

УДК 004.8

08.00.13 - Математические и инструментальные методы экономики (экономические науки)

ПРИМЕНЕНИЕ АВТОМАТИЗИРОВАННОГО СИСТЕМНО-КОГНИТИВНОГО АНАЛИЗА БАНКОВСКИХ БАЗ ДАННЫХ ПО ОПЕРАЦИЯМ С КРЕДИТНЫМИ КАРТАМИ ДЛЯ КОЛИЧЕСТВЕННОЙ ОЦЕНКИ РИСКА МОШЕННИЧЕСТВА

Луценко Евгений Вениаминович
д.э.н., к.т.н., профессор
Web of Science ResearcherID S-8667-2018
Scopus Author ID: 57188763047
РИНЦ id=123162, SPIN-code: 9523-7101
prof.lutsenko@gmail.com <http://lc.kubagro.ru>
<https://www.researchgate.net/profile/Eugene-Lutsenko>
Кубанский Государственный Аграрный университет имени И.Т.Трубилина, Краснодар, Россия

В данной работе ставится, рассматривается и решается актуальная задача достоверного количественного прогнозирования рисков мошенничества по операциям с кредитными картами. Для решения задачи используются реальные данные с портала Kaggle. Применяются автоматизированный системно-когнитивный анализ (АСК-анализ) и его программный инструментарий, в качестве которого настоящее время используется интеллектуальная система «Эйдос». Рассматриваются объект, предмет, проблема, гипотеза, цель, метод, задачи исследования (проблематика исследования, краткое обоснование выбора метода исследования, задачи исследования). Приводится решение следующих задач исследования. Задача-1. Когнитивная структуризация и формализация предметной области. Задача-2. Синтез и верификация моделей (синтез статистических и системно-когнитивных моделей (многопараметрическая типизация), частные критерии знаний, верификация моделей, критерии достоверности моделей, частотные распределения истинных и ложных положительных и отрицательных решений в различных моделях, сравнение достоверности положительных и отрицательных решений в модели с вероятностью случайного угадывания, оценка информационной мощности модели, оценка долей влияния характеристик транзакций в системе детерминации степени риска мошенничества, синтез случайной модели на основе реальной, сравнение частотных распределений истинных и ложных положительных и отрицательных решений в реальной и случайной моделях, сравнение достоверностей положительных и отрицательных решений в реальной и случайной моделях по сравнению со случайным угадыванием, сравнение информационной мощности реальной и

UDC 004.8

08.00.13 - Mathematical and instrumental methods of Economics (Economics)

APPLICATION OF AUTOMATED SYSTEM-COGNITIVE ANALYSIS OF BANK DATABASES ON CREDIT CARD TRANSACTIONS FOR QUANTITATIVE ASSESSMENT OF FRAUD RISK

Lutsenko Evgeniy Veniaminovich
Dr.Sci.Econ., Cand.Tech.Sci., professor
Web of Science ResearcherID S-8667-2018
Scopus Author ID: 57188763047
RSCI id=123162, SPIN-code: 9523-7101
prof.lutsenko@gmail.com <http://lc.kubagro.ru>
<https://www.researchgate.net/profile/Eugene-Lutsenko>
Kuban State Agrarian University named after I.T. Trubilin, Krasnodar, Russia

In this article, we have set, considered and solved the actual problem of reliable quantitative forecasting of fraud risks for credit card transactions. To solve the problem, we used real data from the Kaggle portal. We also used Automated system-cognitive analysis (ASC-analysis) and its software tools, which is currently used as an intelligent system "Eidos". The object, subject, problem, hypothesis, purpose, method, research tasks are considered (research problems, a brief justification for choosing a research method, research tasks). The article also provides a solution of the following research tasks. Task-1. Cognitive structuring and formalization of the subject area. Task-2. Synthesis and verification of models (synthesis of statistical and system-cognitive models (multiparametric typing), private knowledge criteria, model verification, model reliability criteria, frequency distributions of true and false positive and negative decisions in various models, comparison of the reliability of positive and negative decisions in the model with the probability of random guessing, assessment of the information power of the model, assessment of the share of the impact of transaction characteristics in the fraud risk determination system, synthesis of a random model based on real, comparison of the frequency distributions of true and false positive and negative solutions in real and random models, comparison of the reliability of positive and negative solutions in real and random models compared with random guessing, comparison of the information power of real and random models and evaluation of the ratio of the useful signal to noise). Task-3. The solution of the forecasting problem (system identification), integral criteria, what is an integral criterion and what is it needed for, the 1st integral criterion "Sum of knowledge", the 2nd integral criterion "Semantic resonance of knowledge", some mathematical properties of integral criteria, the

случайной моделей и оценка отношения полезного сигнала к шуму). Задача-3. Решение задачи прогнозирования (системная идентификация), интегральные критерии, что такое интегральный критерий и для чего он нужен, 1-й интегральный критерий «Сумма знаний», 2-й интегральный критерий «Семантический резонанс знаний», некоторые математические свойства интегральных критериев, решение задачи идентификации и прогнозирования. Задача-4. Решение задачи принятия решений (результаты многопараметрической типизации классов). Задача-5. Решение задачи исследования моделируемой предметной области путем исследования ее модели (инвертированные SWOT-диаграммы значений характеристик транзакций (семантические потенциалы), кластерно-конструктивный анализ классов, кластерно-конструктивный анализ значений характеристик транзакций, нелокальные нейроны, нелокальная нейронная сеть, 3D-интегральные когнитивные карты, 2D-интегральные когнитивные карты содержательного сравнения классов, 2D-интегральные когнитивные карты содержательного сравнения значений факторов, когнитивные функции, значимость характеристик транзакций и их значений, степень детерминированности классов и классификационных шкал). Все изложение в работе построено на основе подробного численного примера, раскрывающего все перечисленные аспекты решения поставленной задачи

solution of the identification and forecasting problem. Task-4. Solution of the decision-making problem (results of multiparameter class typing). Task-5. Solving the problem of studying the simulated subject area by studying its model (inverted SWOT diagrams of the values of transaction characteristics (semantic potentials), cluster-constructive analysis of classes, cluster-constructive analysis of the values of transaction characteristics, non-local neurons, non-local neural network, 3D-integral cognitive maps, 2D-integral cognitive maps of meaningful comparison of classes, 2D-integral cognitive maps of meaningful comparison of factor values, cognitive functions, the significance of transaction characteristics and their values, the degree of determinism of classes and classification scales). The entire presentation in the work is based on a detailed numerical example that reveals all the listed aspects of solving the problem

Ключевые слова: АВТОМАТИЗИРОВАННЫЙ СИСТЕМНО-КОГНИТИВНЫЙ АНАЛИЗ, АСК-АНАЛИЗ, СИСТЕМА «ЭЙДОС»

Keywords: AUTOMATED SYSTEM-COGNITIVE ANALYSIS, ASC-ANALYSIS, "EIDOS" SYSTEM

DOI: <http://dx.doi.org/10.21515/1990-4665-172-009>

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	3
<i>КОНТЕКСТ</i>	4
<i>СОДЕРЖАНИЕ</i>	4
<i>ОБНОВЛЕНИЕ (05.03.2021)</i>	4
<i>БЛАГОДАРНОСТИ</i>	5
1. ОБЪЕКТ, ПРЕДМЕТ, ПРОБЛЕМА, ГИПОТЕЗА, ЦЕЛЬ, МЕТОД, ЗАДАЧИ ИССЛЕДОВАНИЯ.	5
1.1. ПРОБЛЕМАТИКА ИССЛЕДОВАНИЯ	5
1.2. КРАТКОЕ ОБОСНОВАНИЕ ВЫБОРА МЕТОДА ИССЛЕДОВАНИЯ	6
1.3. ЗАДАЧИ ИССЛЕДОВАНИЯ	9
2. РЕШЕНИЕ ЗАДАЧ ИССЛЕДОВАНИЯ	11
ЗАДАЧА-1. КОГНИТИВНАЯ СТРУКТУРИЗАЦИЯ И ФОРМАЛИЗАЦИЯ ПРЕДМЕТНОЙ ОБЛАСТИ	11
ЗАДАЧА-2. СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛЕЙ	19
2.1. Синтез статистических и системно-когнитивных моделей (многопараметрическая типизация), частные критерии знаний	19
2.2. Верификация моделей	29
2.2.1. Критерии достоверности моделей	29

