

УДК 303.732.4

UDC 303.732.4

АВТОМАТИЗИРОВАННЫЙ СИСТЕМО-КОГНИТИВНЫЙ АНАЛИЗ И ЕГО ПРИМЕНЕНИЕ ДЛЯ УПРАВЛЕНИЯ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИМИ СИСТЕМАМИ В АПК

AUTOMATED SYSTEM-COGNITIVE ANALYSIS AND ITS APPLICATION TO THE MANAGEMENT OF SOCIO-ECONOMIC SYSTEM IN AGRICULTURE

Луценко Евгений Вениаминович
д.э.н., к.т.н., профессор
Кубанский государственный аграрный университет, Россия, 350044, Краснодар, Калинина, 13,
prof.lutsenko@gmail.com

Lutsenko Eugeny Veniaminovich
Dr.Sci.Econ., Cand.Tech.Sci., professor
Kuban State Agrarian University, Krasnodar, Russia

Лойко Валерий Иванович
заслуженный деятель науки РФ,
д.т.н., профессор
Кубанский государственный аграрный университет, Россия, 350044, Краснодар, Калинина, 13,
loyko@kubagro.ru

Loiko Valery Ivanovich
deserved scientist of the Russian Fedration,
Dr.Sci.Tech., professor
Kuban State Agrarian University, Krasnodar, Russia

Макаревич Олег Александрович
к.э.н., доцент
Майкопский государственный технологический университет, Республика Адыгея, Россия

Makarevich Oleg Alexandrovich
Cand.Econ.Sci., associate professor
Adygh State Technological University, Maikop, Adyghea Republic, Russia

Макаревич Лилия Олеговна
соискатель
НОУ ВПО "Санкт-Петербургский институт внешнеэкономических связей, экономики и права", филиал в г. Краснодаре, Россия

Makarevich Liliya Olegovna
competitor
St.-Petersburg institute of foreign economic relations, economy and law, branch in Krasnodar, Krasnodar, Russia

В статье сформулированы проблема и задачи управления социально-экономическими системами в агропромышленном комплексе (АПК) и для их решения предложено применить автоматизированный системно-когнитивный анализ (АСК-анализ). Решение данной проблемы рассматривается на примере многоотраслевой интегрированной производственной системами АПК (МИПС АПК). Кратко рассмотрены теоретические основы, математическая модель, методика численных расчетов и программный инструментарий АСК-анализа, а также основные результаты и перспективы его применения для управления социально-экономическими системами в АПК

The article defines the problem and the points of socio-economic systems in the agro-industrial complex (APC) and proposes using automated system-cognitive analysis (ASCA). The solution to this problem is shown as an example of an integrated multi-industrial agricultural system (IMPI APC). Theoretical basis, mathematical model, technique of numerical calculations and software tools of ASC-analysis, and the main results and prospects of its application to control the socio-economic systems in agriculture are presented

Ключевые слова: СИСТЕМНЫЙ ПОДХОД, СИСТЕМО-КОГНИТИВНЫЙ АНАЛИЗ, МНОГООТРАСЛЕВАЯ КОРПОРАЦИЯ, ПРОГНОЗИРОВАНИЕ, СЕМАНТИЧЕСКАЯ ИНФОРМАЦИОННАЯ МОДЕЛЬ

Keywords: SYSTEMIC APPROACH, SYSTEM-COGNITIVE ANALYSIS, SYSTEMS APPROACH, DIVERSIFIED CORPORATION, FORECASTING, SEMANTIC INFORMATION MODEL

Введение

Агропромышленный комплекс (АПК) представляет собой реальный сектор экономики, во многом определяющий продовольственную безопасность и само существование страны (глобальная цель АПК). Государственная аграрная политика направлена на достижение глобальной цели

АПК путем согласования и корректировки целей хозяйств, отраслей и регионов с использованием в основном экономических рычагов.

Наработанные за прошедшие десятилетия экономические схемы хозяйствования, технологии производства, заготовки, хранения и переработки сельскохозяйственной продукции были ориентированы на приоритеты плановой затратной экономики, экономики недопроизводства, поэтому в настоящее время они быстро теряют адекватность и настоятельно требуют пересмотра. Разработка новых подходов к управлению в АПК, основанных на непредвзятом учете и признании складывающихся макроэкономических реалий, на критериях эффективности и качества рыночной экономики, требует срочных согласованных усилий ученых – аграриев и является одним из приоритетов экономической теории и практики.

Управление в АПК всегда представляло собой проблему, имеющую комплексную межотраслевую природу и состоящую как в рациональном выборе целей производства для конкретных хозяйств, так и в оценке достижимости этих целей и выборе путей их достижения.

Экономической целью каждого хозяйства в АПК в рыночных условиях безусловно является получение максимальной прибыли. Достичь этой цели можно путем производства и реализации на рынке наиболее рентабельных видов продукции в необходимых потребителям объемах. Поэтому для достижения экономической цели хозяйству необходимо решить следующие задачи:

1. Поставить предварительную цель производству – выпустить в рациональных объемах востребованные виды продукции, на основе прогнозирования спроса (конъюнктуры рынка) на сельскохозяйственную продукцию в номенклатуре в натуральном и стоимостном выражении.

2. Выработать реальную цель производства путем корректировки предварительной цели с учетом ограничений: прогнозируемых возможностей производственного потенциала хозяйства (прогнозирования предложения) при различной рентабельности видов продукции. Оценка производственного потенциала хозяйства должна осуществляться по каждому виду продукции в натуральном и стоимостном выражении по количественному и качественному критериям. Для этого необходимо:

а) для каждого пункта выращивания (зоны и подзоны) ранжировать культуры по степени соответствия их генотипа условиям данного пункта;

б) для каждой культуры ранжировать зоны и подзоны выращивания по степени соответствия их условий требованиям генотипа.

3. Выбрать пути достижения целей (способ управления), т.е. реализовать поддержку принятия решений по рациональному выбору агротехнологий (агротехнологических факторов), обеспечивающих перевод объ-

екта управления в заданное целевое состояние, соответствующее поставленной реальной цели производства.

Вышеперечисленные проблемы рассматриваются в работе с точки зрения проектирования информационных технологий и автоматизированных систем управления (АСУ). Решение этих проблем с помощью стандартных математических методов и инструментария математической экономики наталкивается на ряд сложностей:

- специфика объектов АПК, как объектов управления, а именно их слабодетерминированность, многофакторность и активность, а также малоисследованный характер реагирования объекта управления на управляющие факторы;

- комплексный многофакторный характер управляющих воздействий;

- очень большая длительность цикла управления (от одного года до 5-10 и более лет);

- неполнота (фрагментарность), неточность, зашумленность исходной информации;

- сложность доступа к исходной информации, в частности отсутствие или недоступность (часто из-за местнических или узкопонимаемых профессиональных интересов) электронных баз данных, которые могли бы послужить основой для разработки современных систем поддержки принятия управленческих решений.

Автоматизированный системно-когнитивный анализ (АСК-анализ) и его программный инструментарий представляют собой один из возможных подходов к решению сформулированных выше проблемы и задач. АСК-анализ представляет собой системно-когнитивный анализ (СК-анализ), рассматриваемый совместно с его программным инструментарием, в качестве которого в настоящее время выступает интеллектуальная система «Эйдос».

АСК-анализ позволяет непосредственно на основе эмпирических данных в количественной сопоставимой форме выявить силу и направление влияния значений факторов различной природы (измеряемых в различных единицах измерения) на переход моделируемого объекта в различные будущие состояния [1]. При этом метод обеспечивает обработку данных очень большой размерности, является непараметрическим и не требует полных повторностей всех сочетаний значений факторов и будущих состояний моделируемой системы, а также эффективно подавляет шум в исходных данных и выделяет полезный сигнал из шума [1]. АСК-анализ и его инструментарий – это метод исследования не зависящий от предмет-

ной области, т.е. применимый во многих предметных областях¹. Но перечисленные его особенности и возможности очень важны именно при моделировании социально-экономических систем, т.к. с ними невозможно планировать и проводить эксперименты (по крайней мере это очень рискованно и дорого) и исходные данные, описывающие подобные системы, обычно сильно зашумлены, т.е. имеют высокую погрешность.

Ниже рассмотрим:

1. Теоретические предпосылки АСК-анализа.
2. Математическую модель автоматизированного СК-анализа (системная теория информации).
3. Методику численных расчетов АСК-анализа.
4. Специальный программный инструментарий АСК-анализа – Интеллектуальную систему "Эйдос".
5. Опыт и перспективы применения АСК-анализа для управления социально-экономическими системами в АПК.

1. Теоретические предпосылки АСК-анализа

Системно-когнитивный анализ является новым и сравнительно малоизвестным математическим и инструментальным методом экономики, поэтому авторы посчитали целесообразным привести в данной статье краткую информацию об этом методе, полностью основываясь на работе [1], в которой предложен и впервые в развернутом виде описан этот метод.

Системно-когнитивный анализ представляет собой системный анализ, рассматриваемый как метод познания и структурированный по базовым когнитивным (познавательным) операциям (БКОСА). Сам набор БКОСА следует из формализуемой когнитивной концепции, рассматривающей процесс познания как многоуровневую иерархическую систему обработки информации, в которой когнитивные структуры каждого уровня являются результатом интеграции структур предыдущего уровня.

На 1-м уровне этой системы (рис. 1) находятся дискретные элементы потока чувственного восприятия, которые на 2-м уровне интегрируются в чувственный образ конкретного объекта. Те, в свою очередь, на 3-м уровне интегрируются в обобщенные образы классов и факторов, образующие на 4-м уровне кластеры, а на – 5-м конструкторы. Система конструкторов на 6-м уровне образуют текущую парадигму реальности (т.е. человек познает мир путем синтеза и применения конструкторов). На 7-м же уровне обнаруживается, что текущая парадигма не единственно-возможная (рис. 1).

¹ Достаточно сказать, что метод АСК анализа и его инструментарий – интеллектуальная система «Эйдос» были применены при проведении 11 диссертационных работ, в т.ч.: 3-х докторских по экономике, 2-х докторских по техническим наукам, 4-х кандидатских по психологическим наукам, 1-й кандидатской по техническим и 1-й по медицинским наукам и физиологии: <http://lc.kubagro.ru/aidos/index.htm>

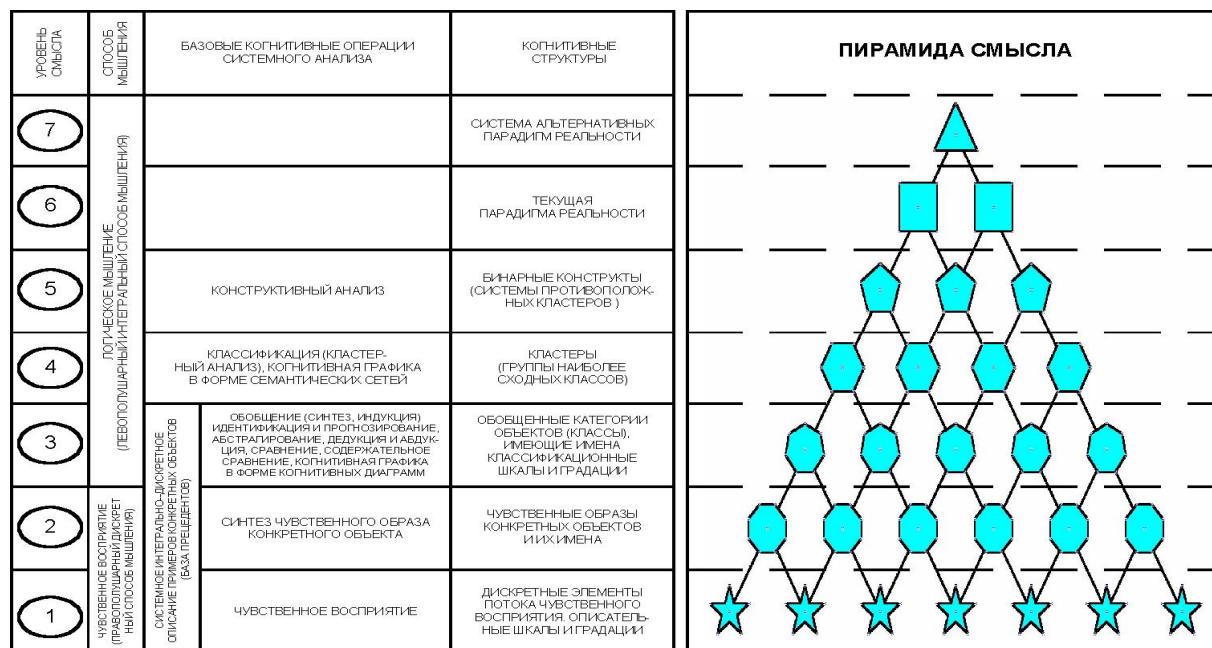


Рис. 1. Обобщенная схема формализуемой когнитивной концепции (иерархия базовых когнитивных операций)

Ключевым для когнитивной концепции является понятие факта – соответствие дискретного и интегрального элементов познания (т.е. элементов разных уровней интеграции-иерархии), обнаруженное на опыте. Факт рассматривается как квант смысла, что служит основой для его формализации. Таким образом, происхождение смысла связывается со своего рода "разностью потенциалов", существующей между смежными уровнями интеграции-иерархии обработки информации в процессах познания. В полном виде когнитивная концепция приведена в работе [1]. Из данной концепции выводятся структура когнитивного конфигуратора, система базовых когнитивных операций и обобщенная схема автоматизированного системного анализа, структурированного до уровня базовых когнитивных операций (СК-анализ) (рис. 2). Между когнитивными структурами разных уровней иерархии существует отношение "дискретное – интегральное", которое служит основой формализации смысла. Когнитивный конфигуратор представляет собой минимальную полную систему когнитивных операций, названных базовые когнитивные операции системного анализа. Всего выявлено 10 таких операций, каждая из которых достаточно элементарна для формализации и программной реализации. Перечислим их: присвоение имен; восприятие; обобщение (синтез, индукция); абстрагирование; оценка адекватности модели; сравнение, идентификация и прогнозирование; дедукция и абдукция; классификация и генерация конструктов; содержательное сравнение; планирование и принятие решений об управлении.

