

УДК 303.732.4

UDC 303.732.4

МЕТОД ВИЗУАЛИЗАЦИИ КОГНИТИВНЫХ ФУНКЦИЙ – НОВЫЙ ИНСТРУМЕНТ ИССЛЕДОВАНИЯ ЭМПИРИЧЕСКИХ ДАННЫХ БОЛЬШОЙ РАЗМЕРНОСТИ

COGNITIVE FUNCTIONS VISUALIZATION METHOD – THE NEW INSTRUMENT FOR THE LARGE DIMENSION EMPIRICAL DATA ANALYSIS

Луценко Евгений Вениаминович
д.э.н., к.т.н., профессор
Кубанский государственный аграрный университет, Краснодар, Россия

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

Трунев Александр Петрович
к. ф.-м. н., Ph.D.
Директор, A&E Trounev IT Consulting, Торонто, Канада

Alexander Trunev
Cand.Phys.-Math.Sci., Ph.D.
Director, A&E Trounev IT Consulting, Toronto, Canada

Бандык Дмитрий Константинович,
Разработчик интеллектуальных систем, Белоруссия

Bandyk Dmitry Konstantinovich,
Artificial intelligence developer, Belarus

В статье рассматривается новый перспективный инструмент АСК-анализа и его программного инструментария – системы «Эйдос» для выявления и наглядной графической визуализации причинно-следственных зависимостей из эмпирических данных большой размерности

The new methods for system-cognitive analysis to identify and present graphical visualization of causal functions from the large dimension empirical data and its software tools - «EIDOS» system are discussed.

Ключевые слова: АВТОМАТИЗИРОВАННЫЙ СИСТЕМНО-КОГНИТИВНЫЙ АНАЛИЗ, БАЗА ЗНАНИЙ, КОГНИТИВНАЯ ФУНКЦИЯ, СЕМАНТИЧЕСКАЯ ИНФОРМАЦИОННАЯ МОДЕЛЬ

Keywords: COMPUTERIZED SYSTEM-COGNITIVE ANALYSIS, KNOWLEDGE BASE, COGNITIVE FUNCTION, SEMANTIC INFORMATION MODEL

***“Истинное знание – это знание причин”
Френсис Бэкон (1561–1626 гг.)***

Проблема, решаемая практически во всех экспериментальных исследованиях, состоит в выявлении причинно-следственных зависимостей из эмпирических данных и формальном представлении этих зависимостей в аналитической форме, т.е. в форме математических функций. Автоматизированный системно-когнитивный анализ (АСК-анализ) [1-30] предоставляет для этого ряд новых возможностей, рассмотрение которых является предметом данной статьи.

Информационные портреты градаций факторов (признаков) были исторически первой формой выявления их влияния на принадлежность объекта к классам, реализованной в системе «Эйдос» изначально, т.е. с самых первых версий этой системы [3], так и в более поздних версиях [4]. Аппарат визуализации полностью редуцированных когнитивных функций средствами системы «Эйдос» был реализован в системе в 2004 [5-7]. В 2009-2010 годах началось широкое использование в научных исследованиях [8-12] *нередуцированных когнитивных функций*, отображаемый внешними системами (например, SigmaPlot, MS Excel) на основе баз данных,

сформированных системой «Эйдос» (в режимах _53 и _683). В 2010 году авторами начали использоваться прямые и обратные, редуцированные и нередуцированные когнитивные функции, отображаемые внешними системами на основе баз данных, подготовленных системой «Эйдос». В начале 2011 года авторами был разработан специальный режим визуализации когнитивных функций системы «Эйдос» [13], обеспечивающий визуализацию прямых и обратных, позитивных и негативных, полностью и частично редуцированных когнитивных функций.

Автоматизированный системно-когнитивный анализ (АСК-анализ) представляет собой новый метод искусственного интеллекта, развитый это системный анализ, автоматизированный путем структурирования по базовым когнитивным операциям системного анализа (*БКОСА*) и включающий: формализуемую когнитивную концепцию, математическую модель, методику численных расчетов и реализующий их программный инструментарий, в качестве которого в настоящее время выступает универсальная когнитивная аналитическая система "Эйдос" [2, 3, 4]. АСК-анализ был предложен в 2002 году одним из авторов [1].

Компоненты АСК-анализа:

- формализуемая когнитивная концепция и следующий из нее когнитивный конфигуратор;
- теоретические основы, методология, технология и методика АСК-анализа;
- математическая модель АСК-анализа, основанная на системном обобщении семантической меры целесообразности информации А. Харкевича;
- методика численных расчетов, в универсальной форме реализующая математическую модель СК-анализа, включающая иерархическую структуру данных и 24 детальных алгоритма 10 БКОСА;
- специальное инструментальное программное обеспечение, реализующее математическую модель и численный метод СК-анализа – Универсальная когнитивная аналитическая система "Эйдос";
- методика, технология и результаты синтеза рефлексивных АСУ активными объектами на основе АСК-анализа.

Этапы АСК-анализа обеспечивают последовательное *повышение степени формализации знаний о предметной области* до уровня, достаточного для представления знаний в автоматизированной системе искусственного интеллекта и решения в ней задач идентификации, прогнозирования и поддержки принятия решений (управления):

- 1) когнитивная структуризация предметной области;
- 2) формализация предметной области (конструирование классификационных и описательных шкал и градаций);

3) подготовка обучающей выборки (ввод данных мониторинга в базу прецедентов);

4) синтез семантической информационной модели (СИМ);

5) повышение эффективности СИМ;

6) проверка адекватности СИМ (измерение внутренней и внешней, дифференциальной и интегральной валидности);

7) системно-когнитивный анализ СИМ, исследование моделируемого объекта путем исследования его модели:

– решение задач идентификации и прогнозирования;

– генерация информационных портретов классов и факторов, т.е. решение обратной задачи прогнозирования, поддержка принятия решений по управлению (результаты отображаются в графической форме двухмерных и трехмерных профилей классов и факторов);

– кластерно-конструктивный анализ классов и факторов (результаты отображаются в форме семантических сетей классов и факторов);

– содержательное сравнение классов и факторов (результаты отображаются в форме когнитивных диаграмм классов и факторов);

– изучение системы детерминации состояний моделируемого объекта, нелокальные нейроны и интерпретируемые нейронные сети;

– построение классических когнитивных моделей (когнитивных карт).

– построение интегральных когнитивных моделей (интегральных когнитивных карт).

Математическая модель АСК-анализа основана на системной теории информации (СТИ).

Системная теория информации (СТИ) – Отличия СТИ от классической теории информации Больцмана-Найквиста-Хартли-Шеннона обусловлены отличиями понятия "*система*" от понятия "*множество*". СТИ рассматривает в качестве элементов не только первичные элементы множества, но и элементы, представляющие собой подсистемы различных уровней иерархии, образующиеся за счет *взаимодействия* первичных элементов, а также учитывает понятие *цели*. В рамках СТИ предложено системное обобщение семантической меры информации Харкевича, которое удовлетворяет принципу соответствия с мерой Хартли в детерминистском случае, как и мера Шеннона в случае равновероятных событий, чем преодолена несогласованность семантической теории информации и классической теории информации Шеннона. Так как данная мера учитывает понятие цели, то она является количественной мерой знаний. В рамках СТИ предложены гипотезы "О возрастании эмерджентности", следующие из нее: "О природе сложности системы", и "О видах системной информации".

Формализуемая когнитивная концепция – когнитивная концепция, предложенная с целью разработки СК-анализа. Из данной концепции выводятся структура когнитивного конфигуратора, система базовых ког-

нитивных операций и обобщенная схема системного анализа, структурированного до уровня базовых когнитивных операций (АСК-анализ).

Рассматривает 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 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.

Ключевым для когнитивной концепции является понятие *факта*, под которым понимается соответствие дискретного и интегрального, экстенционального и интенционального элементов познания (т.е. элементов разных уровней интеграции-иерархии), обнаруженное на опыте. *Факт* рассматривается как квант *смысла*, что является 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. Между когнитивными структурами разных уровней иерархии существует отношение "дискретное – интегральное". Объекты познания каждого уровня описываются как экстенционально, т.е. с использованием элементов более низкого иерархического уровня познания, так и интенционально, т.е. с использованием объекта более высокого иерархического уровня познания. Например, каждый объект исследуемой выборки экстенционально описывается на языке признаков (градаций описательных шкал), а интенционально на языке обобщенных образов классов (градаций классификационных шкал). Каждый класс экстенционально описывается объектами, использованными для синтеза его обобщенного образа, а интенционально – принадлежностью данного класса к некоторому кластеру. Кластеры экстенционально описываются обобщенными образами классов, которые в них входят, а интенционально – конструкторами, представляющими собой оси координат в неортонормированном когнитивном пространстве, отражающим парадигму реальности. Именно это служит основой формализации смысла.

Когнитивный конфигуратор – минимальный полный набор познавательных (когнитивных от: "*cognition*" – "познание", англ.) операций, к которым сводятся различные процессы познания, в т.ч. системный анализ, как метод познания, достаточный для адекватного описания данной предметной области. В *формализуемой когнитивной концепции* выявлено 10

таких *операций*¹, каждая из которых оказалась достаточно элементарной для формализации и программной реализации:

- 1) присвоение имен;
- 2) восприятие;
- 3) обобщение (синтез, индукция);
- 4) абстрагирование;
- 5) оценка адекватности модели;
- 6) сравнение, идентификация и прогнозирование;
- 7) дедукция и абдукция;
- 8) классификация и генерация конструкторов;
- 9) содержательное сравнение;
- 10) планирование и принятие решений об управлении.

Информационный портрет класса – это список факторов, ранжированных в порядке убывания силы их влияния на переход объекта управления в состояние, соответствующее данному классу. Информационный портрет класса отражает систему его детерминации. Генерация информационного портрета класса представляет собой решение обратной задачи прогнозирования, т.к. при прогнозировании по системе факторов определяется спектр наиболее вероятных будущих состояний объекта управления, в которые он может перейти под влиянием данной системы факторов, а в информационном портрете мы, наоборот, по заданному будущему состоянию объекта управления определяем систему факторов, детерминирующих это состояние, т.е. вызывающих переход объекта управления в это состояние. В начале информационного портрета класса идут факторы, оказывающие положительное влияние на переход объекта управления в заданное состояние, затем факторы, не оказывающие на это существенного влияния, и далее – факторы, препятствующие переходу объекта управления в это состояние (в порядке возрастания силы препятствования). Информационные портреты классов могут быть *отфильтрованы* по диапазону факторов, т.е. мы можем отобразить влияние на переход объекта управления в данное состояние не всех отраженных в модели факторов, а только тех, коды которых попадают в определенный диапазон, например, относящиеся к определенным описательным шкалам.

Информационный (семантический) портрет фактора – это список классов, ранжированный в порядке убывания силы влияния данного фак-

¹ Таким образом, в системе «Эйдос» используется как декларативная, так и процедурная (операциональная) формы представления знаний, причем «декларативные» базы знаний формируются в и используются в результате применения некоторых процедур, а эти процедуры реализуются над базами знаний. Вообще важно понимать, что одни и те же знания могут быть вообще не формализованы, а могут быть представлены в различных моделях представления знаний. Например, если интерпретировать «декларативную» матрицу знаний системы «Эйдос» как таблицу принятия решений, то ее можно представить в процедурном виде логических продукций вида: Если А, то В, со степенью истинности I_{AB}. Так одно и тоже значение синуса можно как найти в таблице Брадиса, так и посчитать на калькуляторе, используя разложение этой функции в ряд.

тора на переход объекта управления в состояния, соответствующие данным классам. Информационный портрет фактора называется также его *семантическим портретом*, т.к. в соответствии с концепцией смысла системно-когнитивного анализа, являющейся обобщением концепции смысла Шенка-Абельсона [5], *смысл фактора состоит в том, какие будущие состояния объекта управления он детерминирует*. Сначала в этом списке идут состояния объекта управления, на переход в которые данный фактор оказывает наибольшее влияние, затем состояния, на которые данный фактор не оказывает существенного влияния, и далее состояния – переходу в которые данный фактор препятствует. Информационные портреты факторов могут быть *отфильтрованы* по диапазону классов, т.е. мы можем отобразить влияние данного фактора на переход объекта управления не во все возможные будущие состояния, а только в состояния, коды которых попадают в определенный диапазон, например, относящиеся к определенным классификационным шкалам.

Когнитивная функция представляет собой зависимость вероятностей перехода объекта управления в будущие состояния, соответствующие классам, под влиянием различных значений некоторого фактора.

Когнитивная функции строится для *подматриц* матрицы информативностей (матрицы знаний) системы «Эйдос», образованных различными классификационными и описательными шкалами (одна из подматриц выделена жирной линией и фоном) (таблица 1):

Таблица 1 – К ПОЯСНЕНИЮ ПОНЯТИЯ: «ПОДМАТРИЦЫ МАТРИЦЫ ЗНАНИЙ»

		1-я классификационная шкала			2-я классификационная шкала			3-я классификационная шкала		
		1-я градация	2-я градация	3-я градация	1-я градация	2-я градация	3-я градация	1-я градация	2-я градация	3-я градация
1-я описательная шкала	1-я градация									
	2-я градация									
	3-я градация									
2-я описательная шкала	1-я градация									
	2-я градация									
	3-я градация									
3-я описательная шкала	1-я градация									
	2-я градация									
	3-я градация									

Если взять несколько информационных портретов факторов, соответствующих градациям одной описательной шкалы, отфильтровать их по диапазону градаций некоторой классификационной шкалы и взять из каждого информационного портрета *по одному* состоянию, на переход в которое объекта управления данное значение фактора оказывает наибольшее

влияние, то мы и получим зависимость, отражающую вероятность перехода объекта управления в будущие состояния под влиянием различных значений некоторого фактора, т.е. полностью редуцированную когнитивную функцию.

Когнитивные функции являются наиболее развитым средством изучения причинно-следственных зависимостей в моделируемой предметной области, предоставляемым системой "Эйдос". Необходимо отметить, что на вид функций влияния математической моделью СК-анализа не накладывается никаких ограничений, в частности, они могут быть и *нелинейные*.

Введем определение когнитивной функции: когда функция используется для отображения причинно-следственной зависимости, т.е. информации (согласно концепции Шенка-Абельсона [5]), или *знаний*, если эта информация полезна для достижения целей, то будем называть такую функцию *когнитивной функцией* [1, 5-7], от англ. «*cognition*»².

Смысл когнитивной функциональной зависимости в том, что в значении аргумента содержится определенное количество знаний о том, какое значение примет функция, т.е. когнитивная функция отражает знания о полезных причинно-следственных зависимостях, а не корреляцию.

Кратко рассмотрим **выявление, представление и использование знаний в АСК-анализе и системе «Эйдос»** и соотношение смысла понятия «Когнитивная функция» с содержанием понятий: «Данные, информация и знание», «эмпирическая закономерность, эмпирический закон и научный закон».

Для выявления знаний из эмпирических данных необходимо осознанно и целенаправленно изменять форму их представления таким образом, чтобы *последовательно повышать степень их формализации* до уровня, который позволяет: а) ввести исходные данные в интеллектуальные системы; б) преобразовать их в информацию и знания; в) использовать знания для решения задач прогнозирования и принятия решений. Для этого в АСК-анализе предусмотрены следующие этапы [1]:

1. Когнитивная структуризация предметной области, при которой определяется, что мы хотим прогнозировать и на основе чего (конструирование классификационных и описательных шкал).

2. Формализация предметной области, т.е. 1) разработка градаций классификационных и описательных шкал номинального, порядкового и числового типа; 2) использование разработанных на предыдущих этапах классификационных и описательных шкал и градаций для формального описания (кодирования) исследуемой выборки [30].

3. Синтез и верификация (оценка степени адекватности) модели [31].

² <http://lingvo.yandex.ru/cognition/c%20английского/>

4. Если модель адекватна, то использование ее для решения задач идентификации, прогнозирования и принятия решений, а также для исследования моделируемой предметной области.