2.2.2. Частотные распределения истинных и ложных положительных и отрицательных решений в различных моделях	31
2.2.3. Сравнение достоверности положительных и отрицательных решений в модели с вероятностью случайного угадывания	33
2.2.4. Оценка информационной мощности модели	33
2.2.5. Выводы	34
2.3. Оценка долей влияния характеристик транзакций в системе детерминации степени риска мошенничества	34
2.3.1. Синтез случайной модели на основе реальной	34
2.3.2. Сравнение частотных распределений истинных и ложных положительных и отрицательных решений в реальной и случайной моделях	38
2.3.3. Сравнение достоверностей положительных и отрицательных решений в реальной и случайной моделях по сравнению со случайным угадыванием	41
2.3.4. Сравнение информационной мощности реальной и случайной моделей и оценка отношения полезного сигнала к шуму	42
2.3.5. Выводы	48
Задача-3. РЕШЕНИЕ ЗАДАЧИ ПРОГНОЗИРОВАНИЯ (СИСТЕМНАЯ ИДЕНТИФИКАЦИЯ), ИНТЕГРАЛЬНЫЕ КРИТЕРИИ	48
3.1. Что такое интегральный критерий и для чего он нужен?	48
3.2. 1-й интегральный критерий «Сумма знаний»	48
3.3. 2-й интегральный критерий «Семантический резонанс знаний»	49
3.4. Некоторые математические свойства интегральных критериев	50
3.5. Решение задачи идентификации и прогнозирования	51
Задача-4. РЕШЕНИЕ ЗАДАЧИ ПРИНЯТИЯ РЕШЕНИЙ (РЕЗУЛЬТАТЫ МНОГОПАРАМЕТРИЧЕСКОЙ ТИПИЗАЦИИ КЛАССОВ)	58
Задача-5. РЕШЕНИЕ ЗАДАЧИ ИССЛЕДОВАНИЯ МОДЕЛИРУЕМОЙ ПРЕДМЕТНОЙ ОБЛАСТИ ПУТЕМ ИССЛЕДОВАНИЯ ЕЕ МОДЕЛИ	60
5.1. Инвертированные SWOT-диаграммы значений характеристик транзакций (семантические потенциалы)	60
5.2. Кластерно-конструктивный анализ классов	61
5.3. Кластерно-конструктивный анализ значений характеристик транзакций	63
5.4. Нелокальные нейроны	69
5.5. Нелокальная нейронная сеть	70
5.6. 3D-интегральные когнитивные карты	72
5.7. 2D-интегральные когнитивные карты содержательного сравнения классов	75
5.8. 2D-интегральные когнитивные карты содержательного сравнения значений факторов	77
5.9. Когнитивные функции	79
5.10. Значимость характеристик транзакций и их значений	84
5.11. Степень детерминированности классов и классификационных шкал	88
3. ВЫВОДЫ	89
ЛИТЕРАТУРА	90

Введение

Данные для задачи взяты с портала Kaggle по ссылке: <https://www.kaggle.com/mlg-ulb/creditcardfraud/download> Данные и описание размещены Группой машинного обучения – ULB по лицензии Open Data Commons: <https://opendatacommons.org/licenses/dbcl/1-0/>. На самом портале kaggle задача ставится и описывается группой следующим образом <https://www.kaggle.com/mlg-ulb/creditcardfraud>.

Контекст

«Важно, чтобы компании, выпускающие кредитные карты, могли распознавать мошеннические транзакции по кредитным картам, чтобы с клиентов не взималась плата за товары, которые они не покупали»¹.

Содержание

«Набор данных содержит транзакции, совершенные европейскими держателями кредитных карт **в сентябре 2013 года**.

В этом наборе данных представлены транзакции, которые произошли **за два дня**, из которых у нас **492** мошенничества из **284 807** транзакций. *Набор данных сильно не сбалансирован, на положительный класс (мошенничество) приходится 0,172% всех транзакций.*

Он содержит только числовые входные переменные, которые являются результатом преобразования PCA. *К сожалению, из-за проблем с конфиденциальностью мы не можем предоставить исходные функции и дополнительную справочную информацию о данных.* Характеристики V1, V2,... V28 являются основными компонентами, полученными с помощью PCA, единственными функциями, которые не были преобразованы с помощью PCA, являются «Время» и «Сумма». Функция «Время» содержит секунды, прошедшие между каждой транзакцией и первой транзакцией в наборе данных. Функция «Сумма» - это сумма транзакции, эту функцию можно использовать для обучения, зависящего от стоимости, в зависимости от примера. Функция «Класс» - это переменная ответа, которая принимает значение 1 в случае мошенничества и 0 в противном случае.

Учитывая коэффициент дисбаланса класса, мы рекомендуем измерять точность с помощью области под кривой точности-отзыва (AUPRC). Точность матрицы неточностей не имеет значения для несбалансированной классификации»².

Обновление (05.03.2021)

«Симулятор данных транзакций был выпущен как часть практического руководства по машинному обучению для обнаружения мошенничества с кредитными картами - https://fraud-detection-handbook.github.io/fraud-detection-handbook/Chapter_3_GettingStarted/SimulatedDataset.html . Мы приглашаем всех практиков, интересующихся наборами данных по обнаружению мошенничества, также проверить этот симулятор данных и методологии обнаружения мошенничества с кредитными картами, представленные в книге»³.

¹ Источник информации: <https://www.kaggle.com/mlg-ulb/creditcardfraud>

² Источник информации: <https://www.kaggle.com/mlg-ulb/creditcardfraud>

³ Источник информации: <https://www.kaggle.com/mlg-ulb/creditcardfraud>

Благодарности

«Набор данных был собран и проанализирован в ходе исследовательского сотрудничества Worldline и Группы машинного обучения (<http://mlg.ulb.ac.be>) ULB (Université Libre de Bruxelles) по интеллектуальному анализу больших данных и обнаружению мошенничества.

Более подробная информация о текущих и прошлых проектов по темам, связанным доступны на <https://www.researchgate.net/project/Fraud-detection-5> и на странице [DefeatFraud](#) проекта»⁴.

В этой связи необходимо отметить следующие работы [1-9]».

1. Объект, предмет, проблема, гипотеза, цель, метод, задачи исследования

1.1. Проблематика исследования

Объектом исследования являются транзакции, осуществляемые с кредитными картами.

Транзакция (англ. transaction, от лат. transactio — соглашение, договор) — *минимальная логически осмысленная операция, которая имеет смысл и может быть совершена только полностью*. Написание «транзакция» или «транзакция» зависит от предметной области применения данного термина. В финансовой сфере, к которой относится и наше исследование, используется написание «транзакция», а вариант написания «транзакция» принят в области компьютерных и информационных технологий и баз данных. Мы будем пользоваться обоими вариантами этого термина.

Предметом исследования является количественная оценка степени мошенничества при операциях с кредитными картами на основе характеристик транзакций. Для решения этой проблемы применяются технологии искусственного интеллекта.

Традиционно считается, что мошеннические операции с кредитными картами выявить достаточно сложно. Решению этой проблемы посвящено большое количество исследований и разработок. Однако закономерно возникает вопрос об их эффективности, т.к. *факты говорят об огромных убытках физических и юридических лиц от мошеннических операций с кредитными картами*. Чтобы убедиться в этом достаточно сделать в Google запрос: «losses of individuals and legal entities from fraudulent credit card transactions».

Гипотеза: авторы предполагают, что в банковских базах данных по операциям с кредитными картами, содержится информация о связи между значениями характеристик операций и степенью риска мошенничества.

⁴ Источник информации: <https://www.kaggle.com/mlg-ulb/creditcardfraud>

Проблема заключается в существовании *противоречия, несоответствия* между фактической ситуацией и желательной, целевой ситуацией с мошенничеством при операциях с кредитными картами.

Целью работы является решение поставленной проблемы путем разрешения лежащего в ее основе противоречия, т.е. путем уменьшения убытков от мошеннических операций с кредитными картами.

Для этого необходимо аргументировано ответить на несколько вопросов:

1. Существует ли возможность по значениям характеристик операций с кредитными картами количественно оценить для каждой операции степень риска того, что эта операция является мошеннической.

2. Если ответ на первый вопрос положительный, то исследовать силу и направление влияния каждого значения операций на величину риска того, что данная операция является мошеннической.

Отметим, что в данной работе мы не претендуем на полное рассмотрение всех аспектов решения поставленной проблемы и ограничиваемся только тем, что для этого необходимо сделать *в первую очередь*: *количественно определить для каждой операции риск того, что она является мошеннической*. Рассмотрение вопросов о том, что делать потом, после выявления операций с высокой степенью риска мошенничества, не входит в тематику данной работы.

Для проведения данного исследования необходимо обосновать выбор адекватного для этого метода и программного инструментария.

1.2. Краткое обоснование выбора метода исследования

В качестве метода исследования, решения проблемы и достижения цели предлагается применить новый новационный⁵ метод искусственного интеллекта: Автоматизированный системно-когнитивный анализ (АСК-анализ).

Основной причиной выбора АСК-анализа является то, что он включает теорию и метод количественного выявления в сопоставимой форме силы и направления причинно-следственных зависимостей в неполных зашумленных данных очень большой размерности числовой и не числовой природы, измеряемых в различных типах шкал и единицах измерения.

Очень важным является также то, что АСК-анализ имеет свой развитый и доступный программный инструментарий, в качестве которого в настоящее время выступает Универсальная когнитивная аналитическая система «Эйдос».

⁵ Инновационный – готовый к внедрению, новационный – успешно внедренный в эксплуатацию.