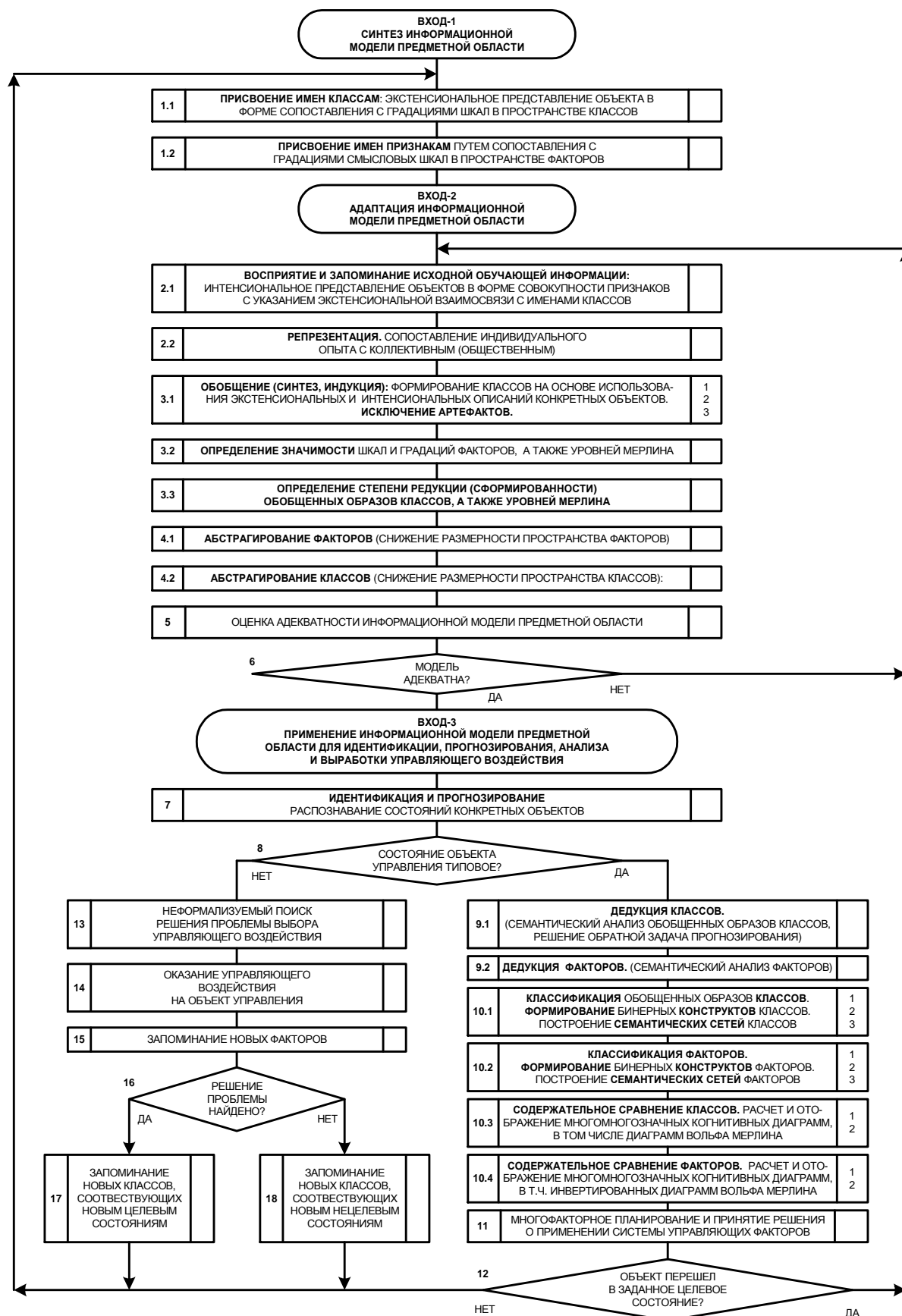


Рис. 2. Обобщенная схема системно-когнитивного анализа

В работе [1] предложены математическая модель, методика численных расчетов, включающая структуры данных и алгоритмы реализации БКОСА, а также программный инструментарий СК-анализа – универсальная когнитивная аналитическая система "Эйдос".

**2. Математическая модель автоматизированного СК-анализа:
системная теория информации**

2.1. Системное обобщение формулы Хартли

Классическая формула Хартли имеет вид:

$$I = \text{Log}_2 W. \tag{1}$$

Будем искать ее системное обобщение в виде:

$$I = \text{Log}_2 W^j, \tag{2}$$

где W – количество элементов в множестве;

φ – коэффициент эмерджентности, названный в [1] коэффициентом эмерджентности Хартли.

В работе [7] предлагается системное обобщение формулы Хартли в виде²:

$$I = \text{Log}_2 \sum_{m=1}^M C_W^m \tag{3}$$

где C_W^m – количество подсистем и m элементов;

m – сложность подсистем;

M – максимальная сложность подсистем.

Так как $C_W^1 = W$, то при $M=1$ система переходит в множество и выражение (3) приобретает вид (1), т.е. для него выполняется принцип соответствия, являющийся обязательным для более общей теории.

Учитывая, что при $M = W$:

$$\sum_{m=1}^M C_W^m = 2^W - 1 \tag{4}$$

² См., также <http://www.twirpx.com/file/780491/>

в этом случае:

$$I = \text{Log}_2(2^W - 1). \quad (5)$$

Выражение (5) дает приближенную оценку максимального количества информации в элементе системы. Из выражения (5) видно, что при увеличении числа элементов W количество информации I очень быстро стремится к W (6) и уже при $W > 4$ погрешность выражения (5) не превышает 1%:

$$\text{при } W \rightarrow \infty \quad I \rightarrow W \quad (6)$$

Приравняв правые части выражений (2) и (3):

$$I = \text{Log}_2 W^j = \text{Log}_2 \sum_{m=1}^M C_W^m \quad (7)$$

получим выражение для коэффициента эмерджентности Хартли:

$$j = \frac{\text{Log}_2 \sum_{m=1}^M C_W^m}{\text{Log}_2 W} \quad (8)$$

Смысл этого коэффициента раскрыт в работе [1]³. Здесь отметим лишь, что при $M \ll W$ система асимптотически переходит в множество $j \ll W$ и (2) \rightarrow (1), как и должно быть согласно принципу соответствия.

С учетом (8) выражение (2) примет вид:

$$I(W, M) = \text{Log}_2 W \frac{\text{Log}_2 \sum_{m=1}^M C_W^m}{\text{Log}_2 W}, \quad (9)$$

или при $M=W$ и больших W , учитывая (4 и 5):

$$I(W, M) = \text{Log}_2 W \frac{W}{\text{Log}_2 W} = W. \quad (10)$$

³ И в ряде других работ: <http://www.twirpx.com/file/780491/>

Выражение (9) и представляет собой искомое системное обобщение классической формулы Хартли, а выражение (10) – его достаточно хорошее приближение при большом количестве элементов в системе W .

2.2. Системное обобщение формулы Харкевича

Классическая формула А.Харкевича имеет вид:

$$I_{ij}(W, M) = \text{Log}_2 \frac{P_{ij}}{P_j}, \quad (11)$$

где P_{ij}	–	условная вероятность перехода объекта в j -е состояние при условии действия на него i -го значения фактора;
P_j	–	безусловная вероятность перехода объекта в j -е состояние (вероятность самопроизвольного перехода или вероятность перехода, посчитанная по всей выборке, т.е. при действии любого значения фактора).

Выражению (11) может быть придан следующий эквивалентный вид, который и будет использоваться ниже:

$$I_{ij}(W, M) = \text{Log}_2 \frac{P_{ij}}{P_i}, \quad (12)$$

- где i – признак (значение фактора): $1 \leq i \leq M$;
- j – состояние объекта или класс: $1 \leq j \leq W$;
- P_{ij} – условная вероятность наблюдения i -го значения фактора у объектов в j -го класса;
- P_i – безусловная вероятность наблюдения i -го значения фактора по всей выборке.

Из (12) видно, что формула Харкевича для семантической меры информации по сути является логарифмом от формулы Байеса для апостериорной вероятности (отношение условной вероятности к безусловной). Вопрос об эквивалентности выражений (11) и (12) рассмотрим ниже.

Известно, что классическая формула Шеннона для количества информации для неравновероятных событий преобразуется в формулу Хартли при условии, что события равновероятны, т.е. удовлетворяет фундаментальному принципу соответствия. Поэтому теория информации Шеннона справедливо считается обобщением теории Хартли для неравновероятных событий. Однако выражения (11) и (12) при подстановке в них реальных

численных значений вероятностей⁴ P_{ij} , P_j и P_i не дают количества информации в *битах*, т.е. для этого выражения не выполняется принцип соответствия, обязательный для более общих теорий. Возможно, в этом состоит причина довольно сдержанного, а иногда даже скептического отношения специалистов по теории информации Шеннона к семантической теории информации Харкевича.

Причина этого, согласно [1], состоит в том, что в выражениях (11) и (12) отсутствуют глобальные параметры, характеризующие размерность конкретных моделей W и M , т.е. А.Харкевич в своем выражении для количества информации не ввел зависимости от мощности пространства будущих состояний объекта W и количества значений факторов M , обуславливающих переход объекта в эти состояния.

В [1] поставлена и решена задача получить такое обобщение формулы Харкевича, которое бы удовлетворяло тому же самому принципу соответствия, что и формула Шеннона, т.е. преобразовывалось в формулу Хартли в предельном детерминистском равновероятном случае, когда каждому классу (состоянию объекта) соответствует один признак (значение фактора), и каждому признаку – один класс, и эти классы (а, значит и признаки), равновероятны, и при этом каждый фактор однозначно, т.е. детерминистским образом определяет переход объекта в определенное состояние, соответствующее классу и между классами и признаками существует взаимно-однозначное соответствие.

Для этой цели в [1] в выражение (12) введен коэффициент Ψ :

$$I_{ij}(W, M) = \text{Log}_2 \left(\frac{P_{ij}}{P_i} \right)^\Psi. \quad (13)$$

В [1] поставлена и решена задача нахождения такого выражения для коэффициента Ψ , названного в этой работе в честь А.Харкевича "коэффициентом эмерджентности Харкевича", которое обеспечивает выполнение для выражения (13) принципа соответствия с классической формулой Хартли (1) и ее системным обобщением (2 и 3) в равновероятном детерминистском случае.

Для этого вероятности P_{ij} , P_j и P_i рассматриваются как пределы, к которым стремятся относительные частоты при увеличении объема выборки, и требуется выразить их через абсолютные частоты наблюдения признаков по классам (табл. 1). В табл. 1 рамкой обведена область значений, переменные определены ранее.

⁴ Вероятности представляют собой пределы, к которым сходятся относительные частоты при неограниченном увеличении объема выборки. В данной работе мы вместо термина «относительные частоты» будем применять термин «вероятности» не забывая об этом различии.

Т а б л . 1 . Матрица абсолютных частот

		Классы					Сумма
		1	...	j	...	W	
Значения факторов	1	N_{11}		N_{1j}		N_{1W}	
	...						
	i	N_{i1}		N_{ij}		N_{iW}	$N_i = \sum_{j=1}^W N_{ij}$
	...						
	M	N_{M1}		N_{Mj}		N_{MW}	
Суммарное количество признаков				$N_j = \sum_{i=1}^M N_{ij}$			$N = \sum_{i=1}^M \sum_{j=1}^W N_{ij}$

2.3. Алгоритм формирования матрицы абсолютных частот

Объекты обучающей выборки описываются векторами (массивами) $\vec{L} = \{L_i\}$ имеющихся у них признаков:

$$\vec{L} = \{L_i\} = \begin{cases} n, & \text{если у объекта есть } i\text{-й признак встречается } n \text{ раз;} \\ 0, & \text{если у объекта нет } i\text{-го признака.} \end{cases}$$

Первоначально в матрице абсолютных частот все значения равны нулю. Затем организуется цикл по объектам обучающей выборки. Если у предъявленного объекта, относящегося к j -му классу есть i -й признак, то в терминах программирования:

$$N_{ij} = N_{ij} + 1; N_i = N_i + 1; N_j = N_j + 1; N = N + 1$$

Здесь можно провести очень интересную и важную аналогию между способом формирования матрицы абсолютных частот и работой многоканальной системы выделения полезного сигнала из шума. Представим себе, что все объекты, предъявляемые для формирования обобщенного образа некоторого класса, в действительности являются различными реализациями одного объекта – "Эйдоса" (в смысле философии Платона), по-разному зашумленного различными случайными обстоятельствами. И наша задача состоит в том, чтобы подавить этот шум и выделить из него то общее и существенное, что отличает объекты данного класса от объектов других классов. Учитывая, что шум чаще всего является "белым" и имеет свойство при суммировании с самим собой стремиться к нулю, а сигнал при этом, наоборот, возрастает пропорционально количеству слагаемых, увеличение

объема обучающей выборки приводит к тому, что отношение сигнал/шум в матрице абсолютных частот все время улучшается, т.е. происходит выделение полезной информации из шума. Примерно таким образом мы начинаем постепенно понимать смысл фразы, которую не сразу расслышали по телефону и несколько раз переспрашивали. При этом в повторах шум не позволяет понять то одну, то другую часть фразы, но, в конце концов, за счет использования памяти и интеллектуальной обработки информации мы понимаем ее всю. Так и объекты, описанные признаками, можно рассматривать как зашумленные фразы, несущие нам информацию об обобщенных образах классов – "Эйдосах", к которым они относятся. И эту информацию мы выделяем из шума при синтезе модели.