Данные – это информация, рассматриваемая безотносительно к ее смысловому содержанию, находящаяся на носителях или в каналах связи и представленная в определенной системе кодирования или на определенном языке (т.е. в формализованном виде).

Информация – это осмысленные данные. Смысл, семантика, содержание (согласно концепции смысла Шенка-Абельсона [5]) – это знание причинно-следственных зависимостей.

Знания – это информация, *полезная* для достижения целей.

Процесс преобразования данных в информацию – это анализ данных, т.е. 1) выявление в них *событий*; 2) выявление причинно-следственных связей (зависимостей) между этими событиями.

Факт наличия причинно-следственных зависимостей может быть установлен методом хи-квадрат, а ее вид – многофакторным анализом. Однако факторный анализ позволяет обрабатывать данные лишь очень небольших размерностей (по числу факторов) и предъявляет чрезвычайно жесткие требования к наличию полных повторностей всех вариантов сочетаний факторов в исходных данных (т.е. данные не должны быть фрагментарными), что на практике выполнить удается крайне редко.

Поэтому большой интерес представляют другие подходы к решению **задачи** выявления в эмпирических данных причинно-следственных зависимостей и их вида, отражения выявленных зависимостей в наглядной графической и аналитической форме.

Рассмотрим вариант решения этой задачи, развиваемый в АСК-анализе и реализованный в системе Эйдос».

Для этого сформулируем *требования* к форме представления данных, информации и знаний, позволяющие оценить *степень их пригодности* для решения задач прогнозирования и принятия решений, а также исследования предметной области (например, кластерного анализа).

Прежде всего, результаты решения вышеперечисленных задач должны быть *инвариантны* относительно:

- *единиц измерения* градаций факторов (признаков);
- *типов шкал*, используемых для формализации классов и факторов (номинальные, порядковые и числовые);
- различных *статистических характеристик* исходной выборки: частотных распределений объектов по классам (обобщенным категориям), частотных распределений градаций факторов, различий в количестве признаков в описаниях объектов исследуемой выборки, различий в суммарном количестве признаков по классам.

Кроме того, форма представления должна обеспечивать решение вышеперечисленных задач с минимальными дополнительными затратами

ручного труда, а это значит, что *вся предварительная обработка должна быть максимально автоматизирована*.

Эти требования можно рассматривать и как *критерии* выбора наиболее подходящей для решения вышеперечисленных задач формы представления данных, информации и знаний.

Рассмотрим **влияние единиц измерения в исходной выборке на результаты решения задач** прогнозирования и принятия решений, а также исследования предметной области (например, кластерного анализа).

Если в исходных данных какие-то значения выражены в больших единицах измерения, то их числовые значения будут малыми, и наоборот, если единицы измерения мелкие, то числовые значения – большие. Большие значения оказывают большее влияние на результаты математической обработки, чем малые, и *это приводит к возникновению зависимости результатов решения задач идентификации, прогнозирования и принятия решений, а также кластерного анализа, от выбранных размерностей исходных данных, что, на взгляд авторов, совершенно недопустимо и указывает на то, что такое решение нельзя признать корректным и даже вообще решением*. По этой же причине некорректно совместно обрабатывать сами исходные данные, представленные в **различных** единицах измерения (натуральных или ценовых), например, складывать расстояния, представленные в километрах и в метрах, а затем прибавлять к ним тонны и килограммы, а затем еще и безразмерные величины, хотя, как ни удивительно, но как показывает опыт на практике это довольно часто делается. Странно, что обычно на это *не обращают никакого внимания* при использовании исходных данных, представленных в различных единицах измерения. Например, даже в таких популярных (причем, совершенно заслуженно) системах, как SPSS, в подсистеме кластерного анализа приводятся примеры кластерного анализа над исходными данными, представленными в различных единицах измерения.

Для решения поставленной задачи в АСК-анализе проводится последовательное повышение степени формализации исходных данных до уровня, обеспечивающего их обработку на компьютере в программной системе. После выполнения когнитивной структуризации и формализации предметной области осуществляется синтез модели. Он включает в себя расчет на основе эмпирических данных, представленных в исследуемой выборке, следующих матриц:

- матрицы абсолютных частот (большинство статических систем этим и ограничиваются);
- матрицы условных и безусловных процентных распределений (в некоторых системах это также делается);
- матрицы информативностей или матрицы знаний (что осуществляется только в АСК-анализе).

Рассмотрим, используя вышеперечисленные критерии, в какой степени эти матрицы *пригодны* для решения задач прогнозирования и принятия решений, а также исследования предметной области (например, кластерного анализа) и какую работу необходимо выполнять вручную и автоматизировать, чтобы повысить их пригодность для этого.

Матрица абсолютных частот отражает, сколько раз каждая градация факторов встречается у объектов каждого класса.

Проблема размерностей при расчете матрицы абсолютных частот решается тем, что *сами размерные исходные данные с использованием шкал различных типов (номинальных, порядковых и числовых) заменяются на факты их встречи*, т.е. на частоты встреч тех или иных их *интервальных значений* [31] в различных группах, соответствующих классам. Фактом является наблюдение определенного экстенционального значения (признака, градации фактора) у объекта исходной выборки, относящегося к некоторой интенциональной категории (классу).

Однако вышеперечисленные задачи решать на основе абсолютных частот можно только в том случае, если по каждому классу в исходных данных было приведено *одинаковое* количество примеров, что на практике встречается крайне редко и является трудно достижимым при сборе исходных данных, за исключением случая жестко спланированного управляемого эксперимента (обычно очень небольшой размерности). Можно, конечно, вручную учитывать это различие, однако реально это возможно сделать только на моделях очень небольшой размерности и требует специальных усилий (работы).

Чтобы результаты решения вышеперечисленных задач *не зависели* от количества примеров по разным классам (т.е. были *инвариантны* относительно формы частотных распределений примеров по классам, частотного распределения признаков и др.) можно перейти от матрицы абсолютных частот к матрице условных и безусловных процентных распределений (матрице *относительных частот* или частостей³).

При неограниченном увеличении объема выборки частоты стремятся (сходятся) к вероятностям, как своим пределам. Способ, которым частоты приближаются к вероятностям, называется *сходимостью* модели. В системе «Эйдос» реализован специальный режим, позволяющий исследовать *сходимость модели*, в том числе *скорость сходимости* и погрешность различия частоты и вероятности при различных объемах исследуемой выборки. Учитывая все это при достаточно больших выборках, по мнению авторов, допустимо вместо термина «частость» использовать термин «ус-

³ **Частота** (абсолютная частота) – количество элементов совокупности, которые имеют данное значение признака. **Частость** (относительная частота) – отношение частоты к общему количеству исследуемых элементов, т.е. объему совокупности.

ловная вероятность», тем более что в аналитических выражениях обычно оперируют именно вероятностями⁴.

Однако и при решении вышеперечисленных задач на основе матрицы условных и безусловных процентных распределений приходится *вручную осуществлять сравнение* условных относительных частот, что реально возможно только на моделях очень малой размерности и требует довольно больших специальных усилий. Поэтому есть смысл *автоматизировать* и это сравнение, так, чтобы в нашем распоряжении была матрица, содержащая уже сами **результаты** сравнения условных относительных частот в количественной форме.

Для того чтобы реализовать эту автоматизацию необходимо *выбрать базу сравнения и способ сравнения*, т.е. ответить на два вопроса:

– *с чем сравнивать* условные относительные частоты: друг с другом или с безусловными частотами;

– *как сравнивать* условные относительные частоты: с помощью вычитания или с помощью деления.

Если в модели есть всего два класса, то можно сравнивать условные относительные частоты как друг с другом, так и с безусловными частотами, т.к. это одинаково как по трудоемкости (затрачиваемым вычислительным ресурсам), так и по результатам сравнения. Если же в модели хотя бы три класса, то уже возникают определенные затруднения в том, как сравнить условные процентные распределения по ним, а если их сотни или тысячи, то это становится даже в теоретическом плане непонятным. Поэтому в [1] предлагается *использовать в качестве базы для сравнения (нормы) условных относительных частот их взвешенное среднее по всей исследуемой выборке или безусловные частоты*.

Что касается вопроса о том, вычитание или деление для этого сравнения использовать, то этот вопрос не является принципиальным, т.к. различие между вычитанием и делением сводится к выбору единиц измерения результатов сравнения: если взять логарифм от отношения, то получится разность логарифмов, которая ведет себя точно также, как разность логарифмируемых выражений.

Переход от матрицы абсолютных частот к матрице условных и безусловных процентных распределений обеспечивает инвариантность результатов решения вышеперечисленных задач от формы частотного распределения примеров по классам, однако при этом никак не решается вопрос о зависимости этих результатов от **размерностей** различных градаций факторов (признаков) и типов шкал, используемых для формализации факторов.

Проблему размерностей можно было бы решить, перейдя к *стандартизированным величинам*⁵ или отношениям условных и безусловных

⁴ Процентные распределения отличаются от вероятностных тем, что вероятности умножены на 100.

вероятностей. Например, *формулу Байеса*⁵ можно рассматривать как дающую количественную оценку степени влияния фактора на наступление некоторого события. Отношение условной вероятности наблюдения некоторого значения фактора в группе (классе) к безусловной вероятности его наблюдения по всей исследуемой выборке также можно рассматривать как количественную меру силы и направления его влияния на переход объекта в состояние, соответствующее классу, т.е. как количественную оценку силы и направления причинно-следственной связи между ними.

Возникает вопрос о том, каким образом формально описать влияние на объект не отдельных значений факторов, а всей их системы. Для того чтобы это сделать введем понятие частных критериев и интегрального критерия.

Частным критерием будем называть выраженное в количественной форме влияние отдельного значения фактора на переход объекта в различные состояния.

Это значит, что отношение условной вероятности наблюдения некоторого значения фактора в группе (классе) к безусловной вероятности его наблюдения по всей исследуемой выборке можно, рассматривать как частный критерий.

Тогда, если значение фактора *способствует* переходу объекта в некоторое состояние, то отношение условной вероятности наблюдения этого значения фактора в группе (классе), соответствующей данному состоянию, будет *больше* безусловной вероятности его наблюдения по всей исследуемой выборке и этот критерий будет иметь значение больше 1.

Если значение фактора *препятствует* переходу объекта в некоторое состояние, то отношение условной вероятности наблюдения этого значения фактора в группе (классе), соответствующей данному состоянию, будет *меньше* безусловной вероятности его наблюдения по всей исследуемой выборке и этот критерий будет иметь значение меньше 1.

Если же значение фактора *никак не влияет на* переход объекта в некоторое состояние, то отношение условной вероятности наблюдения этого значения фактора в группе (классе), соответствующей данному состоянию, будет *равно* безусловной вероятности его наблюдения по всей исследуемой выборке и этот критерий будет иметь значение равное 1.

Интегральным критерием будем называть некоторое аналитическое выражение от частных критериев, которое количественно отражает силу влияния системы факторов на переход объекта в различные состояния.

Моделируемый объект является *линейным*, если результат совместного действия на него совокупности факторов является *суммой* результа-

⁵ При стандартизации вектора каждая его координата заменяется на отношение ее разности со средним по всем координатам к среднеквадратичному отклонению координат от среднего.

⁶ <http://ru.wikipedia.org/wiki/Теорема%20Байеса>

тов влияния на него каждого из этих факторов в отдельности, т.е. выполняется *принцип суперпозиции*⁷ факторов. Чем меньше интенсивность взаимодействия между факторами в объекте, *тем ближе система факторов к множеству* [29] и тем ближе объект к линейному. Таким образом, для линейных объектов можно обоснованно считать, что взаимодействие между факторами в этих объектах отсутствует, т.е. по сути можно считать, что на них действует не система факторов, а множество факторов.

Для *линейных* объектов интегральный критерий, отражающий совместное влияние факторов на объект, можно представить в форме *суммы* влияния каждого из этих факторов в отдельности, т.е. в форме суммы частных критериев, т.е. для линейных объектов оправданно и обоснованно использовать *аддитивный интегральный критерий*.

Приведенные выше количественные меры силы и направления причинно-следственных связей очень неудобны для использования подобных в качестве частных критериев, в основном потому, что в случае *отсутствия* влияния фактора они равны 1. В результате в аддитивном интегральном критерии будет присутствовать некое слагаемое, равное количеству недействующих факторов, и для каждого класса это слагаемое будет свое. В результате подобный интегральный критерий окажется просто непригодным для оценки влияния совокупности факторов на поведение объекта.

Поэтому эти частные критерии необходимо *нормировать* так, чтобы в случае отсутствия влияния он принимали значение равное нулю, а не единице. Есть много вариантов осуществить подобную нормировку, из которых наиболее очевидными являются:

- вычесть 1 из отношения условной вероятности к безусловной;
- взять логарифм от отношения условной вероятности к безусловной.

Первый вариант нормировки приводит к показателям типа ROI⁸ (количественная оценка степени полезности инвестиций) и различным его обобщениям. Второй вариант сразу приводит к семантической мере целесообразности информации А.Харкевича. Из этих вариантов для количественной оценки степени полезности информации для достижения целей предпочтительным является применение меры А.Харкевича [7]. Это связано с тем, что использование логарифма в этой мере позволяет привлечь огромный пласт понятий, связанных с данными, информацией и знаниями, что является для нас очень ценным.

Очень важно, что этот подход позволяет автоматически решить проблему сопоставимой обработки многих факторов, измеряемых в различных единицах измерения, т.к. в этом подходе рассматриваются не сами факторы, какой бы природы они не были и какими бы шкалами не

⁷ <http://slovari.yandex.ru/~книги/БСЭ/Суперпозиции%20принцип/>

⁸ <http://ru.wikipedia.org/wiki/ROI>

формализовались, а количество информации, которое в них содержится о поведении моделируемого объекта [30, 31].

Необходимо также отметить, что представление о полностью линейных объектах (системах) является абстракцией и реально все объекты являются принципиально нелинейными. Вместе с тем для большинства систем нелинейные эффекты можно считать эффектами второго и более высоких порядков и такие системы *в первом приближении* можно считать линейными. Возможны различные модели *взаимодействия факторов*, в частности, развиваемые в форме системного обобщения теории множеств [29]. Этот подход в перспективе может стать одним из вариантов развития теории нелинейных систем.

Отметим, что математическая модель АСК-анализа (системная теория информации) *органично* учитывает принципиальную нелинейность всех объектов. Это проявляется в нелокальности нейронной сети системы «Эйдос» [30], приводящей к зависимости *всех* информативностей от *любого* изменения в исходных данных, а не как в методе обратного распространения ошибки⁹. В результате *значения матрицы информативностей количественно отражают факторы не как множество, а как систему.*

Объект может перейти в некоторое будущее состояние под действием различного количества факторов, но какая бы система факторов не обуславливала (детерминировала) этот переход, в ней не может содержаться информации больше, чем можно получить, точно узнав, что объект переходит в данное состояние. Это количество информации в АСК-анализе называется «Теоретически максимальное количество информации» и определяется только количеством классов (будущих состояний объекта), которые в детерминистском случае равновероятны, т.к. между классами и факторами выполняется взаимнооднозначное соответствие, когда каждое будущее состояние однозначно определяется единственным фактором. Формула А.Харкевича видоизменена в работе [1] таким образом, чтобы удовлетворять принципу соответствия с формулой Р.Хартли в детерминистском случае. Поэтому, чем меньше факторов, тем жестче ими детерминировано поведение объекта, и наоборот, чем больше этих факторов, тем меньше влияние каждого из них на поведение объекта. Например, если переход объекта в некоторое состояние однозначно определяется единственным фактором, то добавление в модель еще одного *точно такого же* фактора приводит к тому, что в сумме эти два фактора будут оказывать тоже самое влияние, которое делится между ними поровну.

Так в математической модели АСК-анализа учитывается *взаимодействие* факторов и отличие *системы* факторов от *множества* факторов [29], являющееся источником нелинейности моделируемого объекта.