Существует много систем искусственного интеллекта. Универсальная когнитивная аналитическая система «Эйдос-X++» отличается от них следующими параметрами:

- является универсальной и может быть применена во многих предметных областях, т.к. разработана в универсальной постановке, не зависящей от предметной области (<http://lc.kubagro.ru/aidos/index.htm>). Система «Эйдос» является автоматизированной системой, т.е. предполагает непосредственное участие человека в реальном времени при решении задач идентификации, прогнозирования, принятия решений и исследования предметной области (автоматические системы работают без такого участия человека);

- находится в полном открытом бесплатном доступе (http://lc.kubagro.ru/aidos/_Aidos-X.htm), причем с актуальными исходными текстами (http://lc.kubagro.ru/_AidosALL.txt): открытая лицензия: [CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/) (<https://creativecommons.org/licenses/by-sa/4.0/>), и это означает, что ей могут пользоваться все, кто пожелает, без какого-либо дополнительного разрешения со стороны первичного правообладателя – автора системы «Эйдос» проф. Е.В.Луценко (отметим, что система «Эйдос» создана полностью с использованием только лицензионного инструментального программного обеспечения и на нее имеется 31 свидетельство РосПатента РФ);

- является одной из первых отечественных систем искусственного интеллекта персонального уровня, т.е. не требует от пользователя специальной подготовки в области технологий искусственного интеллекта: «имеет нулевой порог входа» (есть акт внедрения системы «Эйдос» 1987 года) (<http://lc.kubagro.ru/aidos/aidos02/PR-4.htm>);

- реально работает, обеспечивает устойчивое выявление в сопоставимой форме силы и направления причинно-следственных зависимостей в неполных зашумленных взаимозависимых (нелинейных) данных очень большой размерности числовой и не числовой природы, измеряемых в различных типах шкал (номинальных, порядковых и числовых) и в различных единицах измерения (т.е. не предъявляет жестких требований к данным, которые невозможно выполнить, а обрабатывает те данные, которые есть);

- имеет «нулевой порог входа», содержит большое количество локальных (поставляемых с инсталляцией) и облачных учебных и научных Эйдос-приложений (в настоящее время их 31 и около 300, соответственно: http://aidos.byethost5.com/Source_data_applications/WebAppls.htm) (http://lc.kubagro.ru/aidos/Presentation_Aidos-online.pdf);

- поддерживает on-line среду накопления знаний и обмена ими, широко используется во всем мире (<http://aidos.byethost5.com/map5.php>);

- обеспечивает мультязычную поддержку интерфейса на 51 языке. Языковые базы входят в инсталляцию и могут пополняться в автоматическом режиме;

- наиболее трудоемкие в вычислительном отношении операции синтеза моделей и распознавания реализует с помощью графического процессора (GPU), что на некоторых задачах обеспечивает ускорение решения этих задач в несколько тысяч раз, что реально обеспечивает интеллектуальную обработку больших данных, большой информации и больших знаний (графический процессор должен быть на чипсете NVIDIA);

- обеспечивает преобразование исходных эмпирических данных в информацию, а ее в знания и решение с использованием этих знаний задач классификации, поддержки принятия решений и исследования предметной области путем исследования ее системно-когнитивной модели, генерируя при этом очень большое количество табличных и графических выходных форм (развития когнитивная графика), у многих из которых нет никаких аналогов в других системах (примеры форм можно посмотреть в работе: http://lc.kubagro.ru/aidos/aidos18_LLS/aidos18_LLS.pdf);

- хорошо имитирует человеческий стиль мышления: дает результаты анализа, понятные экспертам на основе их опыта, интуиции и профессиональной компетенции;

- вместо того, чтобы предъявлять к исходным данным практически неосуществимые требования (вроде нормальности распределения, абсолютной точности и полных повторностей всех сочетаний значений факторов и их полной независимости и аддитивности) автоматизированный системно-когнитивный анализ (АСК-анализ) предлагает без какой-либо предварительной обработки осмыслить эти данные и тем самым преобразовать их в информацию, а затем преобразовать эту информацию в знания путем ее применения для достижения целей (т.е. для управления) и решения задач классификации, поддержки принятия решений и содержательного эмпирического исследования моделируемой предметной области.

В чем сила подхода, реализованного в системе Эйдос? В том, что она реализует подход, эффективность которого не зависит от того, что мы думаем о предметной области и думаем ли вообще. Она формирует модели непосредственно на основе эмпирических данных, а не на основе наших представлений о механизмах реализации закономерностей в этих данных. Именно поэтому Эйдос-модели эффективны даже если наши представления о предметной области ошибочны или вообще отсутствуют.

В этом и слабость этого подхода, реализованного в системе Эйдос. Модели системы Эйдос - это феноменологические модели, отражающие эмпирические закономерности в фактах обучающей выборки, т.е. они не отражают причинно-следственного механизма детерминации, а только сам

факт и характер детерминации. Содержательное объяснение этих эмпирических закономерностей формулируется уже экспертами на теоретическом уровне познания в содержательных научных законах.

Система «Эйдос» обеспечивает преобразование исходных данных в информацию, а ее в знания и позволяет решать на основе этих знаний задачи идентификации, прогнозирования, принятия решений и исследования моделируемой предметной области путем исследования ее модели (рисунок 1):



Рисунок 1

1.3. Задачи исследования

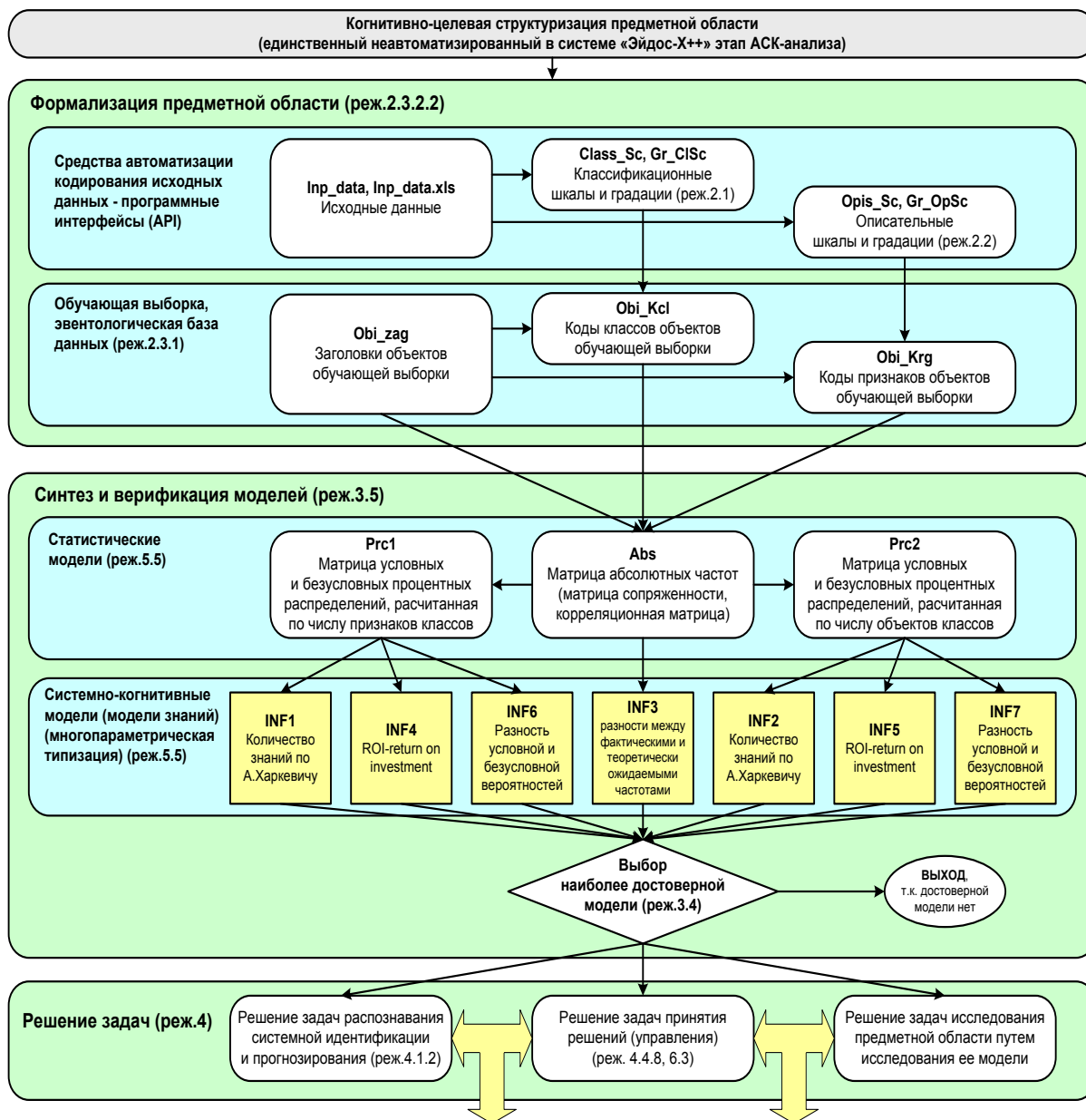
Путем декомпозиции поставленная цель разбивается на ряд задач, решение каждой из которых проще, чем достижение цели и являющихся этапами ее достижения:

- **задача-1:** когнитивная структуризация и формализация предметной области;
- **задача-2:** синтез и верификация моделей;
- **задача-3:** решение задач идентификации и прогнозирования (системная идентификация);
- **задача-4:** решение задачи принятия решений (многопараметрическая типизация классов);
- **задача-5:** решение задачи исследования моделируемой предметной области путем исследования ее модели.