Для выражения (11):

$$P_{ij} = \frac{N_{ij}}{N_i}, \quad (14)$$

Для выражений (12) и (13):

$$P_{ij} = \frac{N_{ij}}{N_j}, \quad (15)$$

Для выражений (11), (12) и (13):

$$P_i = \frac{N_i}{N}; P_j = \frac{N_j}{N};$$

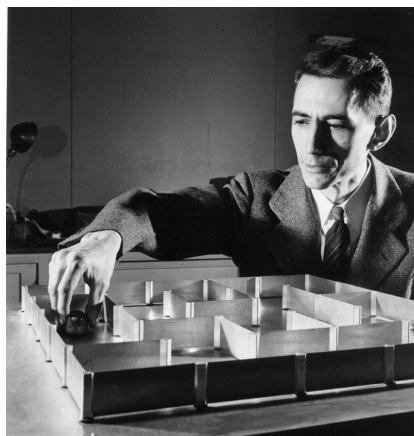
$$N_i = \sum_{j=1}^W N_{ij}; N_j = \sum_{i=1}^M N_{ij}; \quad (16)$$

$$N = \sum_{i=1}^M N_i = \sum_{j=1}^W N_j = \sum_{i=1}^W \sum_{j=1}^M N_{ij},$$

где N – суммарное количество встреч различных факторов у всех объектов исследуемой выборки;

N_i – суммарное количество встреч i -го фактора у всех объектов исследуемой выборки;

- N_j – суммарное количество встреч различных факторов у объектов, перешедших в j -е состояние;
- N_{ij} – суммарное количество наблюдений в исследуемой выборке факта: "действовало i -е значение фактора и объект перешел в j -е состояние".



Клод Шеннон в годы открытий⁵

Отметим, что К.Шеннон при выводе своей знаменитой формулы для средневзвешенного количества информации в сообщении, использовал другое, чем (16), выражение для вероятности, считая, что вероятность выборки определенного конкретного i -го символа из множества этих символов, мощностью N_i , составляет:

$$P_i = \frac{1}{N_i} \quad (a)$$

Представим себе, что в некотором сообщении i -й символ встречается N_i раз. Тогда при выборке конкретного i -го символа мы по Р.Хартли получаем:

$$I = \text{Log}N_i \quad (б)$$

информации. В среднем же при выборке i -го символа мы получаем:

$$I = \frac{\text{Log}N_i}{N_i} \quad (в)$$

Используя в выражении (в) формулу для вероятности выборки конкретного i -го символа (а) получим:

$$I = -P_i \text{Log}P_i \quad (г)$$

В выражении (г) для среднего количества информации в i -м символе учтено, что $\text{Log}N_i = -\text{Log}P_i$, т.к. $\text{Log} \frac{1}{N_i} = \text{Log}1 - \text{Log}N_i = -\text{Log}N_i$. Всего же в сообщении из N символов в среднем будет содержаться:

$$I = -\sum_{i=1}^N P_i \text{Log}P_i \quad (д)$$

Выражение (д) и представляет собой знаменитую формулу К.Шеннона для среднего количества информации в сообщении из N символов.

⁵ Источник: http://internetua.com/upload/tinymce/images/AleX101/shannon.mouse_in_maze.jpg

2.4. Формирование матрицы условных и безусловных вероятностей (относительных частот)

На основе анализа матрицы частот, приведенной в табл. 1, классы можно сравнивать по наблюдаемым частотам признаков только в том случае, если количество объектов по всем классам одинаково, как и суммарное количество признаков по классам. Если они отличаются, то корректно сравнивать классы можно только по условным и безусловным вероятностям наблюдения признаков, посчитанных на основе матрицы частот (см. табл. 1) в соответствии с выражениями (14) и (15), в результате чего получается матрица условных и безусловных вероятностей (процентных распределений) (табл. 2).

Т а б л . 2 . Матрица условных и безусловных вероятностей

		Классы					Безусловная вероятность признака
		1	...	j	...	W	
Значения факторов	1	P_{11}		P_{1j}		P_{1W}	
	...						
	i	P_{i1}		P_{ij}		P_{iW}	P_i
	...						
	M	P_{M1}		P_{Mj}		P_{MW}	
Безусловная вероятность класса				P_j			

При расчете матрицы условных и безусловных вероятностей N_j из табл. 1 берутся из последней строки, где N_j представляет собой "суммарное количество признаков у всех объектов, использованных для формирования обобщенного образа j -го класса". В результате получаем семантическую информационную модель (СИМ).

Эквивалентность выражений (11) и (12) устанавливается, если подставить в них выражения вероятности P_{ij} , P_j и P_i через частоты наблюдения признаков по классам из (14–16). В обоих случаях из выражений (11) и (12) получается одно и то же выражение:

$$I_{ij} = \text{Log}_2 \frac{N_{ij}N}{N_i N_j}, \tag{17}$$

а из (13) выражение (18), с которым мы и будем работать далее.

$$I_{ij} = \text{Log}_2 \left(\frac{N_{ij}N}{N_i N_j} \right)^\Psi \tag{18}$$

При взаимно-однозначном соответствии классов и признаков в равновероятном детерминистском случае имеем (табл. 3.):

Т а б л . 3 . Матрица частот в равновероятном детерминистском случае

		Классы					Сумма
		1	...	j	...	W	
Значения факторов	1	1					1
	...		1				1
	i			1			1
	...				1		1
	M					1	1
Сумма		1	1	1	1	1	N

В этом случае к каждому классу относится один объект, имеющий единственный признак. Откуда получаем для всех *i* и *j* (19):

$$\forall ij : N_{ij} = N_i = N_j = 1. \tag{19}$$

Таким образом, обобщенная формула А.Харкевича (18) с учетом (19) в этом случае приобретает вид:

$$I_{ij} = \text{Log}_2 N^\Psi = \text{Log}_2 W^j, \tag{20}$$

откуда:

$$\Psi = \frac{\text{Log}_2 W^j}{\text{Log}_2 N} \tag{21}$$

или, учитывая выражение для коэффициента эмерджентности Хартли (8): (22)

$$\Psi = \frac{\frac{\text{Log}_2 \sum_{m=1}^M C_W^m}{\text{Log}_2 W}}{\text{Log}_2 N}.$$

Подставив коэффициент эмерджентности Харкевича (21) в выражение (18), получим:

$$\begin{aligned} I_{ij} &= \text{Log}_2 \left(\frac{N_{ij} N}{N_i N_j} \right)^\Psi = \text{Log}_2 \left(\frac{N_{ij} N}{N_i N_j} \right)^{\frac{\text{Log}_2 W^j}{\text{Log}_2 N}} = \\ &= \frac{\text{Log}_2 W^j}{\text{Log}_2 N} \left(\text{Log}_2 \left(\frac{N_{ij}}{N_i N_j} \right) + \text{Log}_2 N \right) = \\ &= \text{Log}_2 \left(\frac{N_{ij}}{N_i N_j} \right)^{\frac{\text{Log}_2 W^j}{\text{Log}_2 N}} + \text{Log}_2 W^j \end{aligned}$$

или окончательно:

$$I_{ij} = \text{Log}_2 \left(\frac{N_{ij}}{N_i N_j} \right)^{\frac{\text{Log}_2 W^j}{\text{Log}_2 N}} + \text{Log}_2 W^j. \quad (23)$$

Отметим, что первая задача получения системного обобщения формул Хартли и Харкевича и вторая задача получения такого обобщения формулы Харкевича, которая удовлетворяет принципу соответствия с формулой Хартли, – это две разные задачи. Первая задача является более общей и при ее решении, которое приведено выше, автоматически решается и вторая задача, которая является, таким образом, частным случаем первой задачи.

Однако представляет самостоятельный интерес и частный случай, в результате которого получается формула Харкевича, удовлетворяющая в равновероятном детерминистском случае принципу соответствия с класси-

ческой формулой Хартли (1), а не с ее системным обобщением (2) и (3). Ясно, что эта формула получается из (23) при $j = 1$.

$$I_{ij} = \text{Log}_2 \left(\frac{N_{ij}}{N_i N_j} \right)^{\frac{\text{Log}_2 W}{\text{Log}_2 N}} + \text{Log}_2 W. \quad (24)$$

Из выражений (21) и (22) видно, что в этом частном случае, т.е. когда система эквивалентна множеству ($M=1$), коэффициент эмерджентности Харкевича приобретает вид:

$$\Psi = \frac{\text{Log}_2 W}{\text{Log}_2 N}. \quad (25)$$

На практике для численных расчетов удобнее пользоваться не выражениями (23) или (24), а формулой (26), которая получается непосредственно из (18) после подстановки в него выражения (25):

$$I_{ij} = \frac{\text{Log}_2 W}{\text{Log}_2 N} \times \text{Log}_2 \frac{N_{ij} N}{N_i N_j} \quad (26)$$

В классическом анализе Шеннона идет речь лишь о передаче символов по одному информационному каналу от одного источника к одному приемнику. При этом исследуется прежде всего передача самого сообщения.

В данной работе решается другая задача – идентифицировать или распознать информационный источник по сообщению от него. Поэтому метод Шеннона был обобщен путем учета в математической модели возможности существования многих источников информации, от которых к приемнику по зашумленному каналу связи приходят не отдельные символы-признаки, а сообщения, состоящие из последовательностей символов (признаков) любой длины.

Следовательно, ставится задача идентификации информационного источника по сообщению от него, полученному приемником по зашумленному каналу. Метод, являющийся обобщением метода Шеннона, позволяет применить классическую теорию информации для построения моделей си-

стем распознавания образов и принятия решений, ориентированных на применение для синтеза адаптивных АСУ сложными объектами.

Для решения поставленной задачи необходимо вычислять не средние информационные характеристики, как в теории Шеннона, а количество информации, содержащееся в конкретном i -м признаке (символе) о том, что он пришел от данного j -го источника информации. Это позволит определить и суммарное количество информации в сообщении о каждом информационном источнике, что дает интегральный критерий для идентификации или прогнозирования состояния объекта.

Логично предположить, что среднее количество информации, содержащейся в системе признаков о системе классов

$$I(Y, X) = \sum_{j=1}^W \sum_{i=1}^M p_{ij} \text{Log}_2 \frac{p_{ij}}{p_i p_j}, \quad (27)$$

является ничем иным, как усреднением (с учетом условной вероятности наблюдения) "индивидуальных количеств информации", которые содержатся в конкретных признаках о принадлежности обладающих ими объектов к конкретным классам (источникам), т.е.:

$$i(x_j, y_i) = \text{Log}_2 \frac{p_{ij}}{p_i p_j}. \quad (28)$$

Это выражение определяет так называемую плотность информации, т.е. количество информации, которое содержится в одном отдельно взятом факте наблюдения i -го символа (признака) на приемнике о том, что этот символ (признак) послан j -м источником.

Если в сообщении содержится M символов, то суммарное количество информации о принадлежности данного сообщения j -му информационному источнику (классу) составляет:

$$i(x_j) = \sum_{i=1}^M \text{Log}_2 \frac{p_{ij}}{p_i p_j}. \quad (29)$$

Необходимо отметить, что применение сложения в выражении (29) вполне корректно и оправданно, так как информация с самого начала вводилась как аддитивная величина, для которой операция сложения является корректной.

Преобразуем выражение (29) к виду, более удобному для применения на практике для численных расчетов. Для этого традиционным для теории информации Шеннона способом выразим вероятности встреч признаков через частоты их наблюдения:

$$P_{ij} = \frac{1}{N_{ij}}; P_i = \frac{1}{N_i}; P_j = \frac{1}{N_j}. \quad (30)$$

Подставив (30) в (29), получим:

$$i(x_j) = \sum_{i=1}^M \text{Log}_2 \frac{N_{ij}}{N_i N_j}. \quad (31)$$

Если ранжировать классы в порядке убывания суммарного количества информации о принадлежности к ним, содержащейся в данном сообщении (т.е. описании объекта), и выбирать первый из них, т.е. тот, о котором в сообщении содержится наибольшее количество информации, то мы получим обоснованную статистическую процедуру, основанную на классической теории информации, оптимальность которой доказывается в фундаментальной лемме Неймана–Пирсона [1].

Подставим значения вероятностей из (30) в (28) и получим выражение для плотности информации Шеннона, выраженное не через вероятности, а через частоты наблюдения символов, которые рассматриваются как признаки объектов, т.е. количество информации, содержащееся в отдельном i -м признаке о том, что другом конце канала связи находится j -й объект:

$$i(x_j, y_i) = \text{Log}_2 \frac{N_{ij}}{N_i N_j}. \quad (32)$$

Сравнивая выражения (23) и (32) видим, что в системном обобщении формулы Харкевича первое слагаемое практически тождественно выражению Шеннона для плотности информации, а второе слагаемое представляет собой плотность информации по Хартли.