⁹ <http://ru.wikipedia.org/wiki/Метод%20обратного%20распространения%20ошибки>

Итак, в матрице информативностей количественно отражены сила и направление влияния каждого значения фактора на переход объекта в каждое из состояний, а также учтено, что совокупность факторов является системой, а не множеством, т.е. учтены взаимодействие факторов и нелинейность моделируемого объекта. Результаты решения задач идентификации, прогнозирования, принятия решений и научного исследования моделируемой предметной области (в частности кластерно-конструктивного анализа), на основе матрицы информативностей **инвариантны** относительно формы частотного распределения объектов исследуемой выборки по классам, единиц измерения значений факторов и типа шкал, используемых для формализации факторов.

Это позволяет корректно использовать в АСК-анализе аддитивный интегральный критерий в форме суммы частных критериев не только для линейных, но и для нелинейных объектов.

Различие между матрицей информативностей и матрицей знаний. Если в модели отражены лишь причинно-следственные связи между факторами и будущими состояниями объекта, но не отражена степень желательности ли нежелательности этих будущих состояний, то мы имеем дело с матрицей информативностей. Если же некоторые из будущих событий классифицируются как желательные, т.е. целевые, а другие как нежелательные, то появляется возможность количественной оценки степени *полезности* информации о действии факторов для перевода объекта в эти состояния, т.е. для преобразования информации в знания.

Процесс преобразования информации в знания – это процесс оценки степени полезности информации для достижения желаемых будущих состояний, т.е. целей.

Таким образом, матрица знаний количественно отражает степень полезности (а также бесполезности и вредности) факторов для достижения целей: она содержит знания в количественной форме о *величине и направлении* влияния каждого значения фактора на перевод объекта в каждое из будущих состояний, как желаемое, так и нежелательное.

Факт – это единство экстенционального и интенционального описания *события*, обнаруженного эмпирическим путем, т.е. по сути, *факт это определение события*. Пример факта: «Кошка кормит котят молоком». Пример определения в науке: «Млекопитающее – это животное (более общее, интенциональное понятие), вскармливающее своих детей молоком (экстенциональный специфический признак)».

Закономерности – это причинно-следственные зависимости, выявленные на исследуемой выборке и распространяемые лишь на саму эту выборку.

Эмпирический закон – это причинно-следственные зависимости, выявленные на исследуемой выборке и распространяемые на некоторую предметную область, более широкую, чем исследуемая выборка, в которой

действуют *те же причины действия* причинно-следственных зависимостей, что и в исследуемой выборке, на которой он обнаружены. Эта более широкая предметная область называется генеральной совокупностью, по отношению к которой исследуемая выборка репрезентативна. Эмпирический закон является феноменологическим, т.е. внешним описанием зависимости последствий от причин, который не раскрывает механизма или способа, которым реализуется эта зависимость.

Научный закон – это содержательная интерпретация механизма действия эмпирического закона, т.е. *способа* преобразования причин в следствия. Научный закон является содержательным *объяснением* и интерпретацией эмпирического закона. Это объяснение, когда оно разрабатывается, не сразу становится научным законом, а сначала имеет статус научной гипотезы и приобретает статус научного закона лишь после того, как *на практике*, т.е. эмпирически, подтверждаются предсказания существования новых, ранее неизвестных явлений, сделанные на основе научной гипотезы. Таким образом, научный закон – это научная гипотеза, адекватность и прогностическая сила которой подтверждены (верифицированы) эмпирически. Процесс преобразования научной гипотезы в научный закон – это процесс подтверждения на практике адекватности этой научной гипотезы.

Необходимо подчеркнуть, что существует принципиальная возможность создания *многих различных моделей*, одинаково адекватно отражающих одну и ту же предметную область. Иногда такие модели и действительно созданы. Тогда возникает вопрос о *критериях выбора* одной модели, в определенном смысле «наилучшей» из многих. Среди этих критериев следует отметить адекватность, удовлетворение принципу соответствия и широту адекватно отражаемой предметной области, а также ее простоту и красоту. Из многих моделей предпочтительная та, которая более адекватна, та, которая адекватно отражает более широкую предметную область и включает в себя на основе принципа соответствия другие известные модели, а также более простая и красивая модель. Однако часто бывает, что разработка многих моделей (научных теорий) весьма затруднительна и есть или известна всего лишь одна-единственная модель. Тогда эта модель автоматически становится наилучшей из всех известных.

Возникает соблазн неоправданно и необоснованно считать, что реальность устроена именно таким образом, какой она отражается в этой наилучшей по сформулированным выше критериям модели или научной теории, т.е. *необоснованно придать онтологический статус абстрактной модели*. В этом состоит широко распространенная малозаметная ошибка познания, называемая «Гипостазирование¹⁰». Однако эта ошибка влечет за собой целый шлейф весьма заметных последствий, важнейшим из которых является отрицание существования фактов, закономерностей и эмпириче-

¹⁰ <http://yandex.ru/yandsearch?text=гипостазирование>

ских законов, не вписывающихся в те или иные научные теории, даже если эти факты в буквальном смысле слова *очевидны*. Например, апологеты воздухоплавания отрицали возможность летательных аппаратов тяжелее воздуха, не смотря на птиц, которые садились и взлетали перед ними (или даже смотря на них, но не осознавая, что они видят). При этом они исходили из того, что принцип действия летательных аппаратов может быть основан только на законе Архимеда, как это следовало из единственной известной им научной теории полета. Однако существуют и другие принципы полета: в частности, баллистический, аэродинамический, ракетный, электромагнитный, на которых может быть основан принцип действия летательных аппаратов тяжелее воздуха, причем эти аппараты ни в коей мере не нарушают закон Архимеда и полностью ему подчиняются.

Признание существования факта не зависит от обнаружения закономерности. Признание существования закономерности не зависит от обнаружения соответствующего эмпирического закона. Признание существования эмпирических законов не зависит от наличия верифицированной содержательной интерпретации или научного закона, а если она есть, то от того, является ли она «правильной» или «неправильной» по тем или иным критериям или по чьему-то мнению. Таким образом, *признание существования факта не зависит от наличия теории, которая его объясняет, и отсутствие такой теории не является основанием для отрицания существования или непризнания существования факта.*

Когнитивные функции представляют собой новый перспективный инструмент отражения и наглядной визуализации закономерностей и эмпирических законов. Разработка содержательной научной интерпретации когнитивных функций представляет собой способ познания природы, общества и человека.

Когнитивные функции могут быть:

- прямые, отражающие зависимость классов от признаков, обобщающие информационные портреты признаков;
- обратные, отражающие зависимость признаков от классов, обобщающие информационные портреты классов;
- позитивные, показывающие чему способствуют система детерминации;
- негативные, отражающие чему препятствуют система детерминации;
- средневзвешенные, отражающие совокупное влияние всех значимых факторов на поведение объекта;
- с различной степенью редукции или степенью детерминации, которая отражает в графической форме (в форме полосы) количество знаний в аргументе о значении функции и является аналогом и обобщением доверительного интервала.

Прямая и обратная, а также позитивная и негативная когнитивные функции *полностью совпадают (тождественны)* друг с другом только для жестко (т.е. полностью) детерминированных систем. Это связано с тем, что матрица знаний, моделирующая полностью детерминированную систему, в которой между значениями аргумента и значениями функции существует взаимнооднозначное соответствие, представляет собой диагональную матрицу [1]. Можно обоснованно предположить, что степень совпадения прямой и обратной когнитивных функций пропорциональна степени детерминированности моделируемой системы. Если *интерпретировать* значения факторов, обуславливающих поведение системы, как ее экстенциональное описание, относящееся к ее прошлому времени, а классы – как интенциональное описание ее будущих состояний, то можно сказать, что степень детерминации поведения системы тем выше, чем более сходным являются влияние на нее прямой и обратной причинности, т.е. если влияние прошлого на будущее совпадет с влиянием будущего на прошлое. Чем сильнее влияние прошлого на будущее отличается от влияния будущего на прошлое, тем слабее детерминированность в поведении системы, тем ближе оно к случайному. При этом рассмотрение вопросов о физическом механизме прямой и обратной причинности, как и самом существовании обратной причинности, не входит в задачи данной работы.

Матрица информативности может быть использована для выявления и визуализации *когнитивных функциональных зависимостей* в фрагментированных и зашумленных данных большой размерности [8]. Кратко поясним суть этого метода. Матрица информативностей рассчитывается на основе системной теории информации [1] непосредственно на основе эмпирических данных и представляет собой таблицу, в которой столбцы соответствуют *обобщенным* образам классов, т.е. будущим состояниям моделируемой системы, строки – значениям факторов, влияющих на эту систему, а на пересечениях строк и столбцов находится количество информации, которое содержится в факте действия значения фактора, соответствующего строке, на переход системы в состояние, соответствующее столбцу. Максимальное количество информации, которое может быть в значении фактора, определяется числом будущих состояний моделируемой системы. Модуль количества информации отражает силу влияния значения фактора, а знак – направление этого влияния, т.е. то, способствует он или препятствует наступлению данного состояния. Если последовательности классов и значений факторов образуют порядковые шкалы или шкалы отношений, т.е. соответственно, на них определены отношения «больше-меньше» или, кроме того, единица измерения, начало отсчета и арифметические операции, то матрица информативностей допускает наглядную

графическую визуализацию, *традиционного* для функций типа, когда значения факторов рассматриваются в качестве значений аргумента, а классы, о наступлении которых в этих значениях факторов содержится *максимальное* количество информации – в качестве значений функции. Другие классы, менее обусловленные данным значением фактора, а также те, наступлению которых это значение препятствует в большей или меньшей степени, также могут отображаться соответствующими цветами, и это также может представлять интерес, т.к. позволяет задействовать мощные способности человека к анализу изображений. Когнитивные функции, представляемые в форме матрицы информативностей, соответствуют очень общему виду функциональной зависимости: *многозначной функции многих аргументов*, т.к. каждое значение фактора влияет на все состояния моделируемого объекта, и каждое его состояние обусловлено всеми значениями факторов. Простой пример визуализации матрицы информативностей, полученной на выборке, отражающей зависимость амплитуды затухающего гармонического колебания от времени, приведен на нижеследующем рисунке 1, взятом из работы [8], в котором степень детерминации значения функции значением аргумента показана различными цветами: теплые цвета – высокая степень детерминации, холодные – низкая.

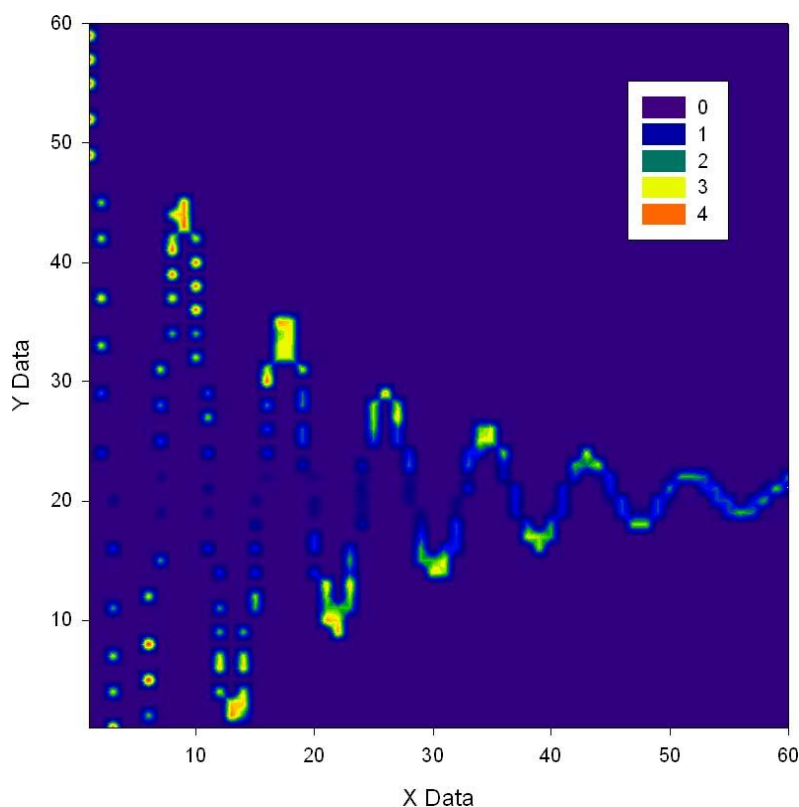


Рисунок 1. Количество информации в значении аргумента о значении функции для нечеткой взаимнооднозначной когнитивной функции

Для визуализации матрицы информативностей использовалась система SigmaPlot for Windows version 10.0. Для преобразования матрицы информативностей в форму, удобную для использования в системе SigmaPlot, применялся режим _683 системы «Эйдос».

На рисунке 2, взятом из работы [10], представлена *когнитивная функция*, отражающая причинно-следственную зависимость объемов производства молока по всем категориям хозяйств АПК от доли оплаты труда в затратах.

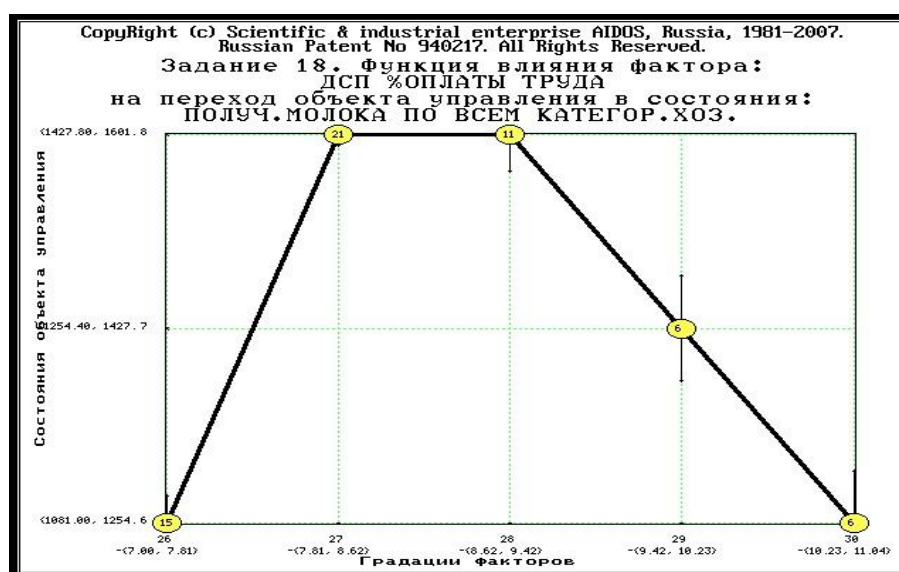


Рисунок 2. Когнитивная функция зависимости объемов производства молока по всем категориям хозяйств АПК от доли в затратах оплаты труда

Прямая полностью редуцированная когнитивная функция, представленная на рисунке 2, визуализирована средствами системы «Эйдос». Количество информации в значениях аргумента о значениях функции показано интервалами. Уравнение регрессии ($R^2=1$), представляющее эту когнитивную функцию в аналитической форме, имеет вид:

$$y = 0,422x^4 + 39,72x^3 - 1414,9x^2 + 14447x - 45629$$

Другие примеры применения аппарата когнитивных функций в экономических исследованиях приведены в работе [28].

Таким образом, использование интервальных значений аргумента и функции позволяет с применением теории информации непосредственно на основе эмпирических данных рассчитать, какое количество информации содержится в каждом значении аргумента о каждом значении функции.

При этом получается, что каждому значению аргумента соответствует не одно, а много значений функции, но соответствуют в различной степени.

Заметим, что ход времени, т.е. процесс преобразования неопределенного многовариантного будущего в определенное безальтернативное настоящее, можно рассматривать как процесс *редукции* многозначных когнитивных функций, отражающих будущее, в однозначные когнитивные функции, отражающие настоящее. *При преобразовании неопределенного будущего в определенное настоящее происходит генерация информации, количество которой соответствует степени уменьшения неопределенности, точно так же, как в процессе измерения или познания.* Поэтому можно обоснованно утверждать, что многозначные когнитивные функции аналогичны по смыслу волновой функции (функция плотности вероятности) квантовой механики (КМ) и квантовой теории поля (КТП), а преобразование многозначной когнитивной функции в однозначную аналогично процессу *редукции* волновой функции в процессе измерения. Процесс редукции волновой функции и другие квантовые и волновые явления тесно связаны с информацией и могут рассматриваться как информационные процессы и явления [1]. Поэтому авторами вводятся понятия нередуцированной, полностью и частично редуцированной когнитивной функции. Отметим, что понятие частично-редуцированного состояния объекта (на примере электрона), по-видимому, впервые введено Ричардом Фейнманом [25]¹¹.