Эти задачи, по сути, являются этапами АСК-анализа.

Рассмотрим их решение в интеллектуальной системе «Эйдос», которая в настоящее время является программным инструментарием АСК-анализа (рисунок 2).

**Последовательность обработки данных, информации и знаний в системе «Эйдос»,
повышение уровня системности данных, информации и знаний,
повышение уровня системности моделей**



О соотношении задач:

- **распознавания, классификации, идентификации и диагностики** (это одно и то же, т.е. синонимы);
- **идентификации и прогнозирования** (при идентификации значения свойств и принадлежность объекта к классу относятся к одному моменту времени, а при прогнозировании значения факторов относятся к прошлому, а переход объекта под действием этих факторов в состояние, соответствующее классу относится к будущему);
- **прогнозирования и принятия решений** (при прогнозировании по значениям факторов, действующих на объект моделирования, определяется в какое будущее состояние он перейдет под их действием. При принятии решений, наоборот, по будущему целевому состоянию объекта моделирования определяются значения факторов, которые обуславливают его переход в это будущее целевое состояние. Таким образом задача принятия решений является обратной по отношению к задаче прогнозирования);
- **принятия решений и исследования моделируемой предметной области** (задача принятия решений является обратной по отношению к задаче прогнозирования только в простейшем случае: в случае использования SWOT-анализа. Однако SWOT-анализ имеет свои ограничения: может быть задано только одно будущее целевое состояние, некоторые рекомендуемые факторы может не быть технологической и финансовой возможности использовать. Поэтому в АСК-анализе и системе «Эйдос» реализован развитый алгоритм принятия решений п.6.3 в котором кроме SWOT-анализа используются также результаты решения задачи прогнозирования и результаты кластерно-конструктивного анализа классов и значений факторов, т.е. некоторые результаты решения задачи исследования предметной области.)

Рисунок 2

2. Решение задач исследования

Задача-1. Когнитивная структуризация и формализация предметной области

При когнитивной структуризации решается, что мы будем рассматривать как объект моделирования, что как влияющие на него факторы, и что как результаты действия этих факторов. Это единственный не автоматизированный в системе «Эйдос» этап АСК-анализа.

В данном случае:

- в качестве объекта моделирования мы рассматриваем операции с кредитными банковскими картами;
- в качестве факторов мы рассматриваем различные характеристики банковских операций, содержащиеся в банковских базах данных;
- в качестве результатов влияния факторов рассматривается риск мошенничества для данной операции.

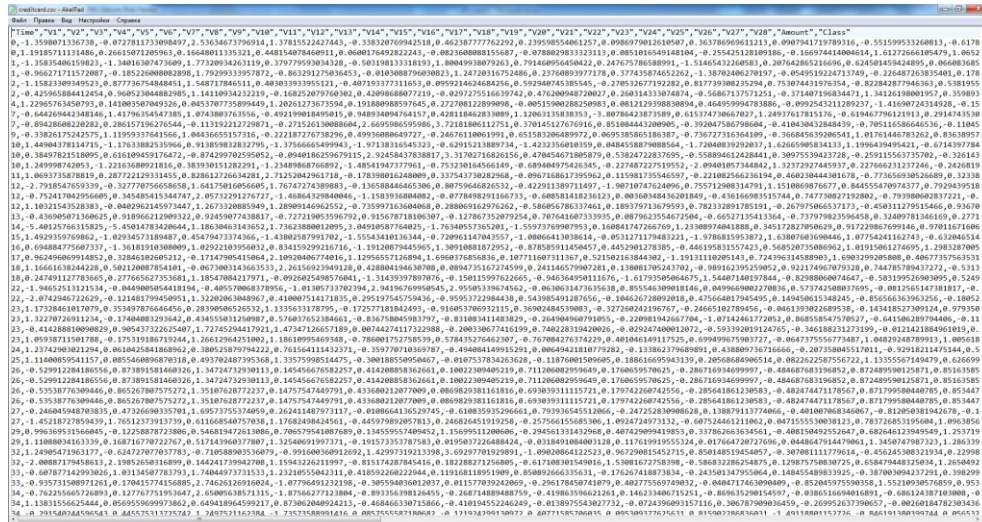
Исходные данные представляют собой банковские базы данных, описывающие транзакции, совершенные *европейскими держателями кредитных карт в сентябре 2013 года*. В этом наборе данных представлены транзакции, которые произошли *за два дня*, из которых у нас **492** мошенничества из **284807** транзакций.

Набор данных сильно не сбалансирован, на положительный класс (мошенничество) приходится 0,172% всех транзакций. Он содержит только числовые входные переменные, которые являются результатом преобразования PCA.

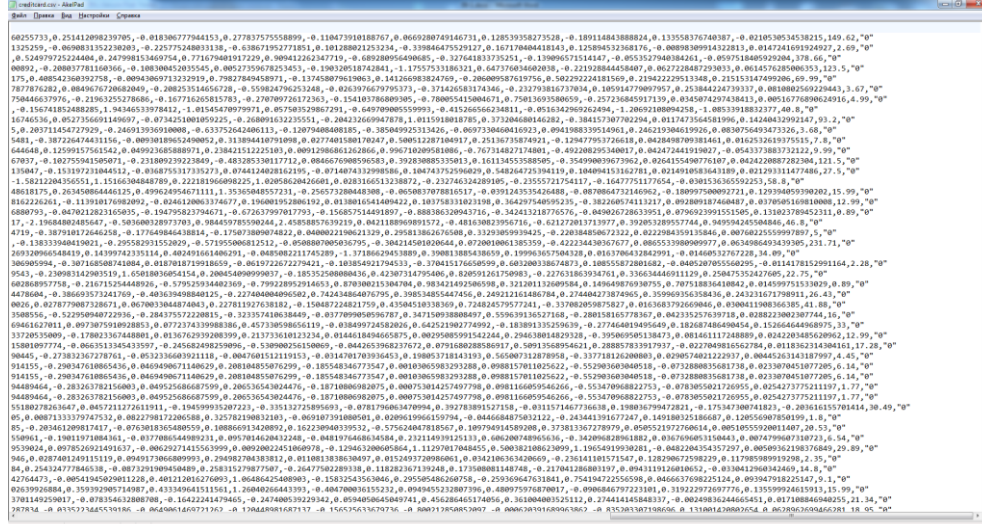
Из-за проблем с конфиденциальностью исходные факторы зашифрованы и дополнительная справочная информация о них не приводится. Характеристики V1, V2,... V28 являются основными компонентами, полученными с помощью PCA, единственными функциями, которые не были преобразованы с помощью PCA, являются «Время» и «Сумма». Функция «Время» содержит секунды, прошедшие между каждой транзакцией и первой транзакцией в наборе данных. Функция «Сумма» - это сумма транзакции, эту функцию можно использовать для обучения, зависящего от стоимости, в зависимости от примера.

Функция «Класс» - это переменная ответа, которая принимает значение 1 в случае мошенничества и 0 в противном случае.

На рисунке 3 приведен небольшой фрагмент исходных данных в том виде, в каком он скачивается с портала Kaggle по прямой ссылке: <https://www.kaggle.com/mlg-ulb/creditcardfraud/download>.



Самая левая часть CSV-файла исходных данных



Самая правая часть CSV-файла исходных данных

Рисунок 3

CSV-файл исходных данных имеет довольно большой объем: 150828752 байтов (примерно 143,8 Мб).

Для ввода в систему «Эйдос» необходимо преобразовать это CSV-файл в XLSX-формат MS Excel с помощью программы XLSX2XLS, т.к. в нем описано 284807 транзакций, а в XLS-файле может быть не более 65536 строк. Кроме того, программа XLSX2XLS-файл обычно преобразует в два раза меньше, чем XLS-файл с той же информацией.

Для самого преобразования применим онлайн csv-xlsx конвертер: <https://online-converting.ru/documents/csv-to-xlsx/>.

Именно данный конвертер выбран потому, что он бесплатный, правильно и быстро работает, но главное, преобразует файлы размером более 100 Мб, чего обычно другие онлайн csv-xlsx конвертеры не обеспечивают.