Различия состоят в том, что в выражении (23) это слагаемое возведено в степень, имеющую смысл коэффициента эмерджентности Харкевича. Поэтому вполне оправданным называть это слагаемое не коэффициентом эмерджентности Харкевича, а коэффициентом эмерджентности Шен-

нона–Харкевича. Необходимо отметить также, что значения частот в этих формулах связаны с вероятностями несколько различным образом (выражения 14–16 и 30).

Из этого также следует, что *выражение (23) представляет собой нелинейную суперпозицию выражений для плотности информации Шеннона и Хартли*, и, таким образом, является обобщающим выражением для плотности информации, которое при различных условиях асимптотически переходит в классические выражения Хартли и Харкевича, а от выражения Шеннона отличается лишь константой, т.е. вторым слагаемым, характеризующим мощность множества состояний объекта в модели.

Это позволяет обоснованно высказать гипотезу о том, что системная теория информации (СТИ), базирующаяся на выражении (23) для плотности информации, является более общей, чем теории Хартли, Шеннона и Харкевича и асимптотически связана с ними через принцип соответствия (рис. 4).

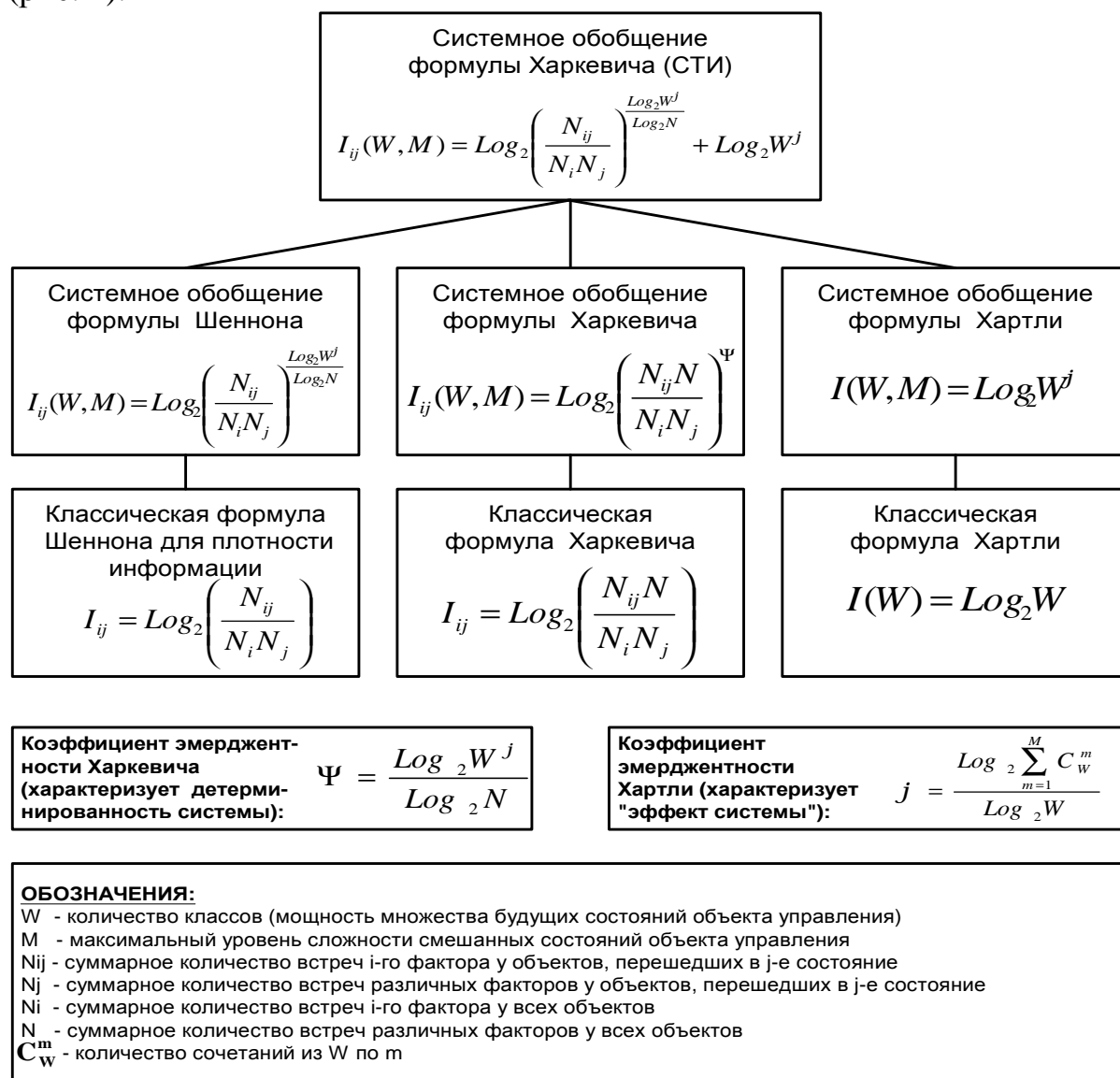


Рис. 3. Генезис системной (эмерджентной) теории информации

**2.5. Взаимосвязь системной меры
целесообразности информации со статистикой χ^2
и новая мера уровня системности предметной области**

Статистика χ^2 представляет собой сумму вероятностей совместного наблюдения признаков и объектов по всей корреляционной матрице или определенным ее подматрицам (т.е. сумму относительных отклонений частот совместного наблюдения признаков и объектов от среднего):

$$C^2 = \sum_{j=1}^W \sum_{i=1}^M \frac{(N_{ij} - t)^2}{t}, \quad (33)$$

где N_{ij} – фактическое количество встреч i -го признака у объектов j -го класса;

t – ожидаемое количество встреч i -го признака у объектов j -го класса.

$$t = \frac{N_i N_j}{N}. \quad (34)$$

Отметим, что статистика χ^2 математически связана с количеством информации в системе признаков объекта о его принадлежности к классу распознавания в соответствии с системным обобщением формулы Харкевича для плотности информации (18):

$$I_{ij} = \text{Log}_2 \left(\frac{N_{ij} N}{N_i N_j} \right)^\Psi, \quad (35)$$

а именно, из (34) и (35) получаем:

$$I_{ij} = \text{Log}_2 \left(\frac{N_{ij}}{t} \right)^\Psi. \quad (36)$$

Из (36) очевидно:

$$I_{ij} = \Psi (\text{Log}_2 N_{ij} - \text{Log}_2 t). \quad (37)$$

Сравнивая выражения (33) и (37), видим, что числитель в выражении (33) под знаком суммы отличается от выражения (37) только тем, что в выражении (37) вместо значений N_{ij} и t взяты их логарифмы. Учитывая, что логарифм является монотонно возрастающей функцией аргумента, то введение логарифма не меняет общего характера поведения функции.

Фактически это означает, что:

$$\begin{cases} \text{если } N_{ij} < t \text{ то } c_{ij} > 0, I_{ij} < 0 \\ \text{если } N_{ij} = t \text{ то } c_{ij} = 0, I_{ij} = 0 \\ \text{если } N_{ij} > t \text{ то } c_{ij} > 0, I_{ij} > 0 \end{cases} \quad (38)$$

Из вышеизложенного следует интерпретация системной меры информации (35) с учетом статистики χ^2 (33): если фактическая вероятность наблюдения i -го признака при предъявлении объекта j -го класса равна ожидаемой (средней), то наблюдение этого признака не несет никакой информации о принадлежности объекта к данному классу. Если она выше средней, то это свидетельствует о том, что предъявлен объект данного класса, если ниже – то другого.

Поэтому наличие статистической связи (информации) между признаками и классами распознавания, т.е. отличие вероятностей их совместных наблюдений от предсказываемого в соответствии со случайным нормальным распределением, приводит к увеличению фактической статистики χ^2 по сравнению с теоретической величиной.

Это дает основания говорить о возможности использования в качестве количественной меры степени выраженности закономерностей в предметной области не матрицы абсолютных частот и меры χ^2 , а новой меры H , основанной на матрице информативностей и системном обобщении формулы Харкевича для количества информации:

$$H = \sqrt[2]{\frac{1}{(W \cdot M - 1)} \sum_{j=1}^W \sum_{i=1}^M (I_{ij} - \bar{I})^2}, \quad (39)$$

где

$$\bar{I} = \frac{1}{W \cdot M} \sum_{j=1}^W \sum_{i=1}^M I_{ij} \quad \text{– средняя информативность признаков по матрице информативностей.}$$

Меру H в выражении (39) в [7] предлагается назвать обобщенным критерием степени сформированности модели Харкевича. Значение данной меры показывает среднее отличие количества информации в факторах

о будущих состояниях активного объекта управления от среднего количества информации в факторе (которое при больших выборках близко к 0). По своей математической форме эта мера сходна с мерами для значимости (интегральной информативности) факторов и степени сформированности образов классов и коррелирует с объемом неортонормированного семантического информационного пространства классов и семантического информационного пространства атрибутов.

С помощью вышеописанной математической модели достигается инвариантность результатов ее синтеза относительно следующих параметров обучающей выборки: суммарное количество и порядок ввода анкет обучающей выборки;

– количество анкет обучающей выборки по каждому классу распознавания;

– суммарное количество признаков во всех анкетах обучающей выборки;

– суммарное количество признаков по классам распознавания; количество признаков и их порядок в отдельных анкетах обучающей выборки.

В результате обеспечивается высокая степень качества решения задач распознавания на неполных и разнородных (в вышеперечисленных аспектах) данных как обучающей, так и распознаваемой выборки, т.е. при таких статистических характеристиках потоков этих данных, которые чаще всего и встречаются на практике и которыми невозможно или очень сложно управлять.

2.6. Получение матрицы знаний (информативностей)

На основе анализа матрицы условных и безусловных вероятностей (табл. 2) наблюдений признаков по классам и всей выборке можно сравнивать признаки друг с другом по их роли для сравнения классов друг с другом и конкретных объектов с обобщенными классами. При этом существует три основных группы признаков:

Г р у п п а п е р в а я . К этой группе относятся признаки, которые в одном классе встречаются, а в других нет. Это детерминистские признаки; обнаружение такого признака у объекта однозначно определяет его принадлежность к соответствующему классу.

Г р у п п а в т о р а я . В этой группе объединены признаки, которые в одном классе встречаются чаще, чем в других. Это статистические признаки; обнаружение такого признака у объекта несет некоторую информацию о его принадлежности к соответствующему классу.

Г р у п п а т р е т ь я . К этой группе относятся признаки, которые в разных классах встречаются с одной и той же вероятностью. Обнаружение этих признаков у объекта не несет никакой информации о его принадлежности к тем или иным классам.

Таким образом, если, используя данные табл. 2, анализировать условные вероятности (или процентные распределения) признаков по классам, то можно вынести правдоподобные суждения о принадлежности объектов, обладающих этими признакам к тем или иным классам.

Однако такой метод сравнения имеет по крайней мере два существенных недостатка.

1. Для того чтобы отнести признак к одной из вышеперечисленных групп, нужно сравнивать вероятности его наблюдения по классам, т.е. каждый раз при таком сравнении выполнять соответствующую необходимую для этого работу.

2. Отнесение признака ко второй группе еще не позволяет использовать этот признак для идентификации объекта; необходимо оценить количество информации, которое содержится в факте обнаружения у объекта этого признака о принадлежности этого объекта к каждому из классов, что требует применения математического и численного методов.

По поводу первого недостатка можно сказать, что для реальных задач большой размерности выполнение этого сравнения вручную практически невозможно, а значит невозможно и использование результатов этого сравнения для решения задач идентификации, прогнозирования и поддержки принятия решений, а тем более для исследования предметной области путем изучения ее модели. Все это обусловлено тем, что результат сравнения вероятностей встречи признака по классам не представляется при ручной обработке в количественной форме некоторого одного числа: частного критерия, величина и знак которого отражали бы результат такого сравнения.

Второй недостаток преодолевается методом, который предложен А.Харкевичем в выражениях (11) и (12) и уточнен в работе [7] в системном обобщении этих выражений (18). В этом методе предложено сравнивать не условные вероятности наблюдения признаков по различным классам друг с другом, а условную вероятность наблюдения признака по классу с безусловной вероятностью его наблюдения по всей выборке.

Это предложение по сути полностью соответствует известному статистическому методу отклонений от средних и нормативному подходу, когда в качестве базы сравнения выбирается норма, т.е. среднее по всей группе. На основе этого подхода формируются и критерии сравнения, т.е. можно сказать, что критериальный подход изначально основан на нормативном.

Если такое сравнение провести по всем признакам и классам, то получится матрица, устраняющая оба указанных недостатка: используя выражение (18) и данные табл. 1 непосредственно прямым счетом получаем матрицу знаний (табл. 4):

Т а б л . 4 . Матрица знаний (информативностей)

		Классы					Значимость фактора
		1	...	j	...	W	
Значения факторов	1	I_{11}		I_{1j}		I_{1W}	$S_1 = \sqrt{\frac{1}{W-1} \sum_{j=1}^W (I_{1j} - \bar{I}_1)^2}$
	...						
	i	I_{i1}		I_{ij}		I_{iW}	$S_i = \sqrt{\frac{1}{W-1} \sum_{j=1}^W (I_{ij} - \bar{I}_i)^2}$
	...						
	M	I_{M1}		I_{Mj}		I_{MW}	$S_M = \sqrt{\frac{1}{W-1} \sum_{j=1}^W (I_{Mj} - \bar{I}_M)^2}$
Степень редукции класса		σ_1		σ_j		σ_W	$H = \sqrt{\frac{1}{(W \cdot M - 1)} \sum_{j=1}^W \sum_{i=1}^M (I_{ij} - \bar{I})^2}$

Здесь \bar{I}_i – это среднее количество знаний в i -м значении фактора:

$$\bar{I}_i = \frac{1}{W} \sum_{j=1}^W I_{ij}.$$

Количественные значения коэффициентов I_{ij} табл. 4 являются знаниями о том, что "объект перейдет в j -е состояние" если "на объект действует i -е значение фактора".