Для визуализации матрицы информативностей (3) первоначально использовалась система SigmaPlot for Windows version 10.0. Для преобразования матрицы информативностей в форму, удобную для использования в системе SigmaPlot, применялись режимы _683 или _53 системы «Эйдос».

В дальнейшем была создана¹² подсистема визуализации когнитивных (каузальных) функций системы «Эйдос» (Подсистема «Эйдос-VCF» или InfVisual) [26], вошедшая в состав базовой системы «Эйдос» (режим _54) и в состав системы «Эйдос-астра» [27], позволяющая осуществлять визуализацию, как всей матрицы информативностей, так и ее фрагментов, а также строить прямые и обратные когнитивные функции различной степени редукции.

¹¹ <http://ru.wikipedia.org/wiki/Фейнман,%20Ричард%20Филлипс>

¹² Система программирования: Turbo Delphi Explorer

На рисунке 3, взятом из работы [8], представлены три *нечеткие многозначные функции*, в которых каждому значению аргумента в различной степени соответствует много различных значений функции. Рисунок 3 построен на основе Оксфордской базы данных содержащей среднемесячные метеорологические данные по температуре в приземном слое воздуха, осадкам и заморозкам в период с января 1853 г по сентябрь 2009 г. (см. <http://www.metoffice.gov.uk/>). Красный цвет на рисунке соответствует *максимальному* количеству информации в номере месяца о соответствующей среднемесячной температуре, т.е. *высокой степени детерминации* в отображаемой причинно-следственной зависимости.

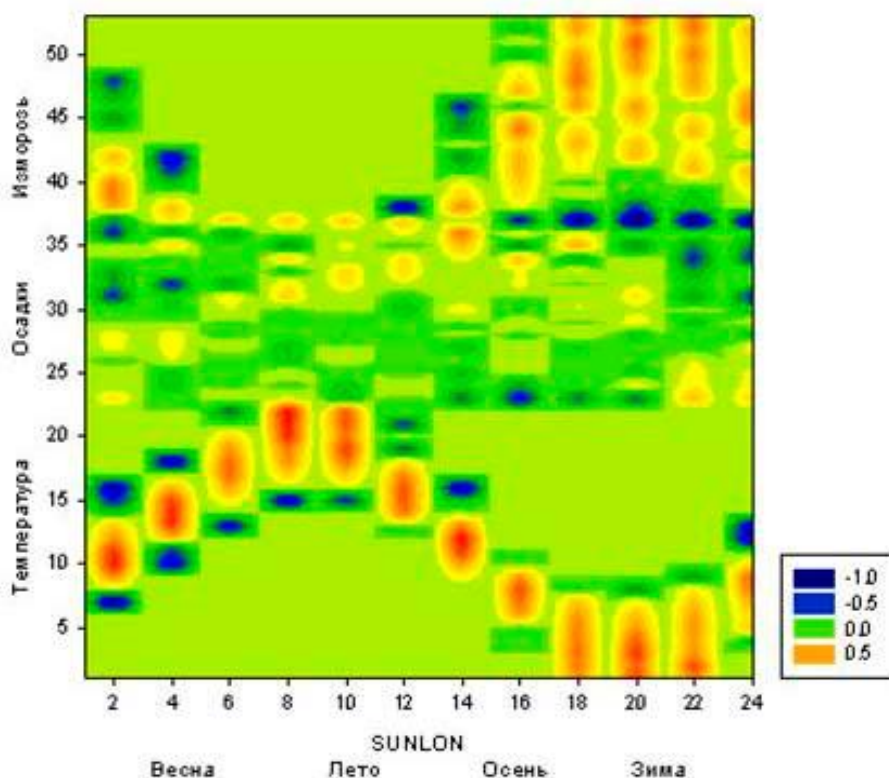


Рисунок 3. Количество информации в значении аргумента о значении функции для нечетких многозначных функций

Для визуализации матрицы информативностей использовалась система SigmaPlot for Windows version 10.0. Для преобразования матрицы информативностей в форму, удобную для использования в системе SigmaPlot, применялся режим _683 системы «Эйдос».

Из этого рисунка видно также, что зимой в Оксфорде заморозки более вероятны, чем летом, а осадки более вероятны летом, чем весной. Данные по осадкам менее информативны в сравнении с температурой, а дан-

ные по заморозкам (иней в воздухе), хотя и являются информативными, но обладают более существенной неоднозначностью, т.к. в разные годы число дней заморозков в данный месяц изменяется в более широких пределах, чем температура.

В базовой системе «Эйдос» реализован режим _53, обеспечивающий подготовку баз данных, содержащих *выборки по подматрицам* матрицы знаний в форме, наиболее удобной для их визуализации во внешних системах SigmaPlot и MS Excel. В данном режиме реализована возможность автоматического определения классификационных шкал и градаций по заданному разделителю, а также задания их вручную. Некоторые экранные формы этого режима приведены на рисунке 4, результаты визуализации – на рисунках 5-7.

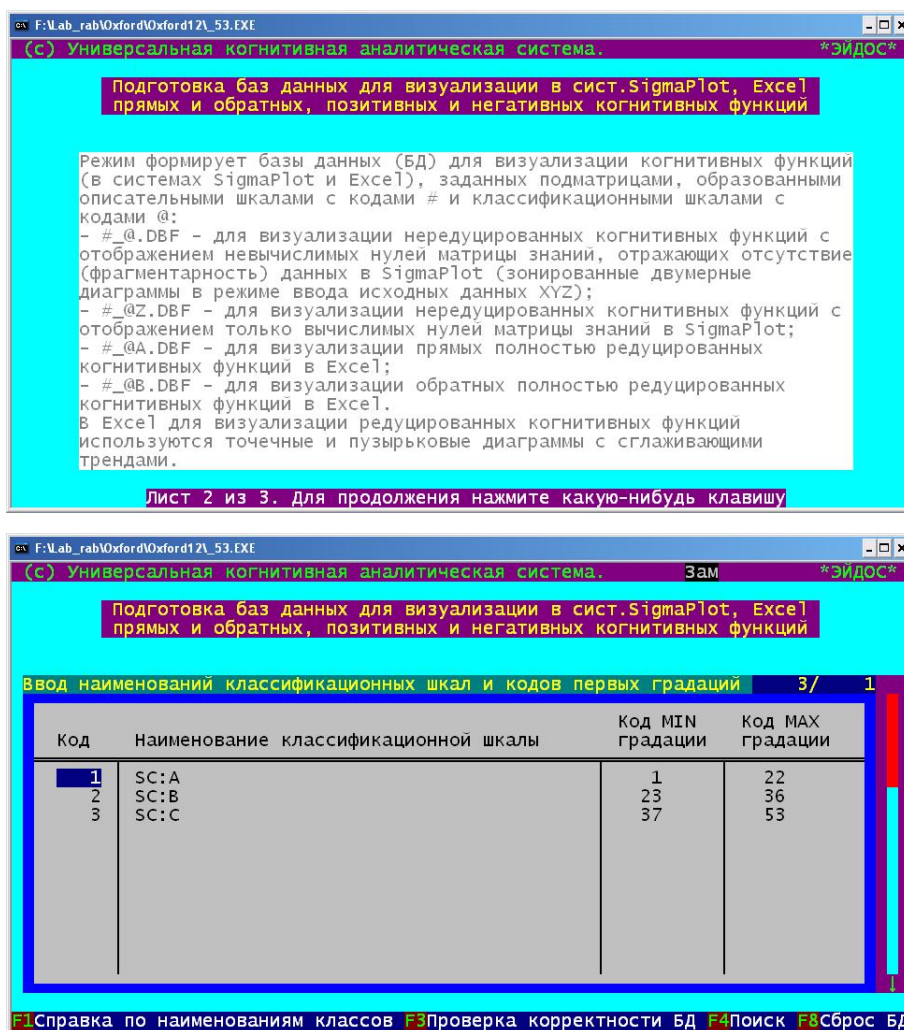


Рисунок 4. Экранные формы режима _53 базовой системы «Эйдос»

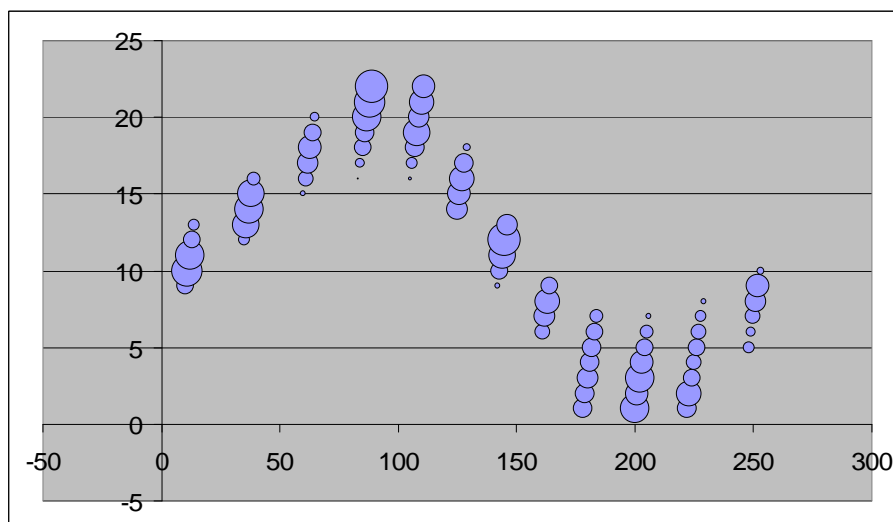


Рисунок 5. Прямая нередуцированная когнитивная функция, построенная в MS Excel на основе базы знаний, сформированной по Оксфордской базе данных

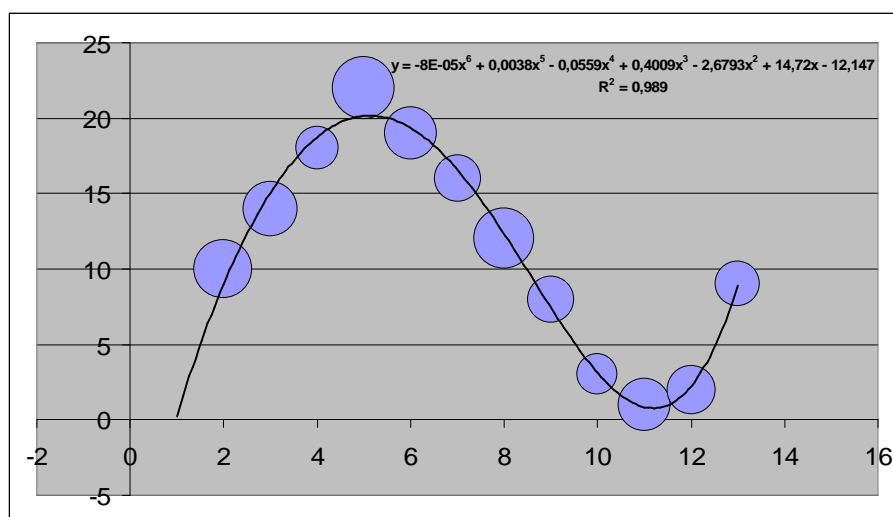


Рисунок 6. Прямая полностью редуцированная когнитивная функция, построенная в MS Excel на основе базы знаний, сформированной по Оксфордской базе данных.

Диаметр пузырьков на рисунках 5 и 6 *пропорционален* количеству информации в значении аргумента о значении функции.

Важно, что на основе матриц, формируемых режимом _53 системы «Эйдос», используя стандартные возможности MS Excel, можно получить *регрессии*, аппроксимирующие найденные зависимости, например для зависимости, представленной на рисунке 6, регрессия имеет вид ($R^2 = 0,989$):

$$y = -8E-05x^6 + 0,0038x^5 - 0,0559x^4 + 0,4009x^3 - 2,6793x^2 + 14,72x - 12,147$$

При визуализации в системе SigmaPlot той же подматрицы, что и на рисунках 5, 6, получим рисунок 7, а с применением подсистемы «Эйдос-VCF» (или InfVisual) – рисунок 8:

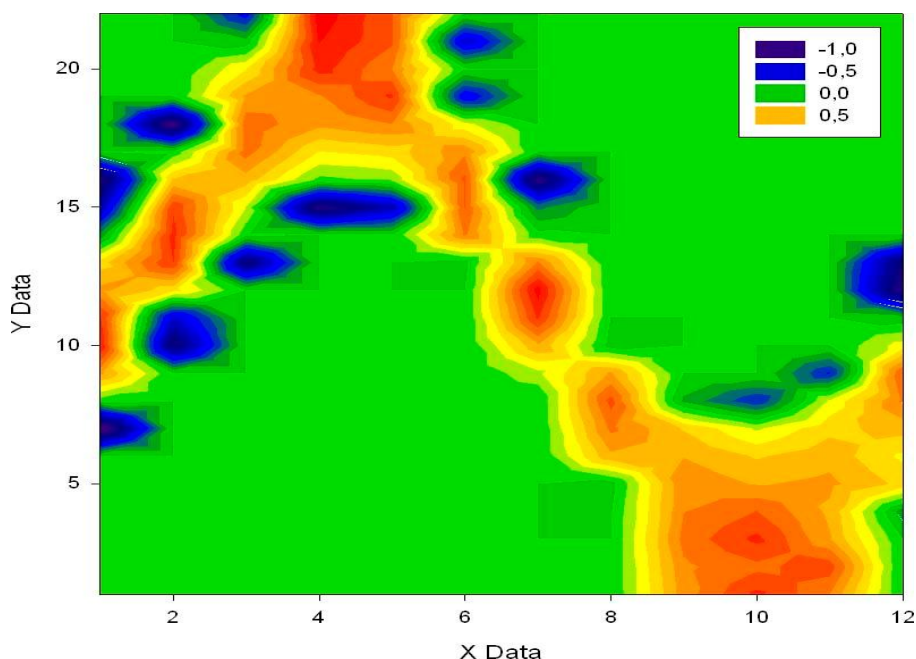


Рисунок 7. Прямая нередуцированная когнитивная функция, построенная в SigmaPlot на основе базы знаний, сформированной по Оксфордской базе данных

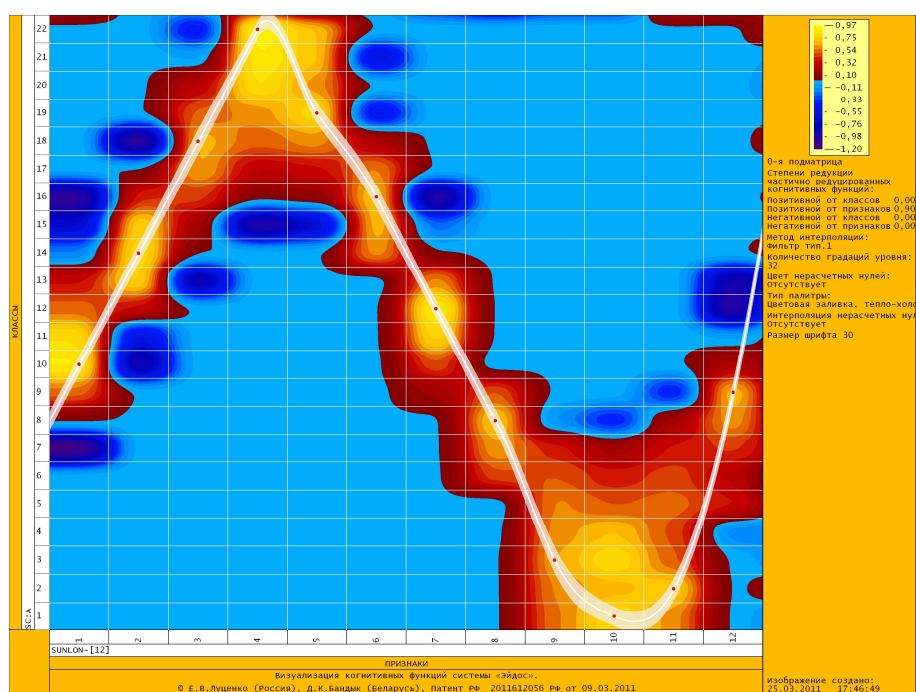


Рисунок 8. Прямая частично редуцированная когнитивная функция, построенная в режиме _54 базовой системы «Эйдос» – подсистеме InfVisual (27) на основе базы знаний, сформированной по Оксфордской базе данных

Подсистема «Эйдос-VCF» (или InfVisual) [26] *предназначена* для визуализации когнитивных (каузальных) на основе матриц знаний базовой системы «Эйдос» и систем окружения и *может применяться* в научных организациях, применяющих системы искусственного интеллекта для решения задач прогнозирования, поддержки принятия решений и научных исследований в различных предметных областях. Она *обеспечивает* выполнение следующих *функций*:

- импорт данных из универсальной когнитивной аналитической системы «Эйдос», интеллектуальной системы научных исследований глобальных геосистем «Эйдос-астра» и других систем окружения;
- задание в диалоге детальных параметров визуализации когнитивных (каузальных) функций;
- визуализацию когнитивных (каузальных) функций: прямых и обратных, позитивных и негативных, с заданной степенью редукции, а также запись полученных изображений в виде файлов.

