within the SECI model // Atas da Conferência da Associação Portuguesa de Sistemas de Informação. – 2014. – Vol. 11. – P. 1–14.

OMG® Unified Modeling Language® (OMG UML®). Version 2.5.1. – Object Management Group, 2017. – 796 p.

Özdemir V., Hekim N. Birth of industry 5.0: Making sense of big data with artificial intelligence, “the internet of things” and next-generation technology policy // Omics: a journal of integrative biology. – 2018. – Vol. 22. – N 1. – P. 65–76.

Radanliev P., De Roure D.C., Nurse J.R.C. et al. New developments in Cyber Physical Systems, the Internet of Things and the Digital Economy-discussion on future developments in the Industrial Internet of Things and Industry 4.0. Preprints (www.preprints.org). – 2019. – 9 p.

Pearl J. Probabilistic Reasoning in Intelligent Systems. Morgan Kauffmann San Mateo. – 1988. – Vol. 88. – 552 p.

Polley D., Magnuson L. (Ed.). Visualizing the topical coverage of an institutional repository using VOSviewer. Data visualization: A guide to visual storytelling for librarians. – Lanham, MD: Rowman & Littlefield, 2016. – 27 p.

Quinlan J.R. Induction of decision trees // Machine learning. – 1986. – Vol. 1. – N 1. – P. 81–106.

Quinlan J.R. C4.5: Programs for Machine learning. – Morgan Kaufmann Publishers, 1993.

Riedmiller M., Braun H. A direct adaptive method for faster backpropagation learning: The RPROP algorithm // IEEE international conference on neural networks. IEEE, 1993. – P. 586–591.

Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain // Psychological review. – 1958. – Vol. 65. – N 6. – P. 386–408.

Rowley J. The wisdom hierarchy: representations of the DIKW hierarchy // Journal of information science. – 2007. – Vol. 33. – N 2. – P. 163–180.

Rumbaugh J., Jacobson I., Booch G. The unified modeling language for Object-Oriented Development // Reference manual, 1999. – 35 p. Режим доступа: <http://cssngce.synthasite.com/resources/UML.pdf>

Schapire R.E., Freund Y., Bartlett P. et al. Boosting the margin: A new explanation for the effectiveness of voting methods // The annals of statistics. – 1998. – Vol. 26. – N 5. – P. 1651–1686.

Schmidhuber J. Deep Learning in Neural Networks: An Overview // Neural Networks. – 2015. – Vol. 61. – P. 85–117.

Schreiber G., Wielinga B., Breuker J. (ed.). KADS: A principled approach to knowledge-based system development. Knowledge-Dased Systems. – Vol. 11. – Academic Press, 1993. – 455 p.

Schuh G., Anderl R., Gausemeier J. et al. Industrie 4.0 Maturity Index : die digitale Transformation von Unternehmen gestalten (Acatech Studie). – München: Herbert Utz Verlag GmbH, 2017. – 55 p.

Sensuse D.I., Cahyaningsih E. Knowledge Management Models: A Summative Review // International Journal of Information Systems in the Service Sector (IJISSS). – 2018. – Vol. 10. – N 1. – P. 71–100.

Sinclair D. A portable diffusion battery // American Industrial Hygiene Association Journal. – 1972. – Vol. 33. – N 11. – P. 729–735.

Sinclair D., Christy D.A., Snyder K. Calculation of aerosol size distribution from diffusion battery measurements—a computer program for the graphical” stripping” method // Aerosol Measurements. – University of Florida Press, 1979. – P. 615–631.

Singhal A. Introducing the Knowledge Graph: things, not strings. 2012. URL: <https://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html>

Stamper R., Liu K., Hafkamp M. et al. Understanding the Roles of Signs and Norms in Organisations // Journal of Behaviour & Information Technology. – 2000. – Vol. 19. – N 1. – P. 15–27.

Sterman J.D. System dynamics modeling: Tools for learning in a complex world // California Management Review. – 2001. – Vol. 43. – N 4. – P. 8–25.

Stewart T.A. The Wealth of Knowledge: Intellectual Capital and the Twenty-First Century Organization; Nicholas Brealey Publishing. – London, UK, 1997.

Šūmane S., Kunda I., Knickel K., et al. Local and farmers' knowledge matters! How integrating informal and formal knowledge enhances sustainable and resilient agriculture // Journal of Rural Studies. – 2018. – Vol. 59. – P. 232–241.