Принципиально важно, что эти весовые коэффициенты не определяются экспертами на основе опыта интуитивным неформализуемым способом, а рассчитываются непосредственно по эмпирическим данным на основе теоретически обоснованной модели, хорошо зарекомендовавшей себя на практике при решении широкого круга задач в различных предметных областях.

Когда количество информации $I_{ij} > 0$ – i -й фактор способствует переходу объекта управления в j -е состояние, когда $I_{ij} < 0$ – препятствует этому переходу, когда же $I_{ij} = 0$ – никак не влияет на это. В векторе i -го фактора (строка матрицы информативностей) отображается, какое количество информации о переходе объекта управления в каждое из будущих состояний содержится в том факте, что данный фактор действует. В векторе j -го состояния класса (столбец матрицы информативностей) отображается, какое количество информации о переходе объекта управления в соответствующее состояние содержится в каждом из факторов.

Таким образом, матрица знаний (информативностей), приведенная в табл. 4, является обобщенной таблицей решений, в которой входы (факторы) и выходы (будущие состояния объекта управления) связаны друг с другом не с помощью классических (Аристотелевых) импликаций, принимающих только значения: "истина" и "ложь", а различными значениями истинности, выраженными в битах, и принимающими значения от положительного теоретически-максимально-возможного ("максимальная степень истинности"), до теоретически неограниченного отрицательного ("степень ложности").

Фактически предложенная модель позволяет осуществить синтез обобщенных таблиц решений для различных предметных областей непосредственно на основе эмпирических исходных данных и продуцировать прямые и обратные правдоподобные (нечеткие) логические рассуждения по неклассическим схемам с различными расчетными значениями истинности, являющимися обобщением классических импликаций.

Таким образом, данная модель позволяет рассчитать, какое количество информации содержится в любом факте о наступлении любого события в любой предметной области, причем для этого не требуется повторности этих фактов и событий. Если данные повторности осуществляются и при этом наблюдается некоторая вариабельность значений факторов, обуславливающих наступление тех или иных событий, то модель обеспечивает многопараметрическую типизацию, т.е. синтез обобщенных образов классов или категорий наступающих событий с количественной оценкой степени и знака влияния на их наступление различных значений факторов. Причем эти значения факторов могут быть как количественными, так и качественными и измеряться в любых единицах измерения, в любом случае в модели оценивается количество информации, которое в них содержится о наступлении событий, переходе объекта управления в определенные состояния или просто о его принадлежности к тем или иным классам.

2.7. Решение задач идентификации (распознавания образов), прогнозирования и поддержки принятия решений

Данная модель позволяет прогнозировать поведение объекта управления при воздействии на него не только одного, но и целой системы факторов:

$$I_j = f(I_{ij}). \quad (40)$$

В теории принятия решений скалярная функция I_j векторного аргумента называется интегральным критерием. Основная проблема состоит в выборе такого аналитического вида функции интегрального критерия, ко-

торый обеспечил бы эффективное решение сформулированной выше задачи АСУ.

Учитывая, что смысл частных критериев (18) состоит в том, что они содержат информацию, а информация по определению является аддитивной функцией, предлагается ввести интегральный критерий, как аддитивную функцию от частных критериев в виде:

$$I_j = (\overset{\mathbf{1}}{I}_{ij}, \overset{\mathbf{1}}{L}_i). \quad (41)$$

В выражении (41) круглыми скобками обозначено скалярное произведение, т.е. свертка. В координатной форме это выражение имеет вид:

$$I_j = \sum_{i=1}^M I_{ij} L_i, \quad (42)$$

где $\overset{\mathbf{1}}{I}_{ij} = \{I_{ij}\}$ – вектор j -го класса-состояния объекта управления;
 вектор состояния предметной области, включающий все
 $\overset{\mathbf{1}}{L}_i = \{L_i\}$ – виды факторов, характеризующих объект управления,
 возможные управляющие воздействия и окружающую
 среду (массив-локатор), т.е.:
 $\overset{\mathbf{r}}{L}_i = \begin{cases} 1, \text{ если } i\text{-й фактор действует;} \\ a_i, \text{ где } 0 < a_i < 1, \text{ если } i\text{-й фактор действует с истинностью } a_i \\ 0, \text{ если } i\text{-й фактор не действует.} \end{cases}$

В простейшем варианте значения координат вектора состояния предметной области принимаются равными либо 1 (действует i -е значение фактора a), либо 0 (фактор не действует). В более развитом варианте i -е значение фактора a интерпретируется как i -я градация описательной шкалы a , т.е. как i -й признак, поэтому в этой интерпретации можно считать, что $L_i = n$, если у объекта обучающей выборки i -й признак встречается n раз. В системе «Эйдос» реализован именно этот вариант.

Таким образом, интегральный критерий представляет собой суммарное количество информации, содержащееся в системе значений факторов различной природы (т.е. факторов, характеризующих объект управления, управляющее воздействие и окружающую среду) о переходе объекта управления в то или иное будущее состояние.

В многокритериальной постановке задача прогнозирования состояния объекта управления при оказании на него заданного многофакторного

управляющего воздействия I_j сводится к максимизации интегрального критерия:

$$j^* = \arg \max_{j \in J} ((I_j, L_i)), \quad (43)$$

т.е. к выбору такого состояния объекта управления, для которого интегральный критерий максимален.

Задача принятия решения о выборе наиболее эффективного управляющего воздействия является обратной задачей по отношению к задаче максимизации интегрального критерия (идентификации и прогнозирования), т.е. вместо того, чтобы по набору факторов прогнозировать будущее состояние объекта, наоборот, по заданному (целевому) состоянию объекта определяется такой набор факторов, который с наибольшей эффективностью перевел бы объект управления в это состояние.

Предлагается еще одно обобщение фундаментальной леммы Неймана–Пирсона, основанное на косвенном учете корреляций между информативностями в векторе состояний при использовании средних по векторам. Соответственно вместо простой суммы количеств информации предлагается использовать корреляцию между векторами состояния и объекта управления, которая количественно измеряет степень сходства этих векторов:

$$I_j = \frac{1}{S_j S_l A} \sum_{i=1}^M (I_{ij} - \bar{I}_j) (L_i - \bar{L}), \quad (44)$$

где \bar{I}_j – средняя информативность по вектору класса;

\bar{L} – среднее по вектору идентифицируемой ситуации (объекта);

S_j – среднеквадратичное отклонение информативностей вектора класса;

S_l – среднеквадратичное отклонение по вектору распознаваемого объекта.

Выражение (44) получается непосредственно из (42) после замены координат перемножаемых векторов их стандартизированными значениями:

$$I_{ij} \rightarrow \frac{I_{ij} - \bar{I}_j}{S_j}, \quad L_i \rightarrow \frac{L_i - \bar{L}}{S_l}.$$

Необходимо отметить, что выражение для интегрального критерия сходства (42) по своей математической форме является корреляцией двух векторов. Это означает, что если эти векторы являются суммой двух сигналов: полезного и белого шума, то при расчете интегрального критерия белый шум практически не будет играть никакой роли, т.е. его корреляция с самими собой равна нулю по определению. Поэтому интегральный критерий сходства объекта со случайным набором признаков с любыми образами классов, или реального объекта с образами классов, сформированными случайным образом, будет равен нулю. Это означает, что выбранный интегральный критерий сходства является высокоэффективным средством подавления белого шума и выделения полезной информации из шума, который неизбежно присутствует в эмпирических данных.

Важно также отметить неметрическую природу предложенного интегрального критерия сходства, благодаря чему его применение является корректным и при неортонормированном семантическом информационном пространстве, каким оно в подавляющем количестве случаев и является, т.е. в общем случае.

Результат прогнозирования поведения объекта управления, описанного данной системой факторов, представляет собой список его возможных будущих состояний, в котором они расположены в порядке убывания суммарного количества информации о переходе объекта управления в каждое из них.

2.8. Сравнение, идентификация и прогнозирование как разложение векторов объектов в ряд по векторам классов (объектный анализ)

Ранее были введены неметрические интегральные критерии сходства объекта, описанного массивом-локатором L_i с обобщенными образами классов I_{ij} (выражения 40–42).

Для непрерывного случая выражение (42) принимает вид:

$$I_j = \int_1^M L(i) I_j(i) di \quad (45)$$

Таким образом, выражение (45) представляет собой обобщение интегрального критерия сходства конкретного объекта и обобщенного класса (42) для непрерывного случая в координатной форме.

Отметим, что коэффициенты ряда Фурье (24) по своей математической форме и смыслу сходны с ненормированными коэффициентами корреляции, т.е. по сути скалярными произведениями для непрерывных функций в координатной форме: выражением (45) между разлагаемой в ряд кривой $f(x)$ и функциями Sin и Cos различных частот и амплитуд [7].

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \cos\left(\frac{np x}{L}\right) + b_n \sin\left(\frac{np x}{L}\right) \right),$$

где :

$$a_0 = \frac{1}{L} \int_{-L}^L f(x) dx; \tag{46}$$

$$a_n = \frac{1}{L} \int_{-L}^L f(x) \cos\left(\frac{np x}{L}\right) dx;$$

$$b_n = \frac{1}{L} \int_{-L}^L f(x) \sin\left(\frac{np x}{L}\right) dx,$$

где $n = \{1, 2, 3, \dots\}$ – натуральное число.

Сравнение выражений (45) и (46) позволяет сделать вывод о том, что процесс идентификации и прогнозирования (распознавания), реализованный в предложенной математической модели, может рассматриваться как разложение вектора-локатора распознаваемого объекта в ряд по векторам информативностей классов распознавания (которые представляют собой произвольные функции, сформированные при синтезе модели путем многопараметрической типизации на основе эмпирических данных).

Например, представим результаты идентификации на рисунке 5:

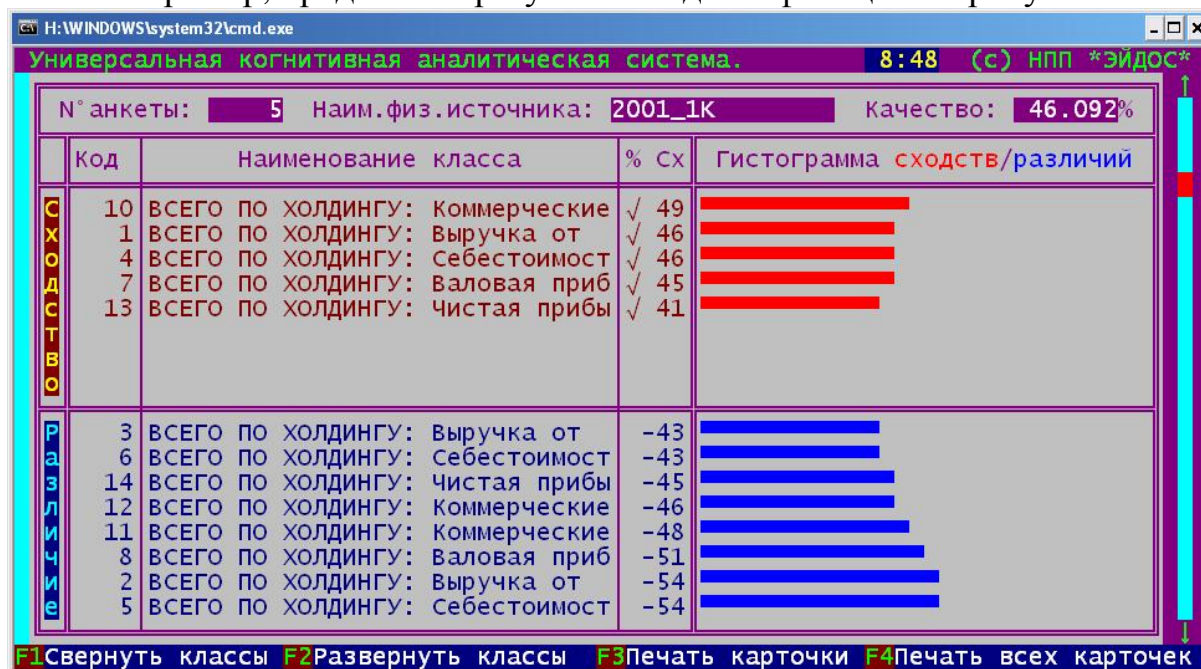


Рис. 4. Пример разложения профиля состояния МИПС АПК, сложившегося в 2001 г., в ряд по образам классов

Продолжая развивать аналогию с разложением в ряд, данный результат идентификации можно представить в векторной аналитической форме:

$\dot{K}_{2001} = 0,49 \cdot \dot{I}(10) + 0,46 \cdot \dot{I}(1) + 0,46 \cdot \dot{I}(4) + 0,45 \cdot \dot{I}(7) + 0,41 \cdot \dot{I}(13) - 0,43 \cdot \dot{I}(3) - 0,43 \cdot \dot{I}(6) - 0,45 \cdot \dot{I}(14) - 0,46 \cdot \dot{I}(12) - 0,48 \cdot \dot{I}(11) - 0,51 \cdot \dot{I}(8) - 0,54 \cdot \dot{I}(2) - 0,54 \cdot \dot{I}(5)$
или в координатной форме, более удобной для численных расчетов:

$$K(i) = \sum_{j=1}^W (I(j) \cdot I(i, j)). \quad (47)$$

Примечание. Предполагается, что $\dot{L}_i = \{L_i\} \approx K(i)$. Таким образом, массив-локатор, характеризующий распознаваемый объект, рассматривается как сумма произведений профилей классов на интегральный критерий сходства массива-локатора с этими профилями (т.е. взвешенная суперпозиция или разложение в ряд по профилям классов).