Для построения изолиний (линий уровня) трехмерных поверхностей когнитивных функций в подсистеме используется триангуляция Делоне¹³ и авторские методы сглаживания, разработанные Д.К.Бандык. Для интерполяции когнитивной функции применен сплайн Акимы¹⁴.

На рисунке 9 приведена главная экранная форма подсистемы «Эйдос-VCF», а на рисунке 10 – экранная форма ее режима настроек:

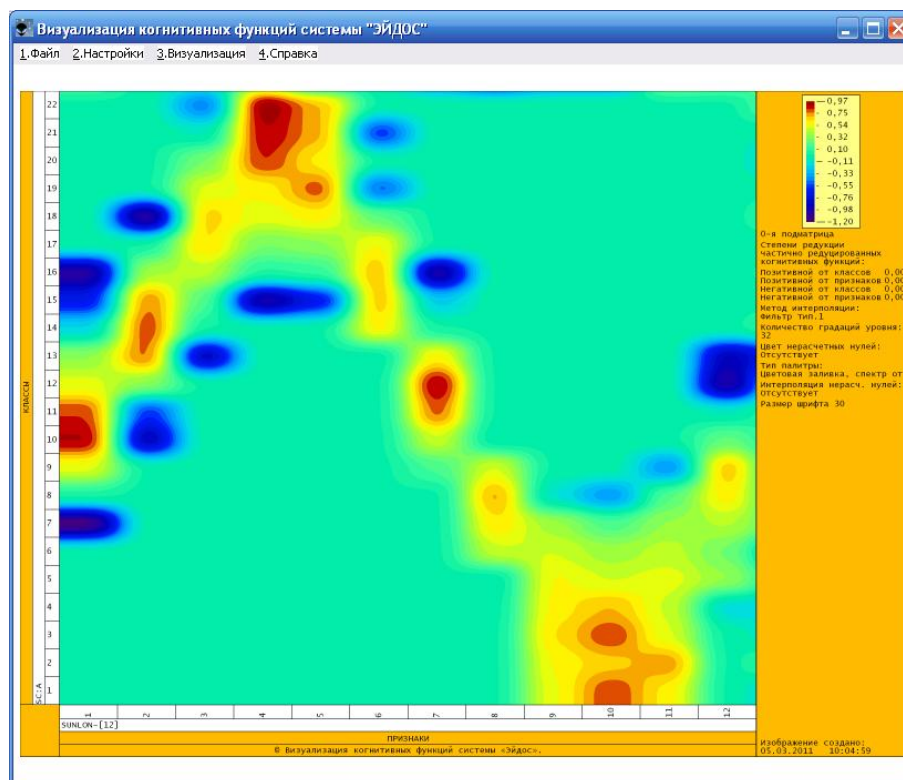


Рисунок 9. Главная экранная форма подсистемы «Эйдос-VCF»

¹³ <http://ru.wikipedia.org/wiki/Триангуляция%20Делоне>

¹⁴ <http://alglib.sources.ru/interpolation/spline3.php>

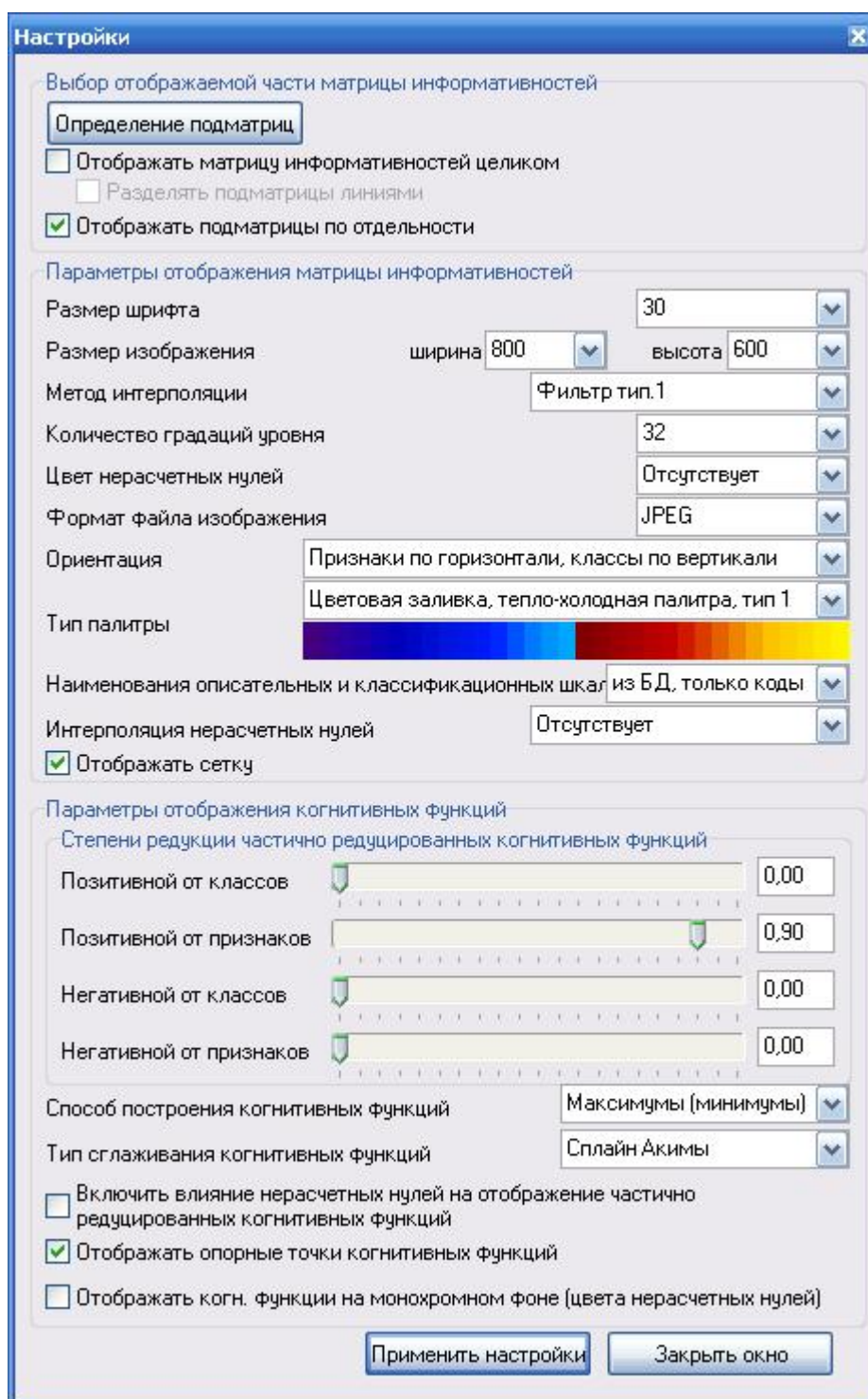


Рисунок 10. Экранная форма режима настроек подсистемы «Эйдос-VCF»

На рисунке 11 приведена экранная форма режима ввода-корректировки наименований классификационных и описательных шкал и градаций подсистемы «Эйдос-VCF», которая появляется при нажатии на клавишу «Определение подматриц»:

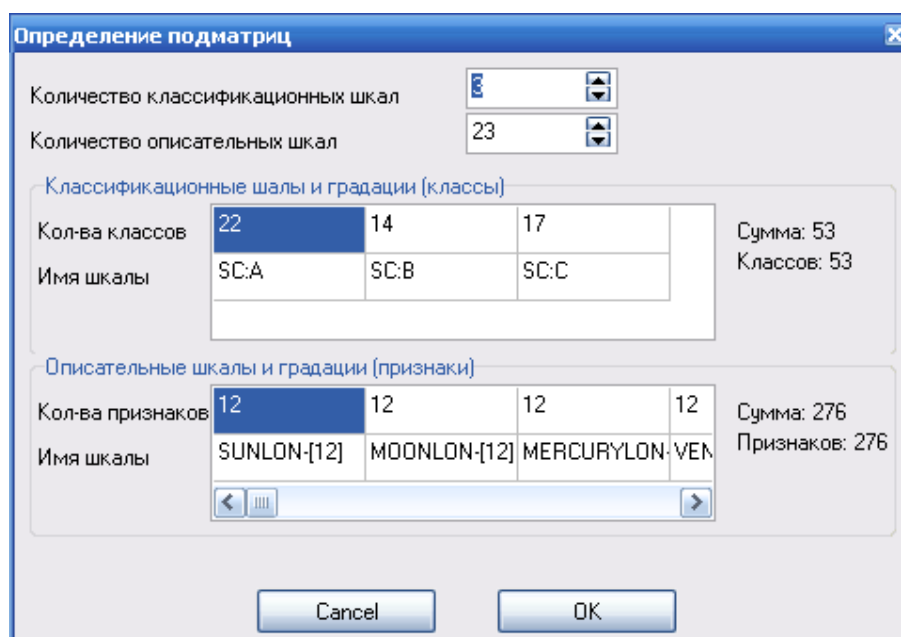
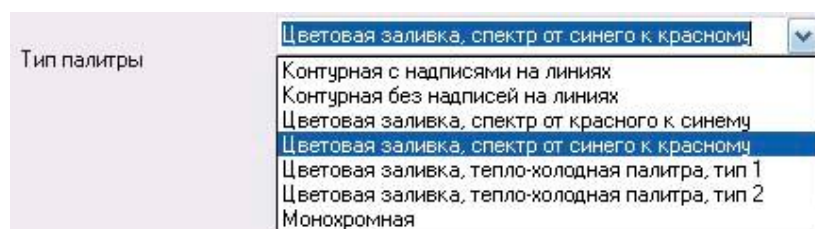


Рисунок 11. Экранная форма режима ввода-корректировки наименований классификационных и описательных шкал и градаций подсистемы «Эйдос-VCF»

В подсистеме «Эйдос-VCF» имеется множество настроек каждого параметра, показанного на окне настроек (рисунок 10) разворачивающимися списками, которые в совокупности позволяют получить самые различные варианты изображений когнитивных функций, из которых можно выбрать наиболее контрастно и удачно отображающие их смысл. Например, параметр: «Тип палитры» может принимать следующие варианты значений:



и при каждом из этих значений когнитивная функция будет выглядеть по-разному. На рисунке 12 приведены различные виды *одной и той же* когнитивной функции, получающиеся при выборе различных значений параметра «Тип палитры»:

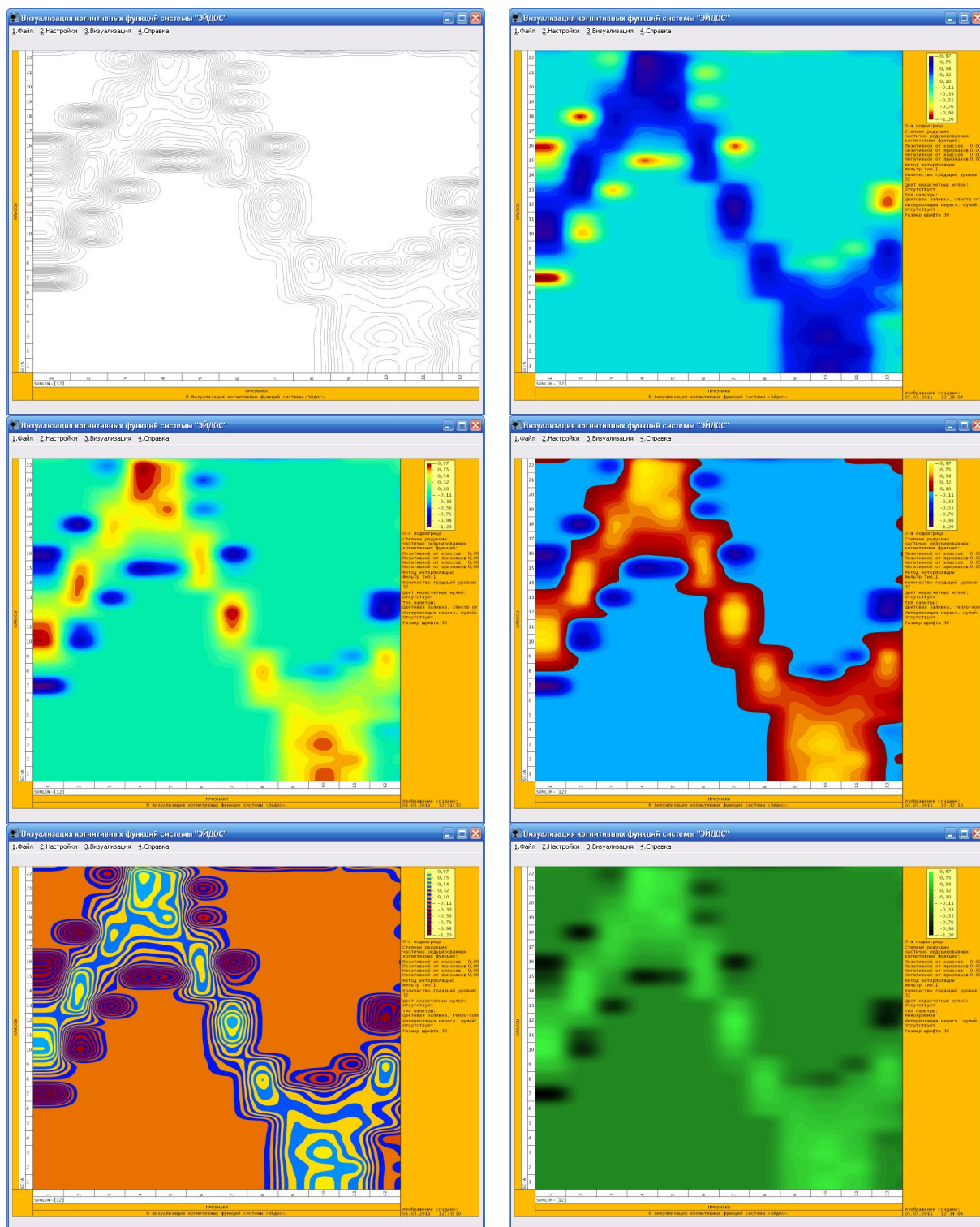


Рисунок 12. Изменение вида когнитивной функции при выборе различных значений параметра «Тип палитры» в подсистеме «Эйдос-VCF»

В работах [8-9, 11-23] и других рассмотрены многочисленные примеры приложения АСК анализа и аппарата когнитивных функций в задачах распознавания событий в поле центральных сил. К числу таких событий относятся, в частности, ряды событий, происходящих на нашей планете, которая движется в Солнечной системе в гравитационном поле Солнца.