В результате он-лайн конвертирования получаем XLSX-файл, фрагмент которого приведен на рисунке 4:

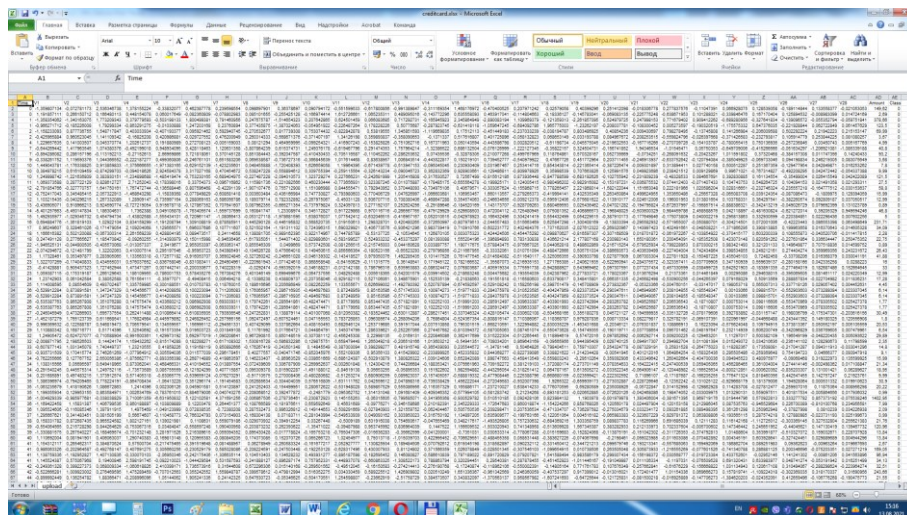


Рисунок 4 . Исходные данные до форматирования (фрагмент)

Для ввода в систему «Эйдос» немного преобразуем файл, приведенный на рисунке 4:

- добавим первую колонку с наименованием транзакции, которая не является шкалой;
- в последней колонке переименуем 1 на «Cheater», 0 на «Not a fraudster», выделим эту колонку желтым фоном для удобства, как принято в системе «Эйдос»;
- сортируем таблицу по последней колонке в порядке возрастания, чтобы описания всех случаев мошенничества шли подряд в первых строках, а затем шли все описания «нормальных» операций. Это потребуется в дальнейшем.

В результате получим файл исходных данных, представленный на рисунке 5:

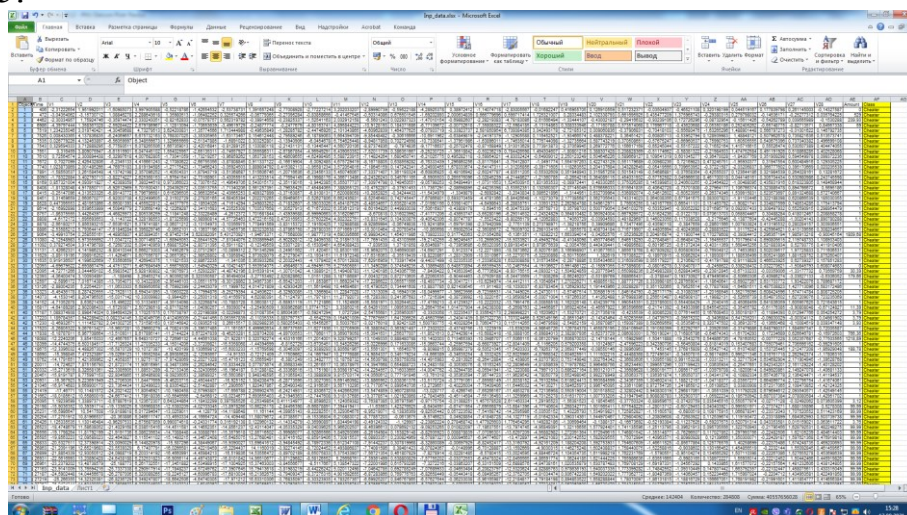


Рисунок 5 . Исходные данные после форматирования (фрагмент)

В результате когнитивной структуризации разработаны классификационные и описательные шкалы и градации (таблицы 1 и 2):

Таблица 1 – Классификационные шкалы

Код	Наименование
1	CLASS

Таблица 2 – Описательные шкалы

Код	Наименование
1	TIME
2	V1
3	V2
4	V3
5	V4
6	V5
7	V6
8	V7
9	V8
10	V9
11	V10
12	V11
13	V12
14	V13
15	V14
16	V15
17	V16
18	V17
19	V18
20	V19
21	V20
22	V21
23	V22
24	V23
25	V24
26	V25
27	V26
28	V27
29	V28
30	AMOUNT

При **формализации предметной области** (см. рисунок 2) на основе исходных данных (таблица 2) **автоматически** разрабатываются градации классификационных и описательных шкал (таблицы 3 и 4), а затем с их помощью кодируются исходные данные, в результате чего получается обучающая выборка (ее еще называют тренировочной выборкой) (таблица 5). Обучающая выборка представляет собой исходные данные, нормализованные с помощью классификационных и описательных шкал и градаций.

Для осуществления автоматизированной формализации предметной области записываем файл исходных Inp_data.xlsx данных в папку:

c:\Aidos-X\AID_DATA\Inp_data\Inp_data.xlsx

и запускаем систему «Эйдос» (рисунок 6):

c:\Aidos-X__START_AIDOS-X.exe.

Затем заходим в режим 2.3.2.2, представляющий собой автоматизированный программный интерфейс (API) системы «Эйдос» с табличными внешними данными, и задаем параметры, приведенные на рисунках 7, 8, 9.

Из этих рисунков мы видим, что весь процесс разработки классификационных и описательных шкал и градаций (таблицы 3 и 4) и формирования обучающей выборки (таблица 5) в данной задаче на среднем компьютере занял 1 час 8 минут 48 секунды.