В выражении (47):

где $I(j)$ – интегральный критерий сходства массива-локатора, описывающего состояние объекта и j -го класса рассчитываемый, согласно выражений (42) или (44):

$$I_j = \frac{1}{S_j S_l M} \sum_{i=1}^M (I_{ij} - \bar{I}_j) (L_i - \bar{L}), \quad (48)$$

$I(i, j)$ – вектор обобщенного образа j -го класса, координаты которого рассчитываются в соответствии с системным обобщением формулы Харкевича (18):

$$I_{ij} = \text{Log}_2 \left(\frac{N_{ij}}{N_i N_j} \right)^{\frac{\text{Log}_2 W^j}{\text{Log}_2 N}} + \text{Log}_2 W^j. \quad (49)$$

Примечание: обозначения $I(i, j)$ и I_{ij} , и т.п. эквивалентны. Смысл всех переменных, входящих в выражения (48) и (49) раскрыт выше.

При дальнейшем развитии данной аналогии естественно возникают вопросы: о полноте, избыточности и ортонормированности системы векторов классов как функций, по которым проводится разложение вектора объекта; о сходимости, т.е. возможности и корректности такого разложения.

В общем случае вектор объекта совершенно не обязательно должен разлагаться в ряд по векторам классов таким образом, чтобы сумма ряда во всех точках совпадала со значениями исходной функции. Это означает, что система векторов классов может быть неполна по отношению к профилю распознаваемого объекта, и, тем более, всех возможных объектов.

Предлагается считать не разлагаемые в ряд, т.е. плохо распознаваемые объекты суперпозицией хорошо распознаваемых объектов ("похожих" на те, которые использовались для формирования обобщенных образов классов), и объектов, которые и не должны распознаваться, так как объекты этого типа не встречались в обучающей выборке и не использовались для формирования обобщенных образов классов и не коррелирует с ними, а также не относятся к представляемой обучающей выборкой генеральной совокупности.

Нераспознаваемую компоненту можно рассматривать либо как шум, либо считать ее полезным сигналом, несущим ценную информацию о неисследованных объектах интересующей нас предметной области (в зависимости от целей и тезауруса исследователей). Использование первого варианта не приводит к осложнениям, так как примененный в математической модели алгоритм сравнения векторов объектов и классов, основанный на вычислении нормированной корреляции Пирсона (сумма произведений), является весьма устойчивым к наличию белого шума в идентифицируемом сигнале. При использовании второго варианта необходимо дообучить систему распознаванию объектов, несущих такую компоненту (в этой возможности и заключается адаптивность модели). Технически этот вопрос решается копированием описаний плохо распознавшихся объектов из распознаваемой выборки в обучающую, их идентификацией экспертами и дообучением системы. Кроме того, может быть целесообразным расширить справочник классов распознавания новыми классами, соответствующими этим объектам, и осуществить пересинтез модели. Это позволяет расширить генеральную совокупность, отражаемую моделью, по отношению к которой обучающая выборка репрезентативна.

Однако на практике гораздо чаще наблюдается противоположная ситуация (можно даже сказать, что она типична), когда система векторов избыточна, т.е. в системе классов распознавания есть очень похожие классы (между которыми имеет место высокая корреляция, наблюдаемая в режиме кластерно-конструктивный анализ). Это означает, что в системе сформировано несколько практически одинаковых образов с разными наименованиями. Для исследователя это само по себе является очень ценной информацией. Однако если исходить только из потребности разложения распознаваемого объекта в ряд по векторам классов (чтобы определить суперпозицией каких образов он является, т.е. "разложить его на компоненты"), то наличие сильно коррелирующих друг с другом векторов представляется неоправданным, так как просто увеличивает размерности данных, внося в них мало нового по существу. Поэтому возникает задача исключения избыточности системы классов распознавания, т.е. выбора из всей системы классов распознавания такого минимального их набора, в котором профили классов минимально коррелируют друг с другом, т.е. ортогональны в фазовом пространстве признаков. Это условие в теории рядов называется

"ортонормируемостью" системы базовых функций, а в факторном анализе связано с идеей выделения "главных компонент".

В предлагаемой математической модели реализованы два варианта выхода из данной ситуации:

- 1) исключение неформирующихся, расплывчатых классов;
- 2) объединение почти идентичных по содержанию (дублирующих друг друга) классов.

Однако выбрать нужный вариант и реализовать его, используя соответствующие режимы, пользователь технологии АСК-анализа должен сам. Вся необходимая и достаточная информация для принятия соответствующих решений предоставляется пользователю инструментария АСК-анализа.

Если считать, что функции образов составляют формально-логическую систему, к которой применима теорема Геделя, то можно сформулировать эту теорему для данного случая следующим образом: "Для любой системы базисных функций в принципе всегда может существовать по крайней мере одна такая функция, что она не может быть разложена в ряд по данной системе базисных функций, т.е. функция, которая является ортонормированной ко всей системе базисных функций в целом". Поэтому для адекватного отражения подобных функций в модели необходимо повышение размерности семантического информационного пространства, т.е. увеличение размерности справочников классов и признаков.

Очевидно, не взаимосвязанными друг с другом могут быть только четко оформленные, детерминистские образы, т.е. образы с высокой степенью редукции ("степень сформированности конструкта"). Поэтому в процессе выявления взаимно-ортогональных базисных образов в первую очередь из модели будут исключены аморфные "расплывчатые" образы, которые связаны практически со всеми остальными образами.

В некоторых случаях результат такого процесса представляет интерес, и это делает оправданным его реализацию. Однако можно предположить, что наличие расплывчатых образов в системе является оправданным, так как в этом случае система образов не будет формальной и подчиняющейся теореме Геделя. Следовательно, система распознавания будет более полна в том смысле, что увеличится вероятность идентификации любого объекта, предъявленного ей на распознавание. Конечно, уровень сходства с аморфным образом не может быть столь высоким, как с четко оформленным, в связи с чем в этом случае более уместно применять термины "ассоциация" или нечеткая, расплывчатая идентификация, чем "однозначная идентификация".

Итак, можно сделать следующий вывод: допустимость в математической модели СК-анализа не только четко оформленных (детерминистских) образов, но и аморфных, нечетких, расплывчатых, рыхлых образов является важным достоинством данной модели. Это обусловлено тем, что

данная модель обеспечивает корректные результаты анализа, идентификации и прогнозирования даже в тех случаях, когда модели идентификации и информационно-поисковые системы детерминистского типа традиционных АСУ практически неработоспособны. В этих условиях данная модель СК-анализа работает как система ассоциативной (нечеткой) идентификации.

Таким образом, в предложенной семантической информационной модели при идентификации и прогнозировании, по сути, осуществляется разложение векторов идентифицируемых объектов по векторам классов распознавания, т.е. выполняется "объектный анализ" (по аналогии с спектральным, гармоническим или Фурье-анализом), что позволяет рассматривать идентифицируемые объекты как суперпозицию обобщенных образов классов различного типа с различными амплитудами (25). При этом вектора обобщенных образов классов, с математической точки зрения, представляют собой произвольные функции и не обязательно образуют полную и не избыточную (ортонормированную) систему функций.

Для любого объекта всегда существует такая система базисных функций, что вектор объекта может быть представлен в форме линейной суперпозиции (суммы) этих базисных функций с различными амплитудами. Это утверждение, по-видимому, является одним из следствий фундаментальной теоремы А.Н. Колмогорова, доказанной им в 1957 г.⁶

Теорема Колмогорова. Любая непрерывная функция от n переменных $F(x_1, x_2, \dots, x_n)$ может быть представлена в виде:

$$F(x_1, x_2, \dots, x_n) = \sum_{j=1}^{2n+1} \left(g_j \sum_{i=1}^n (h_{ij}(x_i)) \right),$$

где g_j и h_{ij} – непрерывные функции, причем h_{ij} не зависят от функции F .

Эта теорема означает, что для реализации функций многих переменных достаточно операций суммирования и композиции функций одной переменной. Удивительно, что в этом представлении лишь функции g_j зависят от представляемой функции F , а функции h_{ij} универсальны. Это означает, что одну и ту же функцию многих переменных F можно разложить в ряд по различным системам базисных функций h_{ij} . Необходимо отметить, что теорема Колмогорова является обобщением теоремы В.И. Арнольда (1957), которая дает решение 13-й проблемы Гильберта.

К сожалению, определение вида функций h_{ij} и g_j для данной функции F представляет собой математическую проблему, для которой пока не найдено общего строгого решения.

В работе [7] предлагается рассматривать приведенную семантическую информационную модель как один из вариантов решения этой про-

⁶ См.: О представлении непрерывных функций нескольких переменных в виде суперпозиций непрерывных функций одного переменного и сложения // Докл. АН СССР, Т. 114, С. 953–956, 1957.

блемы. В этом контексте функция F интерпретируется как образ идентифицируемого объекта, функция h_{ij} – образ j -го класса, а функция g_j – мера сходства образа объекта с образом класса.

3. Методика численных расчетов АСК-анализа

3.1. Детальный список БКОСА и их алгоритмов

В табл. 5 приведен детальный список базовых когнитивных операций системного анализа, которым соответствуют 24 алгоритма. Все эти алгоритмы приведены в полном виде в работе [7].

Т а б л . 5 . Детальный список базовых когнитивных операций системного анализа (БКОСА)

Номер алгоритма	Код БКОСА по схеме СК-анализа	Номер БКОСА	Наименование БКОСА	Полное наименование базовых когнитивных операций системного анализа (БКОСА)		
	1.1	1	Присвоение имен	Присвоение имен классам (интенциональная, интегральная репрезентация)		
	1.2			Присвоение имен атрибутам (экстенциональная, дискретная репрезентация)		
1	2.1	2	Восприятие	Восприятие и запоминание исходной обучающей информации		
2	2.2			Репрезентация. Сопоставление индивидуального опыта с коллективным (общественным)		
3	3.1.1	3	Обобщение (синтез, индукция)	Накопление первичных данных		
4	3.1.2			Исключение артефактов		
5	3.1.3			Расчет истинности смысловых связей между предпосылками и результатами (обобщенных таблиц решений)		
6	3.2			Определение значимости шкал и градаций факторов, уровней Мерлина		
7	3.3			Определение значимости шкал и градаций классов, уровней Мерлина		
8	4.1			4	Абстрагирование	Абстрагирование факторов (снижение размерности семантического пространства факторов)
9	4.2					Абстрагирование классов (снижение размерности семантического пространства классов)
10	5	5	Оценка адекватности	Оценка адекватности информационной модели предметной области		
11	7	6	Сравнение, идентификация и прогнозирование	Сравнение, идентификация и прогнозирование. Распознавание состояний конкретных объектов (объектный анализ)		
12	9.1	7	Анализ, дедукция и абдукция	Анализ, дедукция и абдукция классов (семантический анализ обобщенных образов классов, решение обратной задачи прогнозирования)		
13	9.2			Анализ, дедукция и абдукция факторов (семантический анализ факторов)		
14	10.1.1	8	Классификация и генерация конструктов	Классификация обобщенных образов классов		
15	10.1.2			Формирование бинарных конструктов классов		
16	10.1.3			Визуализация семантических сетей классов		
17	10.2.1			Классификация факторов		
18	10.2.2			Формирование бинарных конструктов факторов		
19	10.2.3			Визуализация семантических сетей факторов		
20	10.3.1	9	Содержательное сравнение	Содержательное сравнение классов		
21	10.3.2			Расчет и отображение многозначных когнитивных диаграмм, в т.ч. диаграмм Мерлина		
22	10.4.1			Содержательное сравнение факторов		
23	10.4.2			Расчет и отображение многозначных когнитивных диаграмм, в т.ч. инвертированных диаграмм Мерлина		
24	11.	10	Планирование и управление	Многовариантное планирование и принятие решения о применении системы управляющих факторов		

3.2. Иерархическая структура данных семантической информационной модели СК-анализа

На рис. 5 приведен обобщенный вид иерархической структуры баз данных семантической информационной модели системно-когнитивного анализа. На этой схеме базы данных обозначены прямоугольниками, а базовые когнитивные операции системного анализа, преобразующие одну базу в другую, – стрелками с надписями. Имеются также базовые когнитивные операции, формирующие выходные графические формы.

Из этой схемы видно, что одни базовые когнитивные операции готовят данные для других операций, относящихся к более высоким уровням иерархии системы процессов познания. Этим определяется возможная последовательность выполнения базовых когнитивных операций.