В качестве примера укажем ряды сейсмических событий, исследованные в работах [8, 11-12, 14], движение полюса Земли [13, 21-23], ряды социально-экономических событий [15-20]. Ниже дан краткий обзор перечисленных работ с акцентом на применение в них метода визуализации когнитивных функций.

Задача о распознавании категорий событий в поле центральных сил

Рассмотрим задачу распознавания категорий по астрономическим данным [8-9, 11-20]. Имеется множество событий A , которому ставится в соответствие множество категорий C_i . Событиями можно считать, например, измерение координат полюса Земли, а категориями – значение координат, лежащее в определенном интервале. Каждое такое событие характеризуется моментом времени и географическими координатами места его происхождения (которые в данной задаче фиксированы). По этим данным можно построить матрицу, содержащую координаты небесных тел, например астрономические углы долготы, широты и расстояния. Будем считать, что заданы частотные распределения N_i – число событий, имеющих отношение к данной категории C_i .

Определим число случаев реализации данной категории, которое приходится на заданный интервал изменения астрономических параметров, имеем в дискретном случае:

$$N_{ij}(x_j, k) = N_i w(\tilde{x}_j, k) \Delta x, \quad x_j < \tilde{x}_j < x_j + \Delta x$$

$$1 \leq i \leq n, \quad 1 \leq j \leq 2m, \quad k = 1, \dots, k_0 \tag{1}$$

Здесь w – плотность распределения событий вдоль нормированной координаты. Нормированная переменная определяется через угловую и радиальную координаты следующим образом:

$$x_{jk} = \begin{cases} J_j(k) / 2p, & 1 \leq j \leq m \\ \frac{r_{\max}(k) - r(k)}{r_{\max}(k) - r_{\min}(k)}, & m + 1 \leq j \leq 2m \end{cases} \tag{2}$$

где r_{\min}, r_{\max} – минимальное и максимальное удаление планеты от центра масс системы, k_0 – число небесных тел, используемых в задаче.

Определим матрицу информативностей согласно [1]

$$I_{ijk} = \log_2 \frac{N_{ij} / \sum_j N_{ij}}{\sum_i N_{ij} / \sum_{i,j} N_{ij}}, \quad N_{ij}(x_{jk}) \neq 0$$

$$I_{ijk} = 0, \quad N_{ij}(x_{jk}) = 0, \tag{3}$$

$$dI_{jk} = \sqrt{\frac{1}{n} \sum_i \left(I_{ijk} - \frac{1}{n} \sum_i I_{ijk} \right)^2} \quad (3)$$

$$1 \leq i \leq n, \quad 1 \leq j \leq 2m, \quad 1 \leq k \leq k_0$$

Первая величина (3) называется информативность признака, а вторая величина является стандартным отклонением информативности или интегральной информативностью (ИИ).

Каждой категории можно сопоставить вектор информативности астрономических параметров размерности $2mk_0$, составленный из элементов матрицы информативности, путем последовательной записи столбцов, соответствующих нормированной координате, в один столбец, т.е.

$$c_{is} = I_{ijk} \Big|_{jk=s}, \quad 1 \leq s \leq 2mk_0 \quad (4)$$

С другой стороны, процесс идентификации, распознавания и прогнозирования может рассматриваться как разложение вектора распознаваемого объекта в ряд по векторам категорий (классов распознавания) [1]. Этот вектор, состоящий из единиц и нулей, можно определить по координатам небесных тел, соответствующих дате и месту происхождения события l в виде

$$a_{ls} = \begin{cases} 1, & (j-1)\Delta x \leq x_{jk}(l) \leq j\Delta x, \quad jk = s \\ 0, & 1 \leq s \leq 2mk_0 \end{cases} \quad (5)$$

Таким образом, если нормированная координата небесного тела из данных по объекту исследуемой выборки попадает в заданный интервал, элементу вектора придается значение 1, а во всех остальных случаях – значение 0. Перечисление координат осуществляется последовательно, для каждого небесного тела.

В случае, когда система векторов (4) является полной, можно точно любой вектор (5) представить в виде линейной комбинации векторов системы (4). Коэффициенты этого разложения будут соответствовать уровню сходства данного события с данной категорией. В случае неполной системы векторов (4) точная процедура заменяется распознаванием или разложением в ряд с некоторой погрешностью. При этом уровень сходства данных события с той или иной категорией можно определить по величине скалярного произведения вектора (4) на вектор (5), т.е. в координатной форме:

$$K_{il} = \frac{1}{|a_l||c_i|} \sum_{s=1}^{2mk_0} a_{ls}(A)c_{is} \quad (6)$$

Отметим, что возможны четыре исхода, при которых можно истинно или ложно отнести или не отнести данное событие к данной категории. Для учета этих исходов распознавание категорий в системе искусственного интеллекта «Эйдос-астра» [4] осуществляется по параметру сходства, который определяется следующим образом [16]:

$$S_i = \frac{1}{N} \sum_{l=1}^N (BT_{il} + T_{il} - BF_{il} - F_{il}) \cdot 100 \% \quad (7)$$

S_i – достоверность идентификации « i -й» категории;

N – количество событий в распознаваемой выборке;

BT_{il} – уровень сходства « l -го» события с « i -й» категорией, к которой он был правильно отнесен системой;

T_{il} – уровень сходства « l -го» события с « i -й» категорией, к которой он был правильно не отнесен системой;

BF_{il} – уровень сходства « l -го» события с « i -й» категорией, к которой он был ошибочно отнесен системой;

F_{il} – уровень сходства « l -го» события с « i -й» категорией, к которой он был ошибочно не отнесен системой.

При таком определении параметр сходства изменяется в пределах от -100% до 100%, как обычный коэффициент корреляции в статистике. При этом ошибки 1-го и 2-го рода (ошибки ложной идентификации и ложной неидентификации) приводят к уменьшению параметра сходства. Очевидно, что параметр сходства должен удовлетворять критерию простой проверки

$$S_i(N_i = 1) = 100 \%$$

Было показано, что процедура распознавания по параметру сходства (7), реализованная в системе искусственного интеллекта «Эйдос-астра» [4], является устойчивой как относительно объема выборки, так и относительно числа ячеек модели. Математическое обоснование этой процедуры дано в монографии [1].

На рис. 13 представлен фрагмент матрицы информативностей, демонстрирующий зависимость координат (X,Y) и угловой скорости (X1, Y1) движения полюса Земли – всего 244 категории, перечисленные в таблице 2, от долготы Солнца в 1963-2006 годов по данным Earth Orientation Centre (<http://hpiers.obspm.fr/eop-pc/>). Как следует из данных, приведенных на рис. 13, координаты полюса в зависимости от долготы Солнца образуют *жгуты*, которые формируются из отдельных годовых движений на протяжении многих лет. Отметим, что категории угловой скорости не образуют достаточно четких нитевидных структур, подобных тем, что образуют категории координат. Согласно существующим представлениям, такое поведение угловой скорости полюса в зависимости от долготы Солнца обусловлено наличием случайной составляющей, связанной с движением атмосферы и океана относительно земной коры.

Таблица 2 – ЕЖЕДНЕВНАЯ ЧАСТОТА ПОЯВЛЕНИЯ КАТЕГОРИЙ КООРДИНАТ И УГЛОВОЙ СКОРОСТИ ДВИЖЕНИЯ ПОЛЮСА ЗЕМЛИ В 1963-2006 ГОДАХ

Категория X	ABS	Категория Y	ABS	Категория X1	ABS	Категория Y1	ABS
A1-X=-0,29609	11	B1-Y=-0,01292	30	A1-X1=-0,006	1	B2-Y1=-0,00554	1
A2-X=-0,28609	8	B2-Y=-0,00292	55	A2-X1=-0,0058	2	B4-Y1=-0,00514	6
A3-X=-0,27609	8	B3-Y=0,00708	114	A3-X1=-0,0056	1	B5-Y1=-0,00494	4
A4-X=-0,26609	7	B4-Y=0,01708	139	A4-X1=-0,0054	5	B6-Y1=-0,00474	20
A5-X=-0,25609	17	B5-Y=0,02708	116	A5-X1=-0,0052	7	B7-Y1=-0,00454	42
A6-X=-0,24609	70	B6-Y=0,03708	95	A6-X1=-0,005	11	B8-Y1=-0,00434	46
A7-X=-0,23609	99	B7-Y=0,04708	104	A7-X1=-0,0048	23	B9-Y1=-0,00414	93
A8-X=-0,22609	140	B8-Y=0,05708	123	A8-X1=-0,0046	32	B10-Y1=-0,00394	130
A9-X=-0,21609	125	B9-Y=0,06708	217	A9-X1=-0,0044	47	B11-Y1=-0,00374	137
A10-X=-0,20609	194	B10-Y=0,07708	248	A10-X1=-0,0042	67	B12-Y1=-0,00354	169
A11-X=-0,19609	199	B11-Y=0,08708	253	A11-X1=-0,004	136	B13-Y1=-0,00334	204
A12-X=-0,18609	188	B12-Y=0,09708	207	A12-X1=-0,0038	150	B14-Y1=-0,00314	258
A13-X=-0,17609	173	B13-Y=0,10708	247	A13-X1=-0,0036	202	B15-Y1=-0,00294	345
A14-X=-0,16609	238	B14-Y=0,11708	274	A14-X1=-0,0034	265	B16-Y1=-0,00274	315
A15-X=-0,15609	378	B15-Y=0,12708	256	A15-X1=-0,0032	309	B17-Y1=-0,00254	390
A16-X=-0,14609	269	B16-Y=0,13708	314	A16-X1=-0,003	354	B18-Y1=-0,00234	434
A17-X=-0,13609	272	B17-Y=0,14708	317	A17-X1=-0,0028	356	B19-Y1=-0,00214	433
A18-X=-0,12609	269	B18-Y=0,15708	346	A18-X1=-0,0026	383	B20-Y1=-0,00194	446
A19-X=-0,11609	340	B19-Y=0,16708	375	A19-X1=-0,0024	335	B21-Y1=-0,00174	440
A20-X=-0,10609	354	B20-Y=0,17708	451	A20-X1=-0,0022	434	B22-Y1=-0,00154	449
A21-X=-0,09609	271	B21-Y=0,18708	427	A21-X1=-0,002	445	B23-Y1=-0,00134	473
A22-X=-0,08609	299	B22-Y=0,19708	432	A22-X1=-0,0018	412	B24-Y1=-0,00114	500
A23-X=-0,07609	302	B23-Y=0,20708	422	A23-X1=-0,0016	398	B25-Y1=-0,00094	586
A24-X=-0,06609	342	B24-Y=0,21708	341	A24-X1=-0,0014	422	B26-Y1=-0,00074	542
A25-X=-0,05609	385	B25-Y=0,22708	372	A25-X1=-0,0012	449	B27-Y1=-0,00054	581
A26-X=-0,04609	379	B26-Y=0,23708	478	A26-X1=-0,001	403	B28-Y1=-0,00034	515
A27-X=-0,03609	515	B27-Y=0,24708	417	A27-X1=-0,0008	510	B29-Y1=-0,00014	587
A28-X=-0,02609	406	B28-Y=0,25708	374	A28-X1=-0,0006	595	B30-Y1=0,00006	521
A29-X=-0,01609	460	B29-Y=0,26708	340	A29-X1=-0,0004	515	B31-Y1=0,00026	614
A30-X=-0,00609	421	B30-Y=0,27708	332	A30-X1=-0,0002	574	B32-Y1=0,00046	612
A31-X=0,00391	441	B31-Y=0,28708	333	A31-X1=0	656	B33-Y1=0,00066	575
A32-X=0,01391	336	B32-Y=0,29708	338	A32-X1=0,0002	554	B34-Y1=0,00086	468
A33-X=0,02391	347	B33-Y=0,30708	356	A33-X1=0,0004	575	B35-Y1=0,00106	427
A34-X=0,03391	347	B34-Y=0,31708	400	A34-X1=0,0006	546	B36-Y1=0,00126	403
A35-X=0,04391	369	B35-Y=0,32708	427	A35-X1=0,0008	458	B37-Y1=0,00146	503
A36-X=0,05391	419	B36-Y=0,33708	369	A36-X1=0,001	473	B38-Y1=0,00166	554
A37-X=0,06391	469	B37-Y=0,34708	370	A37-X1=0,0012	503	B39-Y1=0,00186	456
A38-X=0,07391	382	B38-Y=0,35708	341	A38-X1=0,0014	469	B40-Y1=0,00206	438
A39-X=0,08391	414	B39-Y=0,36708	379	A39-X1=0,0016	485	B41-Y1=0,00226	427
A40-X=0,09391	402	B40-Y=0,37708	450	A40-X1=0,0018	428	B42-Y1=0,00246	328
A41-X=0,10391	410	B41-Y=0,38708	372	A41-X1=0,002	338	B43-Y1=0,00266	271
A42-X=0,11391	350	B42-Y=0,39708	351	A42-X1=0,0022	343	B44-Y1=0,00286	306
A43-X=0,12391	371	B43-Y=0,40708	289	A43-X1=0,0024	344	B45-Y1=0,00306	246
A44-X=0,13391	416	B44-Y=0,41708	301	A44-X1=0,0026	353	B46-Y1=0,00326	197
A45-X=0,14391	267	B45-Y=0,42708	335	A45-X1=0,0028	257	B47-Y1=0,00346	170
A46-X=0,15391	280	B46-Y=0,43708	207	A46-X1=0,003	329	B48-Y1=0,00366	132
A47-X=0,16391	284	B47-Y=0,44708	214	A47-X1=0,0032	239	B49-Y1=0,00386	99
A48-X=0,17391	258	B48-Y=0,45708	241	A48-X1=0,0034	192	B50-Y1=0,00406	60
A49-X=0,18391	257	B49-Y=0,46708	229	A49-X1=0,0036	175	B51-Y1=0,00426	37
A50-X=0,19391	250	B50-Y=0,47708	282	A50-X1=0,0038	132	B52-Y1=0,00446	15
A51-X=0,20391	300	B51-Y=0,48708	181	A51-X1=0,004	106	B53-Y1=0,00466	5
A52-X=0,21391	256	B52-Y=0,49708	131	A52-X1=0,0042	83	B54-Y1=0,00486	8
A53-X=0,22391	294	B53-Y=0,50708	125	A53-X1=0,0044	65	B55-Y1=0,00506	3
A54-X=0,23391	181	B54-Y=0,51708	115	A54-X1=0,0046	41	B56-Y1=0,00526	4
A55-X=0,24391	170	B55-Y=0,52708	122	A55-X1=0,0048	15	B59-Y1=0,00586	2
A56-X=0,25391	190	B56-Y=0,53708	132	A56-X1=0,005	11	B62-Y1=0,00646	1
A57-X=0,26391	79	B57-Y=0,54708	147	A57-X1=0,0052	9	B64-Y1=0,00686	1
A58-X=0,27391	92	B58-Y=0,55708	103	A58-X1=0,0054	5	B65-Y1=0,00706	1
A59-X=0,28391	78	B59-Y=0,56708	68	A59-X1=0,0056	1		
A60-X=0,29391	56	B60-Y=0,57708	14	A60-X1=0,0058	2		
A61-X=0,30391	72						
A62-X=0,31391	23						

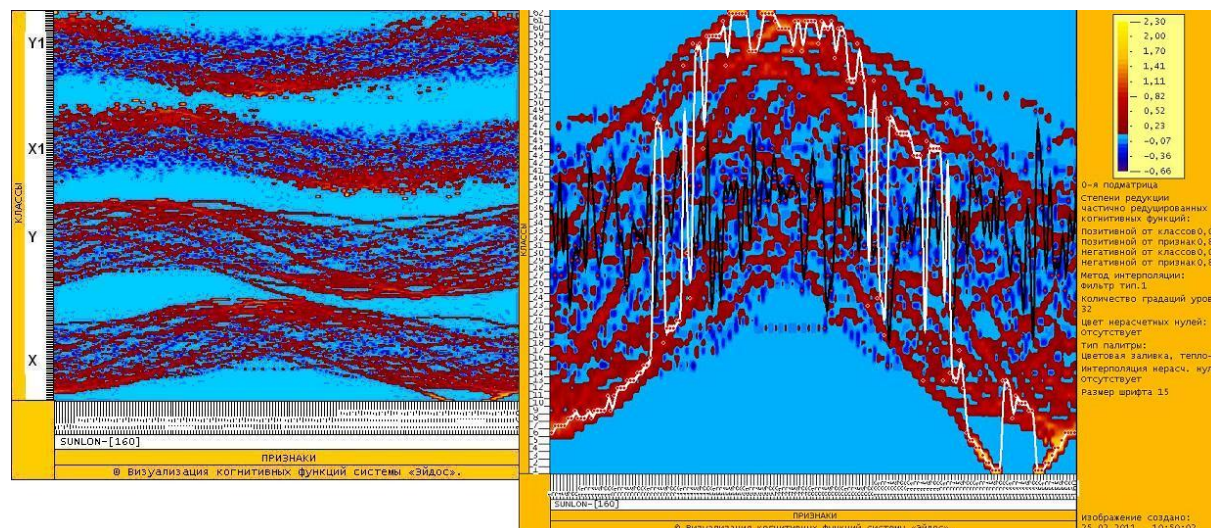


Рисунок 13. Зависимость категорий координат и угловой скорости полюса от долготы Солнца в модели M160 [13] (слева) и прямые позитивные (белая линия) и негативные (черная линия) когнитивные функции координаты X (смещение вдоль меридиана Гринвич).