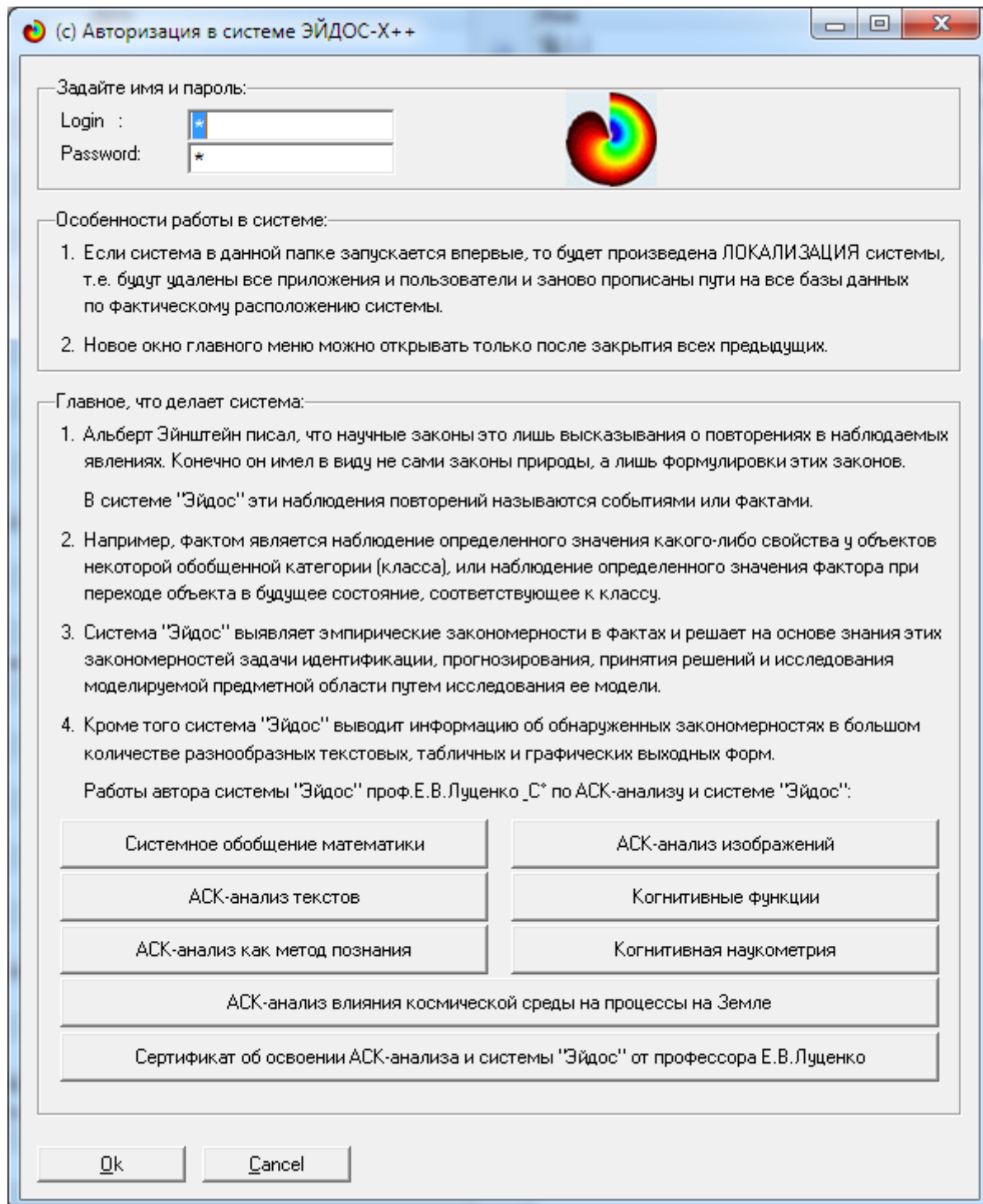


Рисунок 6

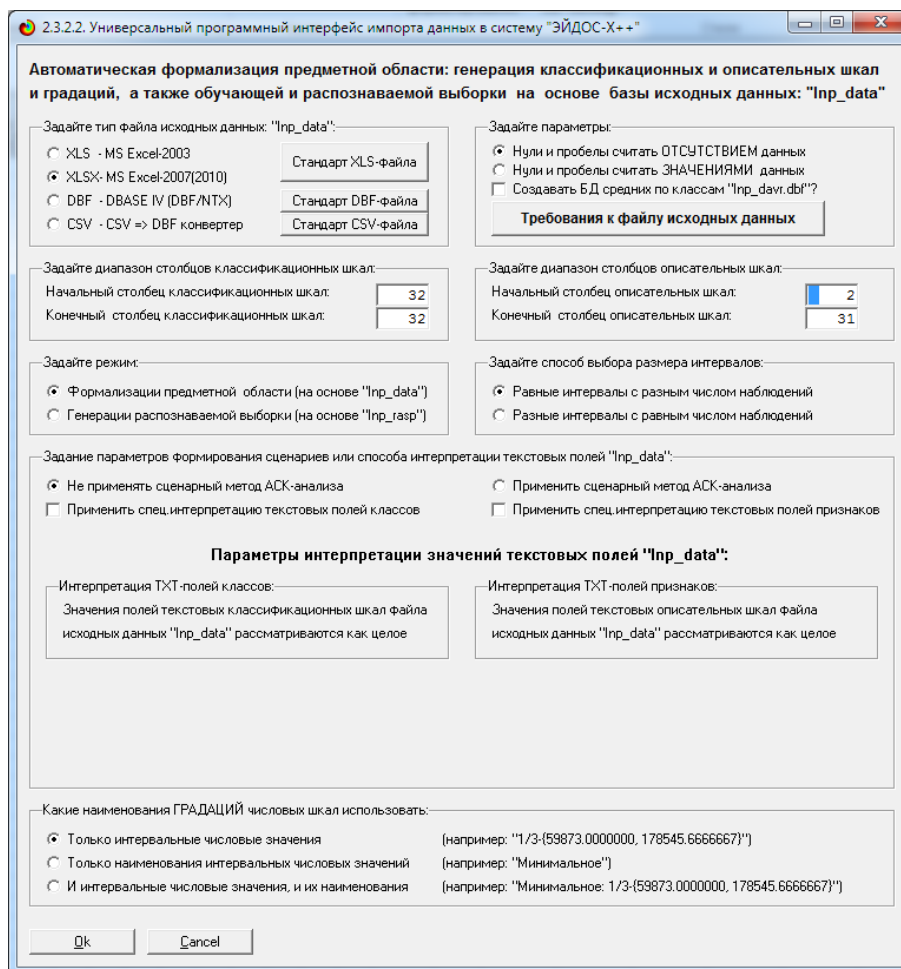


Рисунок 7

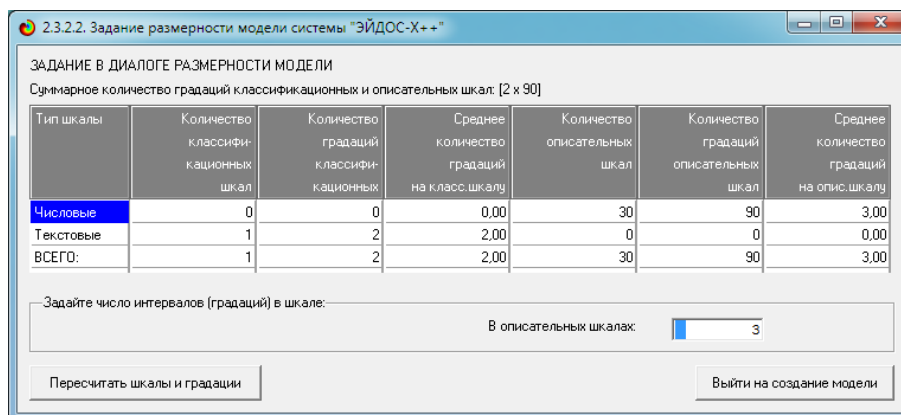


Рисунок 8

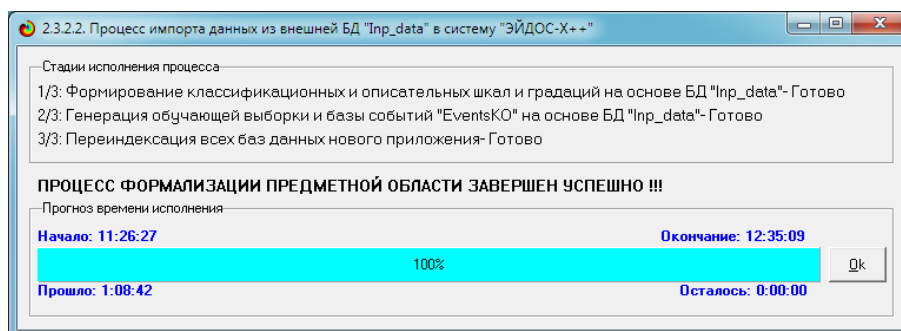


Рисунок 9

Таблица 3 – Классификационные шкалы и градации

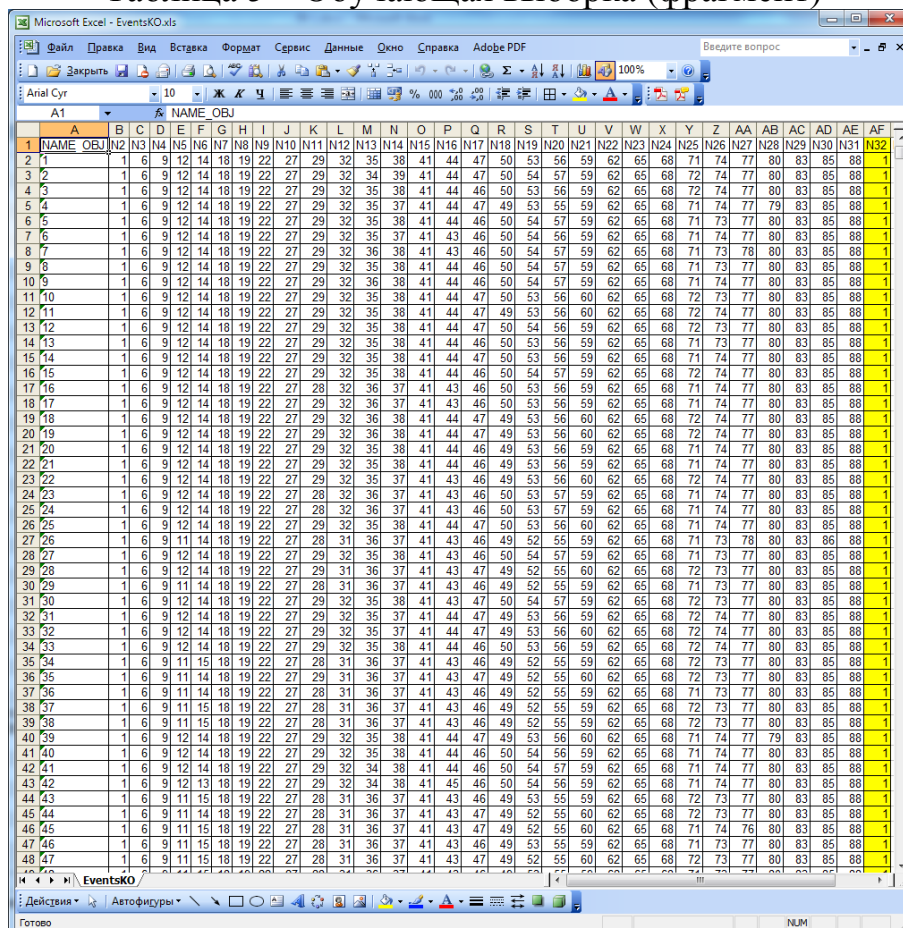
Код	Наименование
1	CLASS-Cheater
2	CLASS-Not a fraudster