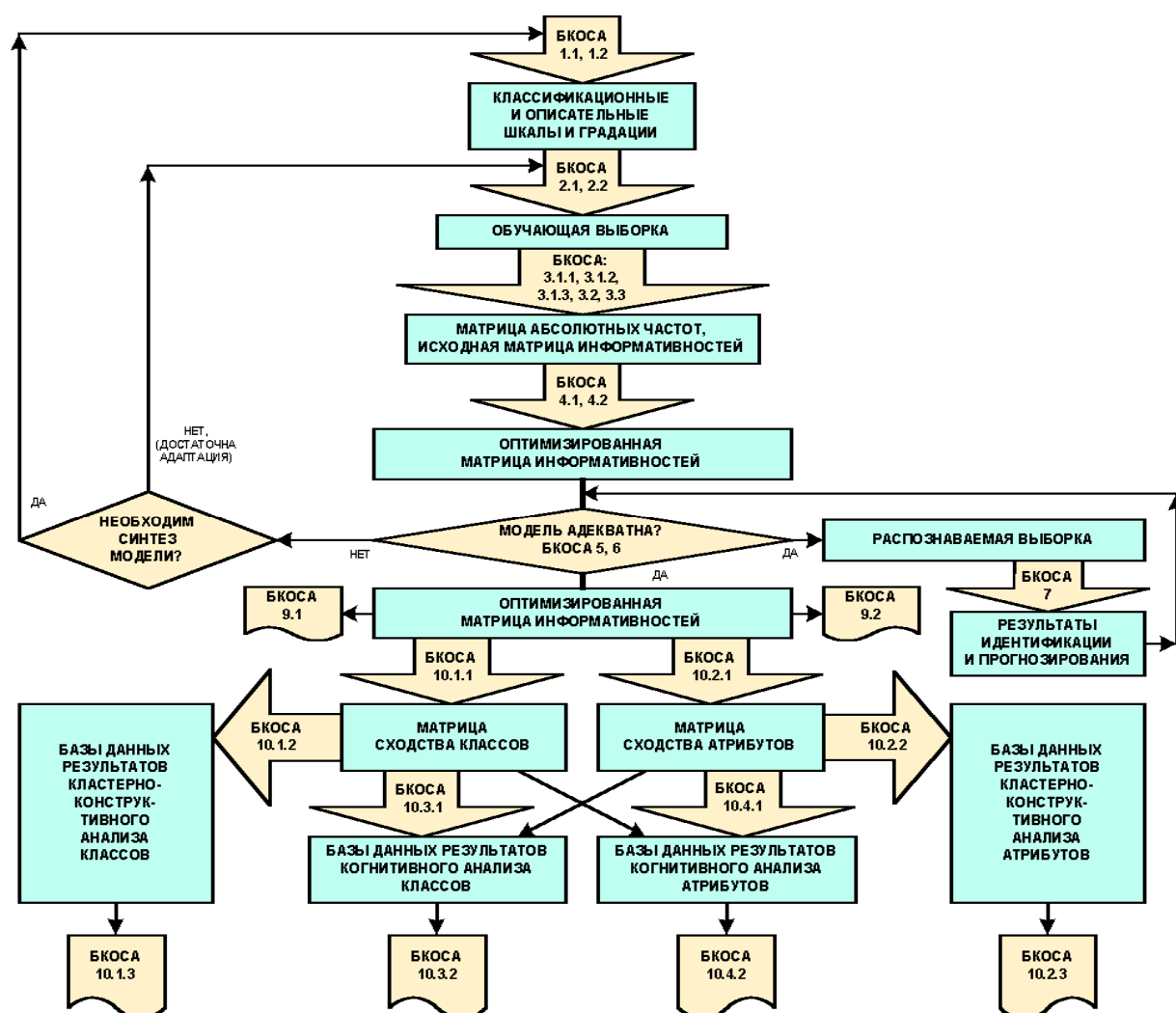


Рис. 5. Иерархическая структура данных СК-анализа

4. Специальный программный инструментарий АСК-анализа – интеллектуальная система "Эйдос"

2.4.1. Цели и основные функции системы "Эйдос"

Универсальная когнитивная аналитическая система "Эйдос" является отечественным лицензионным программным продуктом [7, 15]⁷, созданным исключительно с использованием официально приобретенного лицензионного программного обеспечения. Титульная видеодиаграмма системы приведена на рис. 6.



Рис. 2.6. Титульная видеодиаграмма системы "Эйдос"

Система "Эйдос" является одним из элементов предлагаемого решения проблемы и достижения цели данной работы, так как она обеспечивает выполнение следующих задач.

1. Синтез и адаптация семантической информационной модели предметной области, включая объект активный управления и окружающую среду.

2. Идентификация и прогнозирование состояния активного объекта управления, а также разработка управляющих воздействий для его перевода в заданные целевые состояния.

3. Углубленный анализ семантической информационной модели предметной области.

Таким образом, система "Эйдос" служит инструментарием, решающим проблему данной работы.

⁷ См.: <http://lc.kubagro.ru/aidos>

2.4.1.1. Синтез содержательной информационной модели предметной области

Для разработки информационной модели предметной области необходимо владеть основными принципами ее когнитивной структуризации и формализованного описания. Синтез содержательной информационной модели включает следующие этапы.

1. Формализация (когнитивная структуризация предметной области).
2. Формирование исследуемой выборки и управление ею.
3. Синтез или адаптация модели.
4. Оптимизация модели.
5. Измерение адекватности модели (внутренней и внешней, интегральной и дифференциальной валидности), ее скорости сходимости и семантической устойчивости.

2.4.1.2. Идентификация и прогнозирование состояния объекта управления, выработка управляющих воздействий

Данный вид работ включает:

1. Ввод распознаваемой выборки.
2. Пакетное распознавание.
3. Вывод результатов распознавания и их оценку.

2.4.1.3. Углубленный анализ содержательной информационной модели предметной области

Углубленный анализ выполняется в подсистеме "Типология" и включает:

1. Информационный и семантический анализ классов и признаков.
2. Кластерно-конструктивный анализ классов распознавания и признаков, включая визуализацию результатов анализа в оригинальной графической форме когнитивной графики (семантические сети классов и признаков).
3. Когнитивный анализ классов и признаков (когнитивные диаграммы и диаграммы Мерлина).

2.4.2. Обобщенная структура системы "Эйдос"

Данной обобщенной структуре соответствуют и структура управления и дерево диалога системы (табл. 6):

Т а б л . 6 . Обобщенная структура системы "Эйдос"

Подсистема	Режим	Функция	Операция
1. Формализация ПО	1. Классификационные шкалы и градации		
	2. Описательные шкалы (и градации)		
	3. Градации описательных шкал (признаки)		
	4. Иерархические уровни систем	1. Уровни классов	
		2. Уровни признаков	
	5. Программные интерфейсы для импорта данных	1. Импорт данных из TXT-файлов стандарта DOS-текст	
		2. Импорт данных из DBF-файлов стандарта проф. А.Н.Лебедева	
		3. Импорт из транспонированных DBF-файлов проф. А.Н.Лебедева	
		4. Генерация шкал и обучающей выборки RND-модели	
		5. Генерация шкал и обучающей выборки для исследования чисел	
6. Транспонирование DBF-матриц исходных данных			

		7. Импорт данных из DBF-файлов стандарта Евгения Лебедева	
		8. Системно-когнитивный анализ стандартных графических шрифтов	
		9. Микроструктура системы как фактор управления ее макросвойствами	
	6. Почтовая служба по НСИ	1. Обмен по классам	
		2. Обмен по обобщенным признакам	
		3. Обмен по первичным признакам	
	7. Печать анкеты		
2. Синтез СИМ	1. Ввод-корректировка обучающей выборки		
	2. Управление обучающей выборкой	1. Параметрическое задание объектов для обработки	
		2. Статистическая характеристика, ручной ремонт	
	3. Синтез семантической информационной модели СИМ	3. Автоматический ремонт обучающей выборки	
		1. Расчет матрицы абсолютных частот	
		2. Исключение артефактов (робастная процедура)	
		3. Расчет матрицы информативностей СИМ-1 и сделать ее текущей	
		4. Расчет условных процентных распределений СИМ-1 и СИМ-2	
		5. Автоматическое выполнение режимов 1-2-3-4	
		6. Измерение сходимости и устойчивости модели	1. Сходимость и устойчивость СИМ
			2. Зависимость валидности модели от объема обучающей выборки
		7. Расчет матрицы информативностей СИМ-2 и сделать ее текущей	
		8. Расчет стат. характеристик обучающей выборки	
9. Рассчитать БД разн.факт.и теор.хи-квадрат СИМ-3 и сделать ее текущей			
10. Рассчитать матрицу ROI СИМ-4 и сделать ее текущей			
11. Рассчитать все модели: СИМ-1, СИМ-2, СИМ-3, СИМ-4 на основе Abs.dbf			
4. Почтовая служба по обучающей информации			
3. Оптимизация СИМ	1. Формирование ортонормированного базиса классов		
	2. Исключение признаков с низкой селективной силой		
	3. Удаление классов и признаков, по которым недостаточно данных		
	4. Разделение классов на типичную и нетипичную части		
	5. Генерация сочтенных признаков и перекодирование обучающей выборки		
	6. Удаление малодостоверных данных в БД: ABS, PER, INF по критерию хи-квадрат		
4. Распознавание	1. Ввод-корректировка распознаваемой выборки		
	2. Пакетное распознавание		
	3. Вывод результатов распознавания	1. Разрез: один объект – много классов	
		2. Разрез: один класс – много объектов	
	4. Почтовая служба по распознаваемой выборке		
	5. Построение функций влияния		
	6. Докодирование сочетаний признаков в распознаваемой выборке		
	7. Назначения объектов на классы (задача о назначениях)	1. Задание ограничений на ресурсы по классам	
		2. Ввод затрат на объекты	
		3. Назначения объектов на классы (LC-алгоритм)	
4. Сравнение эффективности LC и RND алгоритмов			
8. Интерактивная идентификация - последовательный анализ Вальда			
9. Сортировка результатов пакетного распознавания			
5. Типология	1. Типологический анализ классов распознавания	1. Информационные (ранговые) портреты (классов)	
		2. Кластерный и конструктивный анализ классов	1. Расчет матрицы сходства образов классов
			2. Генерация кластеров и конструктов классов
			3. Просмотр и печать кластеров и конструктов
			4. Автоматическое выполнение режимов: 1,2,3
			5. Вывод 2d семантических сетей классов
			6. Агломеративная древовидная кластеризация классов
	7. Вывод дерева классов		
	3. Когнитивные диаграммы классов		
	2. Типологический анализ первичных признаков	1. Информационные (ранговые) портреты признаков	
		2. Кластерный и конструктивный анализ признаков	1. Расчет матрицы сходства образов признаков
			2. Генерация кластеров и конструктов признаков
			3. Просмотр и печать кластеров и конструктов
			4. Автоматическое выполнение режимов: 1,2,3
5. Вывод 2d семантических сетей признаков			

		3. Когнитивные диаграммы признаков	
		3. Подготовка баз данных для визуализации когнитивных функций в SigmaPlot и Excel	
		4. Визуализация прямых и обратных, позитивных и негативных когнитивных функций	
6. СК-анализ СИМ		1. Оценка достоверности заполнения объектов	
		2. Измерение адекватности семантической информационной модели	
		3. Измерение независимости классов и признаков	
		4. Просмотр профилей классов и признаков	
		5. Графическое отображение нелокальных нейронов	
		6. Отображение Паретто-подмножеств нейронной сети	
		7. Классические и интегральные когнитивные карты	
		8. Восстановление значений функций по признакам аргумента ⁸	1. Восстановление значений и визуализация 1d-функций 2. Восстановление значений и визуализация 2d-функций 3. Преобразование 2d-матрицы в 1d-таблицу с признаками точек 4. Объединение многих БД: Inp_0001.dbf и т.д. в Inp_data.dbf 5. Помощь по подсистеме (требования к исходным данным)
7. Сервис	1. Генерация (сброс) БД	1. Все базы данных	
		2. НСИ	1. Всех баз данных НСИ 2. БД классов 3. БД первичных признаков 4. БД обобщенных признаков
		3. Обучающая выборка	
		4. Распознаваемая выборка	
		5. Базы данных статистики	
		2. Переиндексация всех баз данных	
		3. Печать БД абсолютных частот	
		4. Печать БД условных процентных распределений СИМ-1 и СИМ-2	
		5. Печать БД информативностей СИМ-1 и СИМ-2	
		6. Интеллектуальная дескрипторная информационно-поисковая система	
		7. Копирование основных баз данных СИМ	
		8. Сделать текущей матрицу информативностей СИМ-1	
		9. Сделать текущей матрицу информативностей СИМ-2	
		10. Конвертировать все РСХ в GIF	
	11. Изменение структур БД к текущему стандарту		
	12. Сделать текущей матрицу информативностей СИМ-3		
	13. Сделать текущей матрицу информативностей СИМ-4		
8. О системе		1. Графическая заставка системы «Эйдос»	
		2. Логотипы мультимodelей	

Подробнее подсистемы, режимы, функции и операции, реализуемые системой "Эйдос", описаны в работе [1].

5. Выводы (опыт и перспективы применения АСК-анализа для управления социально-экономическими системами в АПК)

Авторами проведен ряд исследований по управлению многоотраслевыми интегрированными производственными системами АПК (МИПС АПК), с применением АСК-анализа, технология и методика проведения которых, а также полученные результаты, отражены в работах [1-21].

В частности в статье [5] сформирована главная проблема управления МИПС АПК состоящая в том, что с одной стороны необходимо вырабатывать рекомендации по управлению, для чего необходима адаптивная модель объекта управления, а, с другой стороны, построение такой модели

⁸ См.: Луценко Е.В. Системно-когнитивный анализ функций и восстановление их значений по признакам аргумента на основе априорной информации (интеллектуальные технологии интерполяции, экстраполяции, прогнозирования и принятия решений по картографическим базам данных) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №07(51). С. 130 – 154. – Шифр Информрегистра: 0420900012\0066. – Режим доступа: <http://ej.kubagro.ru/2009/07/pdf/06.pdf>, 1,562 у.п.л.