Прямая когнитивная функция, представленная на правом рис. 13, белой/черной линией (позитивная/негативная функция) характеризует зависимость координаты X смещения полюса вдоль меридиана Гринвич от долготы Солнца. Позитивная функция показывает, что система детерминации способствует увеличению амплитуды колебаний, а негативная функция, напротив, указывает на стабилизацию относительно среднего значения.

Заметим, что просто графическое отображение матрицы информативностей – это нередуцированная когнитивная функция (степень редукции 0). Кривая, соединяющая точки с максимальной информативностью для смежных градаций описательной шкалы – это полностью редуцированная когнитивная функция (степень редукции 1). Она показывает какое состояние объекта наиболее вероятно для каждого значения аргумента, т.е. об осуществлении какого состояния в каждом значении аргумента наибольшее количество информации.

В полностью редуцированной когнитивной функции из самого вида ее графика не видно, в каком значении аргумента больше информации, а в каком меньше в различных значениях аргумента. Эта информация есть в нередуцированной когнитивной функции и отображается она в виде цвета. Но если отображать ее на графике, то кажется удобным использовать для этого полосу около графика полностью редуцированной когнитивной функции: чем эта полоса уже, тем более определенным является ее значение, и чем шире – тем менее определенным. Это соответствует понятию информации как количественной меры снятия неопределенности: много

информации в значении аргумента о значении функции – неопределенность значения функции мала и полоса уже, мало информации – неопределенность больше и полоса шире. Это напоминает и понятие доверительного интервала.

Поскольку есть полностью нередуцированные и полностью редуцированные функции, то вводится и понятие когнитивной функции с промежуточной степенью редукции, т.е. частично редуцированные когнитивные функции. Существует множество функций с различной степенью редукции. Степень редукции задается в программе InfVisual в процессе построения изображения.

На рис. 14 представлен каталог данных визуализации матрицы информативностей в задаче распознавания сейсмических событий по астрономическим данным [8-9, 14], а также с учетом влияния магнитного поля и движения полюса Земли [11-13].

Исследуемая база данных сейсмических событий была сформирована на основе базы данных Международного сейсмологического центра – ISC (см. <http://www.isc.ac.uk/>), содержащей 20489816 записей регистрации различными сейсмостанциями событий землетрясений, произошедших на нашей планете в период с 1 января 1961 года по 31 декабря 2006 г. Из исходной базы было образовано несколько различных БД для исследования влияния астрономических параметров на магнитуду и глубину гипофокуса, на ежедневное число землетрясений, а также на средние параметры сейсмической активности. В работах [11-13] исследована совокупность 128320 событий землетрясений с магнитудой $mb \geq 4$, произошедших на нашей планете в период с 9 февраля 1963 года по 31 декабря 2006 г (всего 16032 дня).

В исходной БД сейсмические события характеризуются магнитудой mb , которой можно сопоставить категорию магнитуды – таблица 3. Поскольку события с одной и той же магнитудой могут повторяться в один день, каждому значению магнитуды сопоставляется несколько типов категорий, а именно:

А – событие с магнитудой mb повторяется один раз;

В – событие с магнитудой mb повторяется два раза;

С – событие с магнитудой mb повторяется три раза.

Кроме того, можно рассмотреть случай, когда, например, категория А усекается, путем отбрасывания некоторых событий. Таким образом, были образованы категории А66, В59 и С53.

Таблица 3 – Частота повторения категорий сейсмической активности

Категория	ABS	Категория	ABS	Категория	ABS
A40-Mb=4,0	1362	B40-Mb=4,0	446	C40-Mb=4,0	130
A41-Mb=4,1	1580	B41-Mb=4,1	660	C41-Mb=4,1	259
A42-Mb=4,2	1796	B42-Mb=4,2	835	C42-Mb=4,2	362
A43-Mb=4,3	2224	B43-Mb=4,3	955	C43-Mb=4,3	436
A44-Mb=4,4	2744	B44-Mb=4,4	1099	C44-Mb=4,4	482
A45-Mb=4,5	3358	B45-Mb=4,5	1223	C45-Mb=4,5	468
A46-Mb=4,6	4119	B46-Mb=4,6	1455	C46-Mb=4,6	515
A47-Mb=4,7	4768	B47-Mb=4,7	1612	C47-Mb=4,7	501
A48-Mb=4,8	4954	B48-Mb=4,8	1817	C48-Mb=4,8	450
A49-Mb=4,9	5008	B49-Mb=4,9	1636	C49-Mb=4,9	447
A50-Mb=5	4904	B50-Mb=5	1428	C50-Mb=5	356
A51-Mb=5,1	4582	B51-Mb=5,1	1206	C51-Mb=5,1	293
A52-Mb=5,2	4134	B52-Mb=5,2	936	C52-Mb=5,2	166
A53-Mb=5,3	3563	B53-Mb=5,3	617	C53-Mb=5,3-6,0	105
A54-Mb=5,4	3010	B54-Mb=5,4	422		
A55-Mb=5,5	2367	B55-Mb=5,5	261		
A56-Mb=5,6	1940	B56-Mb=5,6	180		
A57-Mb=5,7	1460	B57-Mb=5,7	93		
A58-Mb=5,8	1179	B58-Mb=5,8	73		
A59-Mb=5,9	864	B59-Mb=5,9-6,4	69		
A60-Mb=6,0	656				
A61-Mb=6,1	453				
A62-Mb=6,2	319				
A63-Mb=6,3	202				
A64-Mb=6,4	137				
A65-Mb=6,5	87				
A66-Mb=6,6-7,0	68				

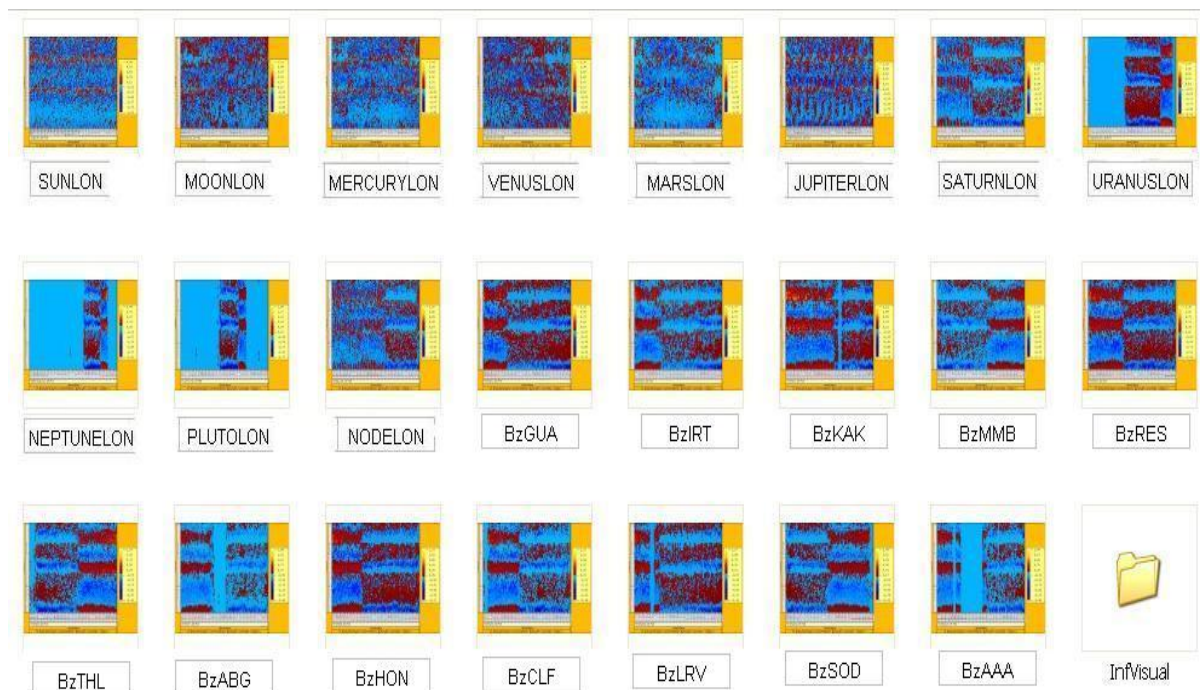


Рисунок 14. Зависимость категорий сейсмических событий А,В,С (в каждом рисунке категории отложены по вертикали снизу вверх) от долготы небесных тел и от вертикальной компоненты индукции магнитного поля на 12 станциях

Анализируя данные, приведенные на рис. 5, можно сделать вывод о том, что влияние долготы Сатурна, Урана, Нептуна, Плутона и Северного узла Луны на сейсмические события на Земле *аналогично* влиянию вертикальной компоненты индукции магнитного поля Земли. Эта аналогия, впервые обнаруженная в работе [12], послужила основой для создания моделей влияния небесных тел на геомагнитное поле и на движение полюса Земли [12-13, 21-22].

На рис. 15 представлен каталог данных визуализации матрицы информативностей в задаче распознавания социальных категорий респондентов по астрономическим данным [15-17]. В этой задаче используется пространство данных (1) различной размерности от $23 \times 2 = 46$ до $23 \times 173 = 3979$, в котором распознается вектор (5), состоящий из 870 [15] социальных категорий. Выделяя среди этих категорий наиболее часто повторяющиеся, приходим к задаче о распознавании 37 категорий [16] (отложены по вертикали на рис. 15) – таблица 4, или только четырех [17].

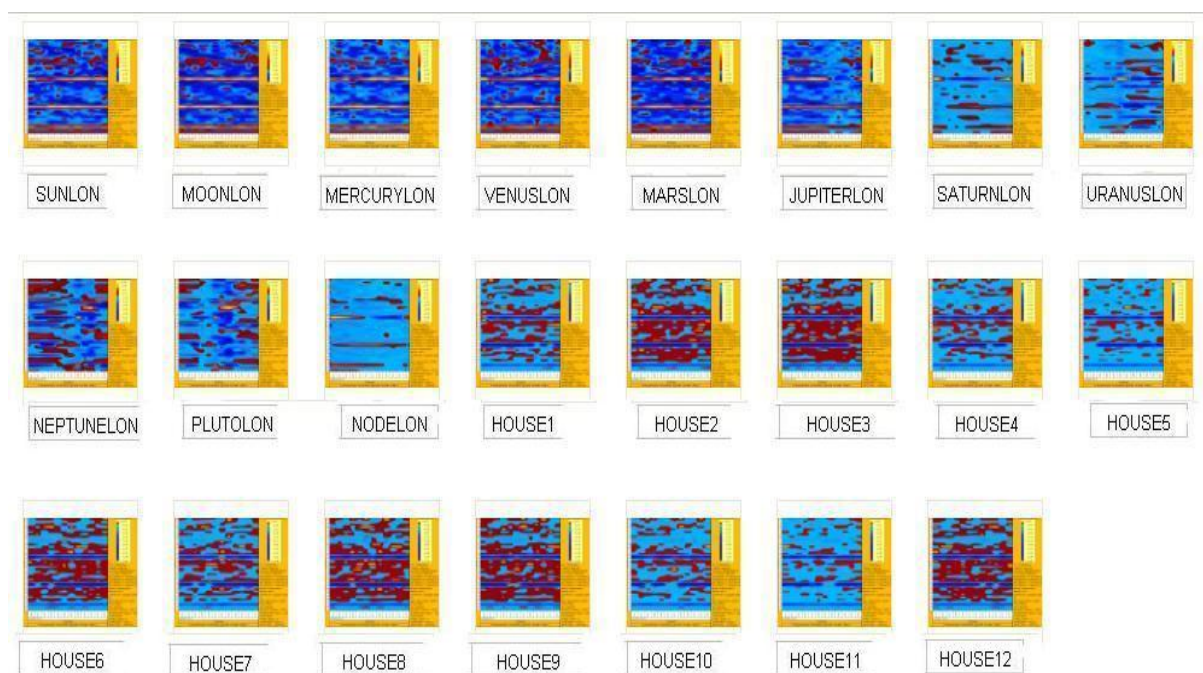


Рисунок 15. Зависимость частоты социальных категорий от угловых параметров небесных тел в модели M12 по данным [15].

Из данных, приведенных на рис. 15, следует, что подматрица информативностей является индивидуальной для каждого астрономического параметра. Это позволяет осуществить процедуру распознавания по параметру сходства (7) с относительно высокой вероятностью около 68,75% [16], что легло в основу метода прогнозирования социальных категорий - астросоциотипологии [23].

Таблица 4 – Частота повторения социальных категорий в БД [16]

NAME	ABS	NAME	ABS
SC:M-	13892	SC:B45-Famous:Greatest hits	1833
SC:Ж-	5226	SC:A29-Parenting	1812
SC:A53-Sports	4608	SC:B173-Sports:Football	1627
SC:A1-Book Collection	4562	SC:B97-Occult Fields:Astrologer	1509
SC:A15-Famous	3456	SC:B21-Relationship:Number of marriages	1461
SC:A42-Medical	3037	SC:B2-Book Collection:Profiles Of Women	1425
SC:A323-Sexuality	2736	SC:A92-Birth	1361
SC:A5-Entertainment	2657	SC:B14-Entertainment:Actor/ Actress	1296
SC:A9-Relationship	2514	SC:?-	1259
SC:A40-Occult Fields	2430	SC:B49-Book Collection:American Book	1204
SC:B111-Sports:Basketball	2403	SC:B26-Personality:Body	1206
SC:B329-Sexuality:Sexual perversions	2415	SC:B189-Medical:Illness	1236
SC:A55-Art	2288	SC:B6-Entertainment:Music	1124
SC:A19-Writers	2281	SC:A99-Financial	1102
SC:A129-Death	2263	SC:B48-Famous:Top 5% of Profession	1104
SC:A25-Personality	2140	SC:A38-Politics	1073
SC:A68-Childhood	2069	SC:A23-Psychological	1041
SC:A31-Business	1858	SC:A108-Education	1027
SC:C330-Sexuality:Sexual perversions: Homosexual m	1858		

Путем формальной замены в задаче [16] социальных категорий на экономические, был развит метод прогнозирования курсов валют по астрономическим данным [18-19]. В таблице 5 приведена ежедневная частота повторения категорий повышения (1) и понижения (2) курсов валют в 2000-2009 гг, вычисленная по данным FOREX. На рис. 16 представлен каталог данных визуализации матрицы информативностей категорий валют из таблицы 5 в зависимости от долготы (LON) и расстояния (DIST) до небесных тел в модели M160.