Таблица 4 – Описательные шкалы и градации

KOD_ATR	NAME_ATR
1	TIME-1/3-{0.0000000, 57597.3333333}
2	TIME-2/3-{57597.3333333, 115194.6666667}
3	TIME-3/3-{115194.6666667, 172792.0000000}
4	V1-1/3-{-56.4075096, -36.7866964}
5	V1-2/3-{-36.7866964, -17.1658832}
6	V1-3/3-{-17.1658832, 2.4549300}
7	V2-1/3-{-72.7157276, -41.1245754}
8	V2-2/3-{-41.1245754, -9.5334232}
9	V2-3/3-{-9.5334232, 22.0577290}
10	V3-1/3-{-48.3255894, -29.0895401}
11	V3-2/3-{-29.0895401, -9.8534909}
12	V3-3/3-{-9.8534909, 9.3825584}
13	V4-1/3-{-5.6831712, 1.8363339}
14	V4-2/3-{-1.8363339, 9.3558389}
15	V4-3/3-{-9.3558389, 16.8753440}
16	V5-1/3-{-113.7433067, -64.2283158}
17	V5-2/3-{-64.2283158, -14.7133250}
18	V5-3/3-{-14.7133250, 34.8016659}
19	V6-1/3-{-26.1605059, 6.9935379}
20	V6-2/3-{-6.9935379, 40.1475817}
21	V6-3/3-{-40.1475817, 73.3016255}
22	V7-1/3-{-43.5572416, 11.1583369}
23	V7-2/3-{-11.1583369, 65.8739154}
24	V7-3/3-{-65.8739154, 120.5894939}
25	V8-1/3-{-73.2167185, -42.1420762}
26	V8-2/3-{-42.1420762, -11.0674339}
27	V8-3/3-{-11.0674339, 20.0072084}
28	V9-1/3-{-13.4340663, -3.7577127}
29	V9-2/3-{-3.7577127, 5.9186410}
30	V9-3/3-{-5.9186410, 15.5949946}
31	V10-1/3-{-24.5882624, -8.4771296}
32	V10-2/3-{-8.4771296, 7.6340033}
33	V10-3/3-{-7.6340033, 23.7451361}
34	V11-1/3-{-4.7974735, 0.8079887}
35	V11-2/3-{-0.8079887, 6.4134510}
36	V11-3/3-{-6.4134510, 12.0189132}
37	V12-1/3-{-18.6837146, -9.8396790}
38	V12-2/3-{-9.8396790, -0.9956435}
39	V12-3/3-{-0.9956435, 7.8483921}
40	V13-1/3-{-5.7918812, -1.4856265}
41	V13-2/3-{-1.4856265, 2.8206283}
42	V13-3/3-{-2.8206283, 7.1268830}
43	V14-1/3-{-19.2143255, -9.3006283}
44	V14-2/3-{-9.3006283, 0.6130689}
45	V14-3/3-{-0.6130689, 10.5267661}
46	V15-1/3-{-4.4989447, -0.0400493}
47	V15-2/3-{-0.0400493, 4.4188462}
48	V15-3/3-{-4.4188462, 8.8777416}
49	V16-1/3-{-14.1298545, -3.6481992}
50	V16-2/3-{-3.6481992, 6.8334562}
51	V16-3/3-{-6.8334562, 17.3151115}
52	V17-1/3-{-25.1627994, -13.6906908}
53	V17-2/3-{-13.6906908, -2.2185823}
54	V17-3/3-{-2.2185823, 9.2535263}
55	V18-1/3-{-9.4987459, -4.6521409}
56	V18-2/3-{-4.6521409, 0.1944642}
57	V18-3/3-{-0.1944642, 5.0410692}
58	V19-1/3-{-7.2135274, -2.9450278}
59	V19-2/3-{-2.9450278, 1.3234718}
60	V19-3/3-{-1.3234718, 5.5919714}
61	V20-1/3-{-54.4977205, -23.1915123}
62	V20-2/3-{-23.1915123, 8.1146960}
63	V20-3/3-{-8.1146960, 39.4209042}
64	V21-1/3-{-34.8303821, -14.1526417}

65	V21-2/3-{-14.1526417, 6.5250988}
66	V21-3/3-{6.5250988, 27.2028392}
67	V22-1/3-{-10.9331437, -3.7877324}
68	V22-2/3-{-3.7877324, 3.3576788}
69	V22-3/3-{3.3576788, 10.5030901}
70	V23-1/3-{-44.8077352, -22.3623529}
71	V23-2/3-{-22.3623529, 0.0830294}
72	V23-3/3-{0.0830294, 22.5284117}
73	V24-1/3-{-2.8366269, -0.3629016}
74	V24-2/3-{-0.3629016, 2.1108238}
75	V24-3/3-{2.1108238, 4.5845491}
76	V25-1/3-{-10.2953971, -4.3570685}
77	V25-2/3-{-4.3570685, 1.5812601}
78	V25-3/3-{1.5812601, 7.5195887}
79	V26-1/3-{-2.6045506, -0.5639185}
80	V26-2/3-{-0.5639185, 1.4767135}
81	V26-3/3-{1.4767135, 3.5173456}
82	V27-1/3-{-22.5656793, -4.5063868}
83	V27-2/3-{-4.5063868, 13.5529056}
84	V27-3/3-{13.5529056, 31.6121981}
85	V28-1/3-{-15.4300839, 0.9958800}
86	V28-2/3-{0.9958800, 17.4218439}
87	V28-3/3-{17.4218439, 33.8478078}
88	AMOUNT-1/3-{0.0000000, 8563.7200000}
89	AMOUNT-2/3-{8563.7200000, 17127.4400000}
90	AMOUNT-3/3-{17127.4400000, 25691.1600000}

Таблица 5 – Обучающая выборка (фрагмент)



Задача-2. Синтез и верификация моделей

2.1. Синтез статистических и системно-когнитивных моделей (многопараметрическая типизация), частные критерии знаний

Математическая модель АСК-анализа и системы «Эйдос» основана на системной нечеткой интервальной математике и обеспечивает сопоставимую обработку больших объемов фрагментированных и зашумленных взаимозависимых данных, представленных в различных типах шкал (номинальных, порядковых и числовых) и различных единицах измерения.

Суть математической модели АСК-анализа состоит в следующем.

Непосредственно на основе эмпирических данных (см. Help режима 2.3.2.2) рассчитывается матрица абсолютных частот (таблица 6).

Таблица 6 – Матрица абсолютных частот (статистическая модель ABS)

		Классы					Сумма
		1	...	<i>j</i>	...	<i>W</i>	
Значения факторов	1	N_{11}		N_{1j}		N_{1W}	
	...						
	<i>i</i>	N_{i1}		N_{ij}		N_{iW}	$N_{i\Sigma} = \sum_{j=1}^W N_{ij}$
	...						
	<i>M</i>	N_{M1}		N_{Mj}		N_{MW}	
Суммарное количество признаков по классу				$N_{\Sigma j} = \sum_{i=1}^M N_{ij}$			$N_{\Sigma\Sigma} = \sum_{i=1}^W \sum_{j=1}^M N_{ij}$
Суммарное количество объектов обучающей выборки по классу				$N_{\Sigma j}$			$N_{\Sigma\Sigma} = \sum_{j=1}^W N_{\Sigma j}$

На ее основе рассчитываются матрицы условных и безусловных процентных распределений (таблица 7).

Отметим, что в АСК-анализе и его программном инструментарии интеллектуальной системе «Эйдос» используется два способа расчета матриц условных и безусловных процентных распределений:

1-й способ: в качестве $N_{\Sigma j}$ используется суммарное количество признаков по классу;

2-й способ: в качестве $N_{\Sigma j}$ используется суммарное количество объектов обучающей выборки по классу.

Таблица 7 – Матрица условных и безусловных процентных распределений (статистические модели PRC1 и PRC2)

		Классы					Безусловная вероятность признака
		1	...	j	...	w	
Значения факторов	1	P_{11}		P_{1j}		P_{1w}	
	...						
	i	P_{i1}		$P_{ij} = \frac{N_{ij}}{N_{\Sigma j}}$		P_{iw}	$P_{i\Sigma} = \frac{N_{i\Sigma}}{N_{\Sigma\Sigma}}$
	...						
	M	P_{M1}		P_{Mj}		P_{Mw}	
Безусловная вероятность класса				$P_{\Sigma j}$			

На практике часто встречается существенная несбалансированность данных, под которой понимается сильно отличающееся количество объектов обучающейся выборки, относящихся к различным классам. Поэтому решать задачу на основе непосредственно матрицы абсолютных частот (таблица 6) было бы очень неразумно и переход от абсолютных частот к условным и безусловным относительным частотам (частостям) является весьма обоснованным и логичным.

Этот переход полностью снимает проблему несбалансированности данных, т.к. в последующем анализе используется не матрица абсолютных частот, а матрицы условных и безусловных процентных распределений и матрицы системно-когнитивных моделей (СК-модели, таблица 9), в частности матрица информативностей.

Этот подход снимает также проблему обеспечения сопоставимости обработки в одной модели исходных данных, представленных в различных видах шкал (номинальных, порядковых и числовых) и в разных единицах измерения.

В системе «Эйдос» это осуществляется всегда при решении любых задач.

Затем на основе таблиц 6 и 7 с использованием частных критериев, знаний приведенных таблице 8, рассчитываются матрицы системно-когнитивных моделей (таблица 9).

Таблица 8– Различные аналитические формы частных критериев знаний, применяемые в АСК-анализе и системе «Эйдос»