затруднительно из-за высокой сложности и динамичности внутренней логики объекта управления, его территориально распределенного и многоотраслевого характера, огромного количества экономических показателей, характеризующих его деятельность на различных уровнях его организации. Кратко рассмотрены подходы к решению основной проблемы на основе «менеджмента знаний», а также соотношение данных, информации и знаний⁹, процедуры выявления знаний из данных, способ использования знаний для прогнозирования и принятия решений, этапы жизненного цикла интеллектуального приложения и основные результаты применения современной автоматизированной технологии управления знаниями: АСК-анализа для управления МИПС АПК. В [6] приводятся основные результаты по решению основной проблемы управления МИПС АПК на основе применения современной автоматизированной технологии управления знаниями: системно-когнитивного анализа (СК-анализ). Подробнее рассматривается технология когнитивных функций СК-анализа, обеспечивающая как выявление знаний из эмпирических данных, так и использование этих знаний для поддержки принятия решений по управлению МИПС АПК в целом на основе управления характеристиками входящих в нее предприятий¹⁰. В работе [7] приводятся примеры постановки и решения задач прогнозирования и поддержки принятия решений (управления) для МИПС АПК на основе его двухуровневой семантической информационной модели. В [8] в общем виде сформулирована проблема управления МИПС АПК. Предлагается общий метод решения сформулированной проблемы путем применения системно-когнитивного подхода. В [9] предлагается общий метод решения сформулированной проблемы путем применения системно-когнитивного подхода. Описывается первый этап синтеза модели: когнитивная структуризация объекта управления и классификация частных моделей, входящих в многоуровневую семантическую информационную модель МИПС АПК. В [10] в обобщенной форме рассмотрена методология применения системно-когнитивного анализа для синтеза многоуровневой семантической информационной модели управления МИПС АПК и решения на ее основе задач прогнозирования, поддержки принятия решений по управлению и научных исследований. В работах [11] и [12] анализируются исходные данные для построения двухуровневой семантической информационной модели управления МИПС АПК, ставится и решается задача их автоматизированного преобразования к виду, непосредственно воспринимаемому системой "Эйдос" с помощью одного из ее стандартных интерфейсов. Приводятся алгоритм и исходный текст программы, обеспечивающей эти функции, а также результаты ее работы и автоматически сформированные системой "Эйдос" справочники классов и

⁹ См., также: <http://www.twirpx.com/file/793311/>

¹⁰ См., также: <http://www.twirpx.com/file/775236/>

факторов, а также обучающая выборка для частных моделей, входящих в двухуровневую семантическую информационную модель управления МИПС АПК. В статье [13] на простом примере описывается смысл семантической информационной модели СК-анализа. Приводятся результаты синтеза и верификации системы частных моделей, входящих в двухуровневую семантическую информационную модель управления МИПС АПК. В [14] приводятся примеры постановки и решения задач прогнозирования и поддержки принятия решений (управления) для МИПС АПК на основе двухуровневой семантической информационной модели. В [15] проводится исследование двухуровневой семантической информационной модели МИПС АПК, которое корректно считать исследованием самого объекта моделирования и управления, так как верификация модели показала ее высокую адекватность.

В [16] формулируется проблема краткосрочного прогнозирования значений экономических показателей МИПС АПК, на основе применения системно-когнитивного анализа (СК-анализ) и его инструментария – интеллектуальной системы «Эйдос», осуществляется формальная постановка задачи и формализация предметной области, т.е. разработка классификационных и описательных шкал и градаций и формирование обучающей выборки. В [17] описывается процедура синтеза четырех моделей МИПС АПК, отличающихся частыми критериями взаимосвязи между прошлыми показателями предприятий, входящих в корпорацию и будущими состояниями корпорации в целом, производится верификация всех частных моделей с использованием двух интегральных критериев, осуществляется прогнозирование будущих состояний корпорации по их системе детерминации.

В [18] формулируется проблема краткосрочного прогнозирования трендов экономических показателей МИПС АПК, на основе применения системно-когнитивного анализа (СК-анализ) и его инструментария (интеллектуальной системы «Эйдос») осуществляется формальная постановка задачи и формализация предметной области, т.е. разработка классификационных и описательных шкал и градаций и формирование обучающей выборки, а также синтез и верификация моделей, включающих четыре базы знаний, отличающихся частными критериями знаний. В статье [19] описывается процедура синтеза четырех моделей МИПС АПК, отличающихся частными критериями взаимосвязи между трендами прошлых показателей предприятий, входящих в корпорацию и будущими состояниями корпорации в целом, производится верификация всех частных моделей с использованием двух интегральных критериев, осуществляется прогнозирование будущих состояний корпорации по их системе детерминации.

В статье [20] формулируется проблема краткосрочного прогнозирования сценариев изменения трендов экономических показателей многоотраслевой корпорации, на основе прошлых сценариев с применением си-

стемно-когнитивного анализа (СК-анализ) и его инструментария (интеллектуальной системы «Эйдос»). Осуществляется формальная постановка задачи и формализация предметной области, т.е. разработка классификационных и описательных шкал и градаций и формирование обучающей выборки. Описывается процедура синтеза четырех моделей корпорации, отличающихся частными критериями взаимосвязи между прошлыми и будущими сценариями трендов показателей корпорации. Производится верификация всех частных моделей с использованием двух интегральных критериев. Решается задача прогнозирования динамики будущих состояний корпорации на основе ее динамики в прошлом. В статье [21] проводится системно-когнитивный анализ когнитивных моделей (баз знаний) краткосрочного прогнозирования сценариев изменения трендов экономических показателей многоотраслевой корпорации, на основе прошлых сценариев.

Но главный вывод, который, по мнению авторов обоснованно можно сделать по результатам данной работы состоит в том, что математическая модель, методика численных расчетов (структуры данных и алгоритмы), специальный программный инструментарий СК-анализа (система "Эйдос"), а также методика и технология их применения являются адекватным инструментом для управления социально-экономическими системами в АПК. Таким образом, системно-когнитивный анализ и реализующий его программный инструментарий – универсальная когнитивная аналитическая система "Эйдос" являются адекватным средством решения поставленной в работе проблемы и задач.

Литература

Обобщающие работы¹¹

1. Луценко Е.В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с.
2. Луценко Е.В., Лойко В.И., Семантические информационные модели управления агропромышленным комплексом. Монография (научное издание). – Краснодар: КубГАУ. 2005. – 480 с.
3. Луценко Е. В., Лойко В.И., Великанова Л.О. Прогнозирование и принятие решений в растениеводстве с применением технологий искусственного интеллекта: Монография (научное издание). – Краснодар: КубГАУ, 2008. – 257 с.
4. Макаревич О.А. Управление агропромышленным холдингом с применением технологий искусственного интеллекта: Монография (научное издание). – М: "Финансы и статистика", 2009. – 215 с.

Методологические работы

5. Луценко Е.В. Автоматизированные технологии управления знаниями в агропромышленном холдинге / Е.В. Луценко, В.И. Лойко, О.А. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №08(52). С. 98 – 109. –

¹¹ Для удобства читателей некоторые из этих работ приведены на сайте: <http://lc.kubagro.ru>

Шифр Информрегистра: 0420900012\0088. – Режим доступа: <http://ej.kubagro.ru/2009/08/pdf/07.pdf>, 0,75 у.п.л.

6. Луценко Е.В. Управление агропромышленным холдингом на основе когнитивных функций связи результатов работы холдинга и характеристик его предприятий / Е.В. Луценко, В.И. Лойко, О.А. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №10(54). С. 248 – 260. – Шифр Информрегистра: 0420900012\0111. – Режим доступа: <http://ej.kubagro.ru/2009/10/pdf/15.pdf>, 0,812 у.п.л.

7. Макаревич О.А. Применение технологий искусственного интеллекта для прогнозирования и управления в агропромышленном холдинге / О.А. Макаревич, Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2010. – №05(59). С. 149 – 157. – Шифр Информрегистра: 0421000012\0093. – Режим доступа: <http://ej.kubagro.ru/2010/05/pdf/10.pdf>, 0,562 у.п.л.

8. Макаревич О.А. Управление агропромышленным холдингом на основе системно-когнитивного подхода / О.А. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2010. – №04(58). С. 233 – 242. – Шифр Информрегистра: 0421000012\0069. – Режим доступа: <http://ej.kubagro.ru/2010/04/pdf/13.pdf>, 0,625 у.п.л.

9. Луценко Е.В. Системно-когнитивный подход к построению многоуровневой семантической информационной модели управления агропромышленным холдингом / Е.В. Луценко, В.И. Лойко, О.А. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №07(41). С. 194 – 214. – Шифр Информрегистра: 0420800012\0095. – Режим доступа: <http://ej.kubagro.ru/2008/07/pdf/11.pdf>, 1,312 у.п.л.

10. Луценко Е.В. Методология применения системно-когнитивного анализа для синтеза многоуровневой семантической информационной модели агропромышленного холдинга и решения на ее основе задач прогнозирования, поддержки принятия управленческих решений и научных исследований / Е.В. Луценко, В.И. Лойко, О.А. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №01(45). С. 11 – 29. – Шифр Информрегистра: 0420900012\0006. – Режим доступа: <http://ej.kubagro.ru/2009/01/pdf/02.pdf>, 1,188 у.п.л.

Работы, посвященные синтезу моделей и их применению для решения задач прогнозирования, управления и исследования предметной области

11. Луценко Е.В. Исследование характеристик исходных данных по агропромышленному холдингу и разработка программного интерфейса их объединения и стандартизации (формализация предметной области) / Е.В. Луценко, В.И. Лойко, О.А. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №07(41). С. 215 – 246. – Шифр Информрегистра: 0420800012\0094. – Режим доступа: <http://ej.kubagro.ru/2008/07/pdf/12.pdf>, 2 у.п.л.

12. Макаревич О.А. Программный интерфейс стандартизации исходных данных по агропромышленному холдингу / О.А. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2010. – №04(58). С. 243 – 251. – Шифр Информрегистра: 0421000012\0066. – Режим доступа: <http://ej.kubagro.ru/2010/04/pdf/14.pdf>, 0,562 у.п.л.

13. Луценко Е.В. Синтез и верификация двухуровневой семантической информационной модели агропромышленного холдинга / Е.В. Луценко, В.И. Лойко, О.А. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №08(42). С. 1 – 15. – Шифр Информрегистра: 0420800012\0120. – Режим доступа: <http://ej.kubagro.ru/2008/08/pdf/01.pdf>, 0,938 у.п.л.

14. Луценко Е.В. Решение задач прогнозирования и поддержки принятия решений (управления) для агропромышленного холдинга на основе его двухуровневой семантической информационной модели / Е.В. Луценко, В.И. Лойко, О.А. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №08(42). С. 16 – 34. – Шифр Информрегистра: 0420800012\0119. – Режим доступа: <http://ej.kubagro.ru/2008/08/pdf/02.pdf>, 1,188 у.п.л.

15. Луценко Е.В. Исследование двухуровневой семантической информационной модели агропромышленного холдинга / Е.В. Луценко, В.И. Лойко, О.А. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №08(42). С. 35 – 75. – Шифр Информрегистра: 0420800012\0118. – Режим доступа: <http://ej.kubagro.ru/2008/08/pdf/03.pdf>, 2,562 у.п.л.

Работы, посвященные прогнозированию значений, трендов и сценариев изменения экономических показателей

16. Оперативное прогнозирование значений экономических показателей многоотраслевой корпорации с применением технологий искусственного интеллекта (часть 1-я: постановка задачи и формализация предметной области) / Е.В. Луценко, В.И. Лойко, О.А. Макаревич, Л.О. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №07(71). С. 692 – 705. – Шифр Информрегистра: 0421100012\0271. – Режим доступа: <http://ej.kubagro.ru/2011/07/pdf/49.pdf>, 0,875 у.п.л.

17. Оперативное прогнозирование значений экономических показателей многоотраслевой корпорации с применением технологий искусственного интеллекта (часть 2-я: синтез и верификация модели) / Е.В. Луценко, В.И. Лойко, О.А. Макаревич, Л.О. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №07(71). С. 706 – 719. – Шифр Информрегистра: 0421100012\0268. – Режим доступа: <http://ej.kubagro.ru/2011/07/pdf/50.pdf>, 0,875 у.п.л.

18. Луценко Е.В. Оперативное прогнозирование трендов экономических показателей многоотраслевой корпорации с применением технологий искусственного интеллекта (часть 1-я: постановка задачи и формализация предметной области) / Е.В. Луценко, В.И. Лойко, Л.О. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №09(73). С. 466 – 477. – Шифр Информрегистра: 0421100012\0378. – Режим доступа: <http://ej.kubagro.ru/2011/09/pdf/43.pdf>, 0,75 у.п.л.

19. Луценко Е.В. Оперативное прогнозирование трендов экономических показателей многоотраслевой корпорации с применением технологий искусственного интеллекта (часть 2-я: синтез и верификация модели) / Е.В. Луценко, В.И. Лойко, Л.О. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №09(73). С. 478 – 487. – Шифр Информрегистра: 0421100012\0376. – Режим доступа: <http://ej.kubagro.ru/2011/09/pdf/44.pdf>, 0,625 у.п.л.

20. Макаревич Л.О. АСК-анализ и прогнозирование сценариев изменения трендов экономических показателей многоотраслевой корпорации (часть 1) / Л.О. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2012. – №02(76). С. 124 – 145. – Режим доступа: <http://ej.kubagro.ru/2012/02/pdf/11.pdf>, 1,375 у.п.л.

21. Макаревич Л.О. АСК-анализ и прогнозирование сценариев изменения трендов экономических показателей многоотраслевой корпорации (часть 2) / Л.О. Макаревич // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2012. – №02(76). С. 180 – 194. – Режим доступа: <http://ej.kubagro.ru/2012/02/pdf/15.pdf>, 0,938 у.п.л.