Таблица 5 – Частота повторения категорий повышения (1) и понижения (2) курсов валют в 2000-2009 годах [19].

NAME	ABS	NAME	ABS
A72-GBP/USD2	3624	A52-NZD/USD2	3444
A71-GBP/USD1	3582	A51-NZD/USD1	3762
A82-EUR/GBP2	3675	A112-EUR/CHF2	3558
A81-EUR/GBP1	3531	A111-EUR/CHF1	3648
A92-USD/CHF2	3672	A32-EUR/USD2	3549
A91-USD/CHF1	3534	A31-EUR/USD1	3657
A62-GBP/CHF2	3534	A22-USD/CAD2	3705
A61-GBP/CHF1	3672	A21-USD/CAD1	3501
A12-AUD/USD2	3462	A42-USD/JPY2	3627
A11-AUD/USD1	3744	A41-USD/JPY1	3579
A102-GBP/JPY2	3516	A121-EUR/JPY1	3768
A101-GBP/JPY1	3708	A122-EUR/JPY2	3438

Из этих данных следует, что подматрицы информативностей категорий курсов валют изменяются индивидуально в зависимости от астрономических параметров небесных тел, что позволило создать метод прогнозирования курсов валют [18-19].

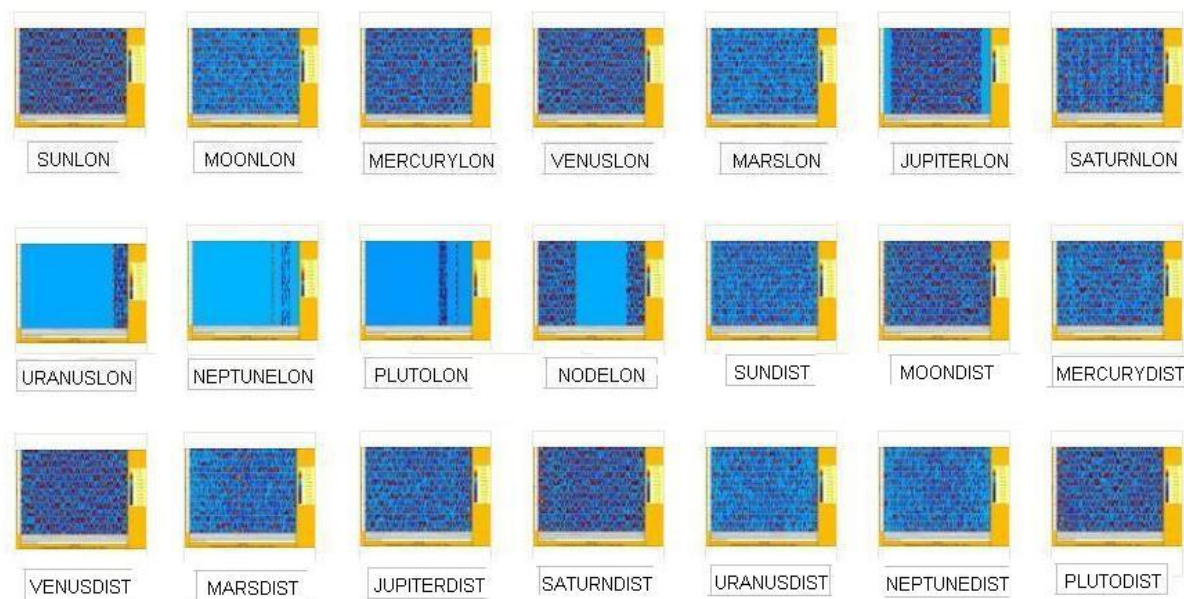


Рисунок 16. Зависимость частоты категорий повышения/понижения курсов валют от астрономических параметров небесных тел в модели M160 по данным [19].

Полученные результаты позволяют предположить, что существует *принцип аналогии* [9], объединяющий глобальные социальные [15-17], экономические [18-20] и природные процессы [8, 11-22], происходящие на нашей планете, сформулировать теорему о распознавании событий в поле центральных сил [18, 23-24], а также создать метод исследования указанных процессов на основе АСК-анализа и аппарата когнитивных функций.

Развитый в автоматизированном системно-когнитивном анализе аппарат выявления и визуализации причинно-следственных зависимостей в форме когнитивных функций позволяет очень наглядно буквально увидеть такие объективно существующие явления и закономерности, о существовании которых еще недавно в науке вообще не было известно и которые весьма проблематично обнаружить другими методами, в том числе аналитическими.

Здесь необходимо особо отметить, что данный подход *предлагает* такое распределение функций по выявлению и анализу причинно-следственных зависимостей между человеком и интеллектуальной системой, при котором на каждую из сторон возлагаются именно те функции, которые в настоящее время (при современном уровне развития технологии и сознания человека) ею выполняются лучше, чем другой стороной. В частности, в настоящее время есть смысл использовать превосходные возможности человека по выявлению закономерностей в изображениях. Но для этого необходимо соответствующим образом подготовить и представить ему эти изображения, что осознанно и целенаправленно реализовано авторами в методе визуализации когнитивных функций.

Это позволяет обоснованно говорить о том, что автоматизированный системно-когнитивный анализ и его программный инструментарий – сис-

тема «Эйдос-астра» и базовая система «Эйдос» представляют собой новый инструмент исследования в астрономии и геофизике, своего рода «математический телескоп», открывающий качественно новые, ранее недоступные возможности исследования. История науки наглядно демонстрирует, что появление более совершенных инструментов исследования, обеспечивающих качественно новые возможности исследования, ранее всегда приводило к возникновению новых направлений в науке. Так создание микроскопа позволило открыть целый мир микроорганизмов и привело к возникновению микробиологии, создание оптического телескопа позволило Галилео Галилею сразу же открыть спутники Юпитера и привело к созданию оптической астрономии, создание радиотелескопа привело к возникновению радиоастрономии, и.т.д. Авторы считают, что применение систем искусственного интеллекта для анализа баз данных, содержащих информацию об огромном количестве событий на Земле в различных глобальных системах, позволяет выявить в этих данных влияние небесных тел Солнечной системы на эти события и, *позволяет открыть существование новых, ранее неизвестных объективно существующих явлений и закономерностей*. По сути это означает, что применение технологий искусственного интеллекта для исследования влияния небесных тел Солнечной системы на глобальные геосистемы: ноосферу, биосферу, атмосферу, магнитосферу, геосферу и другие, представляет собой новое перспективное направление исследований и разработок в науке.

Выводы. Таким образом, *когнитивные функции являются адекватным математическим инструментом для формального представления причинно-следственных зависимостей*. Когнитивные функции представляют собой многозначные интервальные функции многих аргументов, в которых различные значения функции в различной степени соответствуют различным значениям аргументов, причем количественной мерой этого соответствия выступает знания, т.е. информация о причинно-следственных зависимостях в эмпирических данных, полезная для достижения целей. Многочисленные исследования [8-23] подтверждают, что предложенный авторами *метод и программный инструментарий визуализации когнитивных функций* позволяют наглядно увидеть такие причинно-следственные закономерности предметной области, которые другими методами выявить и описать весьма проблематично.

Библиографический список

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. – 605с¹⁵.
2. Луценко Е.В. 30 лет системе «Эйдос» – одной из старейших отечественных универсальных систем искусственного интеллекта, широко применяемых и развивающихся и в настоящее

¹⁵ Для удобства читателей ряд работ из списка приведен на сайте автора: <http://lc.kubagro.ru/>

- время / Е.В. Луценко // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №10(54). – Шифр Информрегистра: 0420900012\0110. – Режим доступа: <http://ej.kubagro.ru/2009/10/pdf/04.pdf>
3. Луценко Е.В. Универсальная когнитивная аналитическая система "ЭЙДОС". Пат. № 2003610986 РФ. Заяв. № 2003610510 РФ. Оpubл. от 22.04.2003.
 4. Patent 2008610097¹⁶, Russia, System for Typification and Identification of the Social Status of Respondents Based on the Astronomical Data at the Time of Birth - "AIDOS-ASTRO" / E.V. Lutsenko, A.P. Trunev, V.N. Shashin; Application № 2007613722, January 9, 2008.
 5. Луценко Е.В. Системно-когнитивный анализ как развитие концепции смысла Шенка – Абельсона / Е.В. Луценко // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №03(5). – Режим доступа: <http://ej.kubagro.ru/2004/03/pdf/04.pdf>
 6. Луценко Е.В. АСК-анализ как метод выявления когнитивных функциональных зависимостей в многомерных зашумленных фрагментированных данных / Е.В. Луценко // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2005. – №03(11). – Режим доступа: <http://ej.kubagro.ru/2005/03/pdf/19.pdf>
 7. Луценко Е.В. Когнитивные функции как адекватный инструмент для формального представления причинно-следственных зависимостей / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2010. – №09(63). – Режим доступа: <http://ej.kubagro.ru/2010/09/pdf/01.pdf>
 8. Трунев А.П. Прогнозирование сейсмической активности и климата на основе семантических информационных моделей / А.П. Трунев, Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №09(53). – Шифр Информрегистра: 0420900012\0098. – Режим доступа: <http://ej.kubagro.ru/2009/09/pdf/09.pdf>
 9. Луценко Е.В. «Эйдос-астра» – интеллектуальная система научных исследований влияния космической среды на поведение глобальных геосистем / Е.В. Луценко, А.П. Трунев // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2010. – №07(61). – Режим доступа: <http://ej.kubagro.ru/2010/07/pdf/17.pdf>
 10. Шеляг М.М. Системно-когнитивный анализ структуры затрат и объемов производства продукции в АПК (по материалам Краснодарского края) / М.М. Шеляг // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2010. – №04(58). – Шифр Информрегистра: 0421000012\0064. – Режим доступа: <http://ej.kubagro.ru/2010/04/pdf/25.pdf>
 11. Трунев А.П. Системно-когнитивный анализ и прогнозирование сейсмической активности литосферы Земли, как глобальной активной геосистемы / А.П. Трунев, Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2010. – №01(55). – Шифр Информрегистра: 0421000012\0001. – Режим доступа: <http://ej.kubagro.ru/2010/01/pdf/22.pdf>
 12. Трунев А.П. Семантические информационные модели глобальной сейсмической активности при смещении географического и магнитного полюса / А.П. Трунев, Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2010. – №02(56). – Шифр Информрегистра: 0421000012\0023. – Режим доступа: <http://ej.kubagro.ru/2010/02/pdf/15.pdf>
 13. Трунев А.П. Автоматизированный системно-когнитивный анализ влияния тел Солнечной системы на движение полюса Земли и визуализация причинно-следственных зависимостей в виде когнитивных функций / А.П. Трунев, Е.В. Луценко, Д.К. Бандык // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №01(65). – Режим доступа: <http://ej.kubagro.ru/2011/01/pdf/20.pdf>

¹⁶ <http://lc.kubagro.ru/aidos/2008610097.jpg>

14. Трунев А.П. Прогнозирование землетрясений по астрономическим данным с использованием системы искусственного интеллекта / А.П. Трунев, Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №08(52). – Шифр Информрегистра: 0420900012\0086. – Режим доступа: <http://ej.kubagro.ru/2009/08/pdf/13.pdf>
15. Луценко Е.В. Типизация и идентификация респондентов в социологии по их астрономическим показателями на момент рождения. / Е.В. Луценко, А.П. Трунев, В.Н. Шашин // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2007. – №01(25). – Шифр Информрегистра: 0420700012\0014. – Режим доступа: <http://ej.kubagro.ru/2007/01/pdf/14.pdf>
16. Луценко Е.В. Астросоциотипология и спектральный анализ личности по астросоциотипам с применением семантических информационных мультимodelей / Е.В. Луценко, А.П. Трунев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №01(35). – Шифр Информрегистра: 0420800012\0002. – Режим доступа: <http://ej.kubagro.ru/2008/01/pdf/10.pdf>
17. Луценко Е.В. Artificial intelligence system for identification of social categories of natives based on astronomical parameters / Е.В. Луценко, А.П. Трунев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №03(37). – Режим доступа: <http://ej.kubagro.ru/2008/03/pdf/07.pdf>
18. Трунев А.П. Проблема распознавания событий в поле центральных сил и прогнозирование курсов валют / А.П. Трунев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №06(50). – Шифр Информрегистра: 0420900012\0057. – Режим доступа: <http://ej.kubagro.ru/2009/06/pdf/08.pdf>
19. Трунев А.П. Прогнозирование курсов валют по астрономическим данным с использованием системы искусственного интеллекта / А.П. Трунев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №07(51). – Шифр Информрегистра: 0420900012\0068. – Режим доступа: <http://ej.kubagro.ru/2009/07/pdf/15.pdf>
20. Трунев А.П. Корреляция фондового индекса s & p 500 с астрономическими и геофизическими параметрами (Системно-когнитивный анализ взаимосвязи ноосферы, литосферы, магнитосферы и космической среды) / А.П. Трунев, Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2010. – №03(57). – Шифр Информрегистра: 0421000012\0039. – Режим доступа: <http://ej.kubagro.ru/2010/03/pdf/13.pdf>
21. Трунев А.П. Моделирование электромагнитного и гравитационного влияния небесных тел солнечной системы на смещение географического полюса и магнитное поле Земли / А.П. Трунев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2010. – №07(61). – Шифр Информрегистра: 0421000012\0152. – Режим доступа: <http://ej.kubagro.ru/2010/07/pdf/16.pdf>
22. Трунев А.П. Моделирование влияния небесных тел на движение полюса Земли / А.П. Трунев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2010. – №10(64). – Режим доступа: <http://ej.kubagro.ru/2010/10/pdf/22.pdf>
23. Трунев А. П., Луценко Е. В. Астросоциотипология: Монография (научное издание). – Краснодар: КубГАУ, 2008, – 279 с.

24. Трунев А. П., Луценко Е. В. Системно-когнитивный анализ взаимосвязей между астрономическими и социальными событиями в астросоциотипологии /8-я Международная ФАМ конференция по финансово-актуарной математике и смежным вопросам, Красноярск, 2009.
25. Ричард Фейнман, «Характер физических законов». М., Наука, 1987 г., 160 с.
26. Патент РФ¹⁷: Луценко Е.В., Бандык Д.К. Подсистема визуализации когнитивных (каузальных) функций системы «Эйдос» (Подсистема «Эйдос-VCF»). Пат. № 2011612056 РФ. Заяв. № 2011610347 РФ 20.01.2011. Опубл. от 09.03.2011.
27. Патент РФ¹⁸: Луценко Е.В., Трунев А.П., Шашин В.Н., Бандык Д.К. Интеллектуальная система научных исследований влияния космической среды на глобальные геосистемы «Эйдос-астра» (ИСНИ «Эйдос-астра»). Пат. № 2011612054 РФ. Заяв. № 2011610345 РФ 20.01.2011. Опубл. от 09.03.2011.
28. 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). – Шифр Информрегистр: 0420900012\0111. – Режим доступа: <http://ej.kubagro.ru/2009/10/pdf/15.pdf>
29. Луценко Е.В. Обобщенный коэффициент эмерджентности Хартли как количественная мера синергетического эффекта объединения булеанов в системном обобщении теории множеств / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №02(66). – Режим доступа: <http://ej.kubagro.ru/2011/02/pdf/45.pdf>
30. Луценко Е.В. Системная теория информации и нелокальные интерпретируемые нейронные сети прямого счета / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2003. – №01(1). – Режим доступа: <http://ej.kubagro.ru/2003/01/pdf/11.pdf>
31. Луценко Е.В. Типовая методика и инструментарий когнитивной структуризации и формализации задач в СК-анализе / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №01(3). – Режим доступа: <http://ej.kubagro.ru/2004/01/pdf/16.pdf>
32. Луценко Е.В. Математический метод СК-анализа в свете идей интервальной бутстрепной робастной статистики объектов нечисловой природы / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №01(3). – Режим доступа: <http://ej.kubagro.ru/2004/01/pdf/13.pdf>

¹⁷ <http://lc.kubagro.ru/aidos/2011612056.jpg>

¹⁸ <http://lc.kubagro.ru/aidos/2011612054.jpg>