Tolle K.M., Tansley D.S. W., Hey A.J.G. The fourth paradigm: Data-intensive scientific discovery [point of view] // Proceedings of the IEEE. – 2011. – Vol. 99. – N 8. – P. 1334–1337.

Trivelli L., Apicella A., Chiarello F. et al. From precision agriculture to Industry 4.0: Unveiling technological connections in the agrifood sector // British Food Journal. – 2019. – Vol. 121. – N 8. – P. 1730–1743.

Tzortzaki A.M., Mihiotis A. A review of knowledge management theory and future directions // Knowledge and Process Management. – 2014. – Vol. 21. – N 1. – P. 29–41.

Van Eck N., Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping // Scientometrics. – 2010. – Vol. 84. – P. 523–538.

Van Eck N., Waltman L., Noyons E. et al. Automatic term identification for bibliometric mapping // Scientometrics. – 2010. – Vol. 82. – P. 581–596.

Van Eck N., Waltman L. Text mining and visualization using VOSviewer // ArXiv:1109.2058. 2011. <http://arxiv.org/abs/1109.2058>

Vom Brocke J., Rosemann M. Handbook on Business Process Management. Strategic Alignment, Governance, People and Culture. – New York: Springer Berlin Heidelberg (International handbooks on information systems), 2015. – 865 p.

Walter A., Finger R., Huber R. et al. Opinion: Smart farming is key to developing sustainable agriculture // Proceedings of the National Academy of Sciences. – 2017. – Vol. 114. – N 24. – P. 6148–6150.

Wang L., Wang G. Big data in cyber-physical systems, digital manufacturing and industry 4.0 // International Journal of Engineering and Manufacturing (IJEM). – 2016. – Vol. 6. – N 4. P. 1–8.

Welbank M. An overview of knowledge acquisition methods // Interacting with computers. – 1990. – Vol. 2. – N 1. – P. 83–91.

Wielinga B.J., Schreiber A.T., Breuker J.A. KADS: A modelling approach to knowledge engineering // Knowledge acquisition. – 1992. – Vol. 4. – N 1. – P. 5–53.

Wiig K. Knowledge Management Methods: Practical Approaches to Managing Knowledge. – Schema Press: Arlington, TX, USA, 1995. – 490 p.

Winklmayr W., Wang H.C., John W. Adaptation of the Twomey algorithm to the inversion of cascade impactor data // Aerosol Science and Technology. – 1990. – Vol. 13. – N 3. – P. 322–331.

Witten I.H., Frank E., Hall M.A., Pal C.J. Data Mining: Practical machine learning tools and techniques. – Morgan Kaufmann, 2016. – 621 p.

Wolfert S., Ge L., Verdouw C., Bogaardt M.-J. Big data in smart farming – a review // Agricultural Systems. – 2017. – Vol. 153. – P. 69–80.

Wortmann F., Flüchter K. Internet of things // Business & Information Systems Engineering. – 2015. – Vol. 57. – N 3. – P. 221–224.

Wu X. Zhu X., Wu G.Q. et al. Data mining with big data // IEEE transactions on knowledge and data engineering. – 2013. – Vol. 26. – N 1. – P. 97–107.

Zhu N., Liu X, Liu Z et al. Deep learning for smart agriculture: Concepts, tools, applications, and opportunities // International Journal of Agricultural and Biological Engineering. – 2018. – Vol. 11. – N 4. – P. 32–43.

МОНОГРАФИЯ

В.К. КАЛИЧКИН, Р.А. КОРЯКИН,
П.К. КУЦЕНОГИЙ

ИНТЕЛЛЕКТУАЛЬНЫЕ
СИСТЕМЫ В СЕЛЬСКОМ ХОЗЯЙСТВЕ
(теоретический анализ возможного приложения)

Корректор *Г.Н. Ягунова*
Дизайн, электронная верстка *Н.Ю. Бориско*
Дизайн обложки: *С.А. Косев*

Выход в свет 18.12.2020. Формат 60×84¹/₁₆.
Объем 18,5 печ. л. Тираж 500 экз. Заказ № 147

Отпечатано в Сибирском федеральном
научном центре агробιοтехнологий Российской академии наук
630501, Краснообск, Новосибирский район,
Новосибирская область, здание СФНЦА РАН, а/я 463