Наименование модели знаний и частный критерий	Выражение для частного критерия	
	через относительные частоты	через абсолютные частоты
ABS , матрица абсолютных частот, N_{ij} - фактическое число встреч i -го признака у объектов j -го класса; \bar{N}_{ij} - теоретическое число встреч i -го признака у объектов j -го класса; N_i – суммарное количество признаков в i -й строке; N_j – суммарное количество признаков или объектов обучающей выборки в j -м классе; N – суммарное количество признаков по всей выборке (таблица 1)	---	$N_i = \sum_{j=1}^M N_{ij}; N_j = \sum_{i=1}^W N_{ij}; N = \sum_{i=1}^W \sum_{j=1}^M N_{ij}$ $N_{ij} - \text{фактическая частота};$ $\bar{N}_{ij} = \frac{N_i N_j}{N} - \text{теоретическая частота}$
PRC1 , матрица условных P_{ij} и безусловных P_i процентных распределений, в качестве N_j используется суммарное количество признаков по классу	---	$P_{ij} = \frac{N_{ij}}{N_j}; P_i = \frac{N_i}{N}$
PRC2 , матрица условных P_{ij} и безусловных P_i процентных распределений, в качестве N_j используется суммарное количество объектов обучающей выборки по классу		
INF1 , частный критерий: количество знаний по А.Харкевичу, 1-й вариант расчета вероятностей: N_j – суммарное количество признаков по j -му классу. Вероятность того, что если у объекта j -го класса обнаружен признак, то это i -й признак	$I_{ij} = \Psi \times \log_2 \frac{P_{ij}}{P_i} \Psi \times \log_2 \frac{N_j}{\bar{N}_{ij}} = \Psi \times \log_2 \frac{N_j N_i}{N_i N_j}$	
INF2 , частный критерий: количество знаний по А.Харкевичу, 2-й вариант расчета вероятностей: N_j – суммарное количество объектов по j -му классу. Вероятность того, что если предъявлен объект j -го класса, то у него будет обнаружен i -й признак.		
INF3 , частный критерий: Хи-квадрат : разности между фактическими и теоретически ожидаемыми абсолютными частотами	---	$I_{ij} = N_{ij} - \bar{N}_{ij} = N_{ij} - \frac{N_i N_j}{N}$
INF4 , частный критерий: ROI - Return On Investment, 1-й вариант расчета вероятностей: N_j – суммарное количество признаков по j -му классу	$I_{ij} = \frac{P_{ij}}{P_i} - 1 = \frac{P_{ij} - P_i}{P_i} \frac{N_j}{\bar{N}_{ij}} - 1 = \frac{N_{ij} N_i}{N_i N_j} - 1$	---
INF5 , частный критерий: ROI - Return On Investment, 2-й вариант расчета вероятностей: N_j – суммарное количество объектов по j -му классу		
INF6 , частный критерий: разность условной и безусловной вероятностей, 1-й вариант расчета вероятностей: N_j – суммарное количество признаков по j -му классу	$I_{ij} = P_{ij} - P_i$	$I_{ij} = \frac{N_{ij}}{N_j} - \frac{N_i}{N}$
INF7 , частный критерий: разность условной и безусловной вероятностей, 2-й вариант расчета вероятностей: N_j – суммарное количество объектов по j -му классу		

Обозначения к таблице 3:*i* – значение прошлого параметра;*j* – значение будущего параметра;*N_{ij}* – количество встреч *j*-го значения будущего параметра при *i*-м значении прошлого параметра;*M* – суммарное число значений всех прошлых параметров;*W* – суммарное число значений всех будущих параметров.*N_i* – количество встреч *i*-м значения прошлого параметра по всей выборке;*N_j* – количество встреч *j*-го значения будущего параметра по всей выборке;*N* – количество встреч *j*-го значения будущего параметра при *i*-м значении прошлого параметра по всей выборке.*I_{ij}* – частный критерий знаний: количество знаний в факте наблюдения *i*-го значения прошлого параметра о том, что объект перейдет в состояние, соответствующее *j*-му значению будущего параметра;*Ψ* – нормировочный коэффициент (Е.В.Луценко, 2002), преобразующий количество информации в формуле А.Харкевича в биты и обеспечивающий для нее соблюдение принципа соответствия с формулой Р.Хартли;*P_i* – безусловная относительная частота встречи *i*-го значения прошлого параметра в обучающей выборке;*P_{ij}* – условная относительная частота встречи *i*-го значения прошлого параметра при *j*-м значении будущего параметра.

В таблице 8 приведены формулы:

- для сравнения **фактических и теоретических абсолютных частот**;
- для сравнения **условных и безусловных относительных частот** («вероятностей»).

И это сравнение в таблице 8 осуществляется двумя возможными способами: путем **вычитания** и путем **деления**.

Когда мы сравниваем фактические и теоретические абсолютные частоты путем вычитания у нас получается частный критерий знаний хи-квадрат (СК-модель INF3), когда же мы сравниваем их путем деления, то у нас получается частный критерий количество информации по А.Харкевичу (СК-модели INF1, INF2) или коэффициент возврата инвестиций ROI - Return On Investment (СК-модели INF4, INF5) в зависимости от способа нормировки.

Когда мы сравниваем условные и безусловные относительные частоты путем вычитания у нас получается частный критерий знаний коэффициент взаимосвязи (СК-модели INF6, INF7), когда же мы сравниваем их путем деления, то у нас получается частный критерий количество информации по А.Харкевичу (СК-модели INF1, INF2).

Таким образом, мы видим, что **все частные критерии знаний тесно взаимосвязаны друг с другом**. Особенно интересна связь знаменитого критерия хи-квадрат Пирсона с замечательной мерой количества информации А.Харкевича и с известным в экономике коэффициентом ROI.

Вероятность рассматривается как предел, к которому стремится относительная частота (отношение количества благоприятных исходов к числу испытаний) при **неограниченном** увеличении количества испытаний. Ясно, что вероятность – это математическая абстракция, которая никогда не встречается на практике (также как и другие

математические и физические абстракции, типа математической точки, материальной точки, бесконечно малой и т.п.). На практике встречается только относительная частота. Но она может быть весьма близкой к вероятности. Например, при 480 наблюдений различие между относительной частотой и вероятностью (погрешность) составляет около 5%, при 1250 наблюдениях – около 2.5%, при 10000 наблюдениях – 1%.

Таблица 9 – Матрица системно-когнитивной модели

		Классы					Значимость фактора
		1	...	<i>j</i>	...	<i>W</i>	
Значения факторов	1	I_{11}		I_{1j}		I_{1W}	$\sigma_{1\Sigma} = \sqrt[2]{\frac{1}{W-1} \sum_{j=1}^W (I_{1j} - \bar{I}_1)^2}$
	...						
	<i>i</i>	I_{i1}		I_{ij}		I_{iW}	$\sigma_{i\Sigma} = \sqrt[2]{\frac{1}{W-1} \sum_{j=1}^W (I_{ij} - \bar{I}_i)^2}$
	...						
Степень редукции класса	<i>M</i>	I_{M1}		I_{Mj}		I_{MW}	$\sigma_{M\Sigma} = \sqrt[2]{\frac{1}{W-1} \sum_{j=1}^W (I_{Mj} - \bar{I}_M)^2}$
		$\sigma_{\Sigma 1}$		$\sigma_{\Sigma j}$		$\sigma_{\Sigma W}$	$H = \sqrt[2]{\frac{1}{(W \cdot M - 1)} \sum_{j=1}^W \sum_{i=1}^M (I_{ij} - \bar{I})^2}$

Суть этих методов в том, что вычисляется количество информации в значении фактора о том, что объект моделирования перейдет под его действием в определенное состояние, соответствующее классу. Это позволяет сопоставимо и корректно обрабатывать разнородную информацию о наблюдениях объекта моделирования, представленную в различных типах измерительных шкал и различных единицах измерения.

На основе системно-когнитивных моделей, представленных в таблице 9 (отличаются частыми критериями, приведенными в таблице 8), решаются задачи идентификации (классификации, распознавания, диагностики, прогнозирования), поддержки принятия решений (обратная задача прогнозирования), а также задача исследования моделируемой предметной области путем исследования ее системно-когнитивной модели.

Отметим, что как значимость значения фактора, степень детерминированности класса и ценность или качество модели в АСК-анализе рассматривается вариабельность значений частных критериев этого значения фактора, класса или модели в целом (таблица 9).

*Численно эта вариабельность может измеряться разными способами, например средним отклонением модулей частных критериев от среднего, дисперсией или среднеквадратичным отклонением или его квадратом. В системе «Эйдос» принят последний вариант, т.к. эта величина совпадает с **мощностью** сигнала, в частности мощностью информации, а в АСК-анализе все модели рассматриваются в как источник информации об объекте моделирования.*

Поэтому есть все основания уточнить традиционную терминологию АСК-анализа (таблица 10):

Таблица 10 – Уточнение терминологии АСК-анализа

№	Традиционные термины (синонимы)	Новый термин	Формула
1	1. Значимость значения фактора (признака). 2. Дифференцирующая мощность значения фактора (признака). 3. Ценность значения фактора (признака) для решения задачи идентификации и других задач	Корень из информационной мощности значения фактора	$\sigma_{i\Sigma} = \sqrt[2]{\frac{1}{W-1} \sum_{j=1}^W (I_{ij} - \bar{I}_i)^2}$
2	1. Степень детерминированности класса. 2. Степень обусловленности класса.	Корень из информационной мощности класса	$\sigma_{\Sigma j} = \sqrt[2]{\frac{1}{M-1} \sum_{i=1}^M (I_{ij} - \bar{I}_j)^2}$
3	1. Качество модели. 2. Ценность модели. 3. Степень сформированности модели. 4. Количественная мера степени выраженности закономерностей в моделируемой предметной области	Корень из информационной мощности модели	$H = \sqrt[2]{\frac{1}{(W \cdot M - 1)} \sum_{j=1}^W \sum_{i=1}^M (I_{ij} - \bar{I})^2}$

Отметим, что впервые количественное выражение для корня информационной мощности модели предложено проф. Е.В.Луценко в работе [10] еще в 2002 году.⁶

Для синтеза 3 статистических и 7 системно-когнитивных моделей запускаем режим 3.5 с параметрами, приведенными на рисунке 10.

⁶ https://elibrary.ru/download/elibrary_18632909_40810830.pdf, формула (3.81) на стр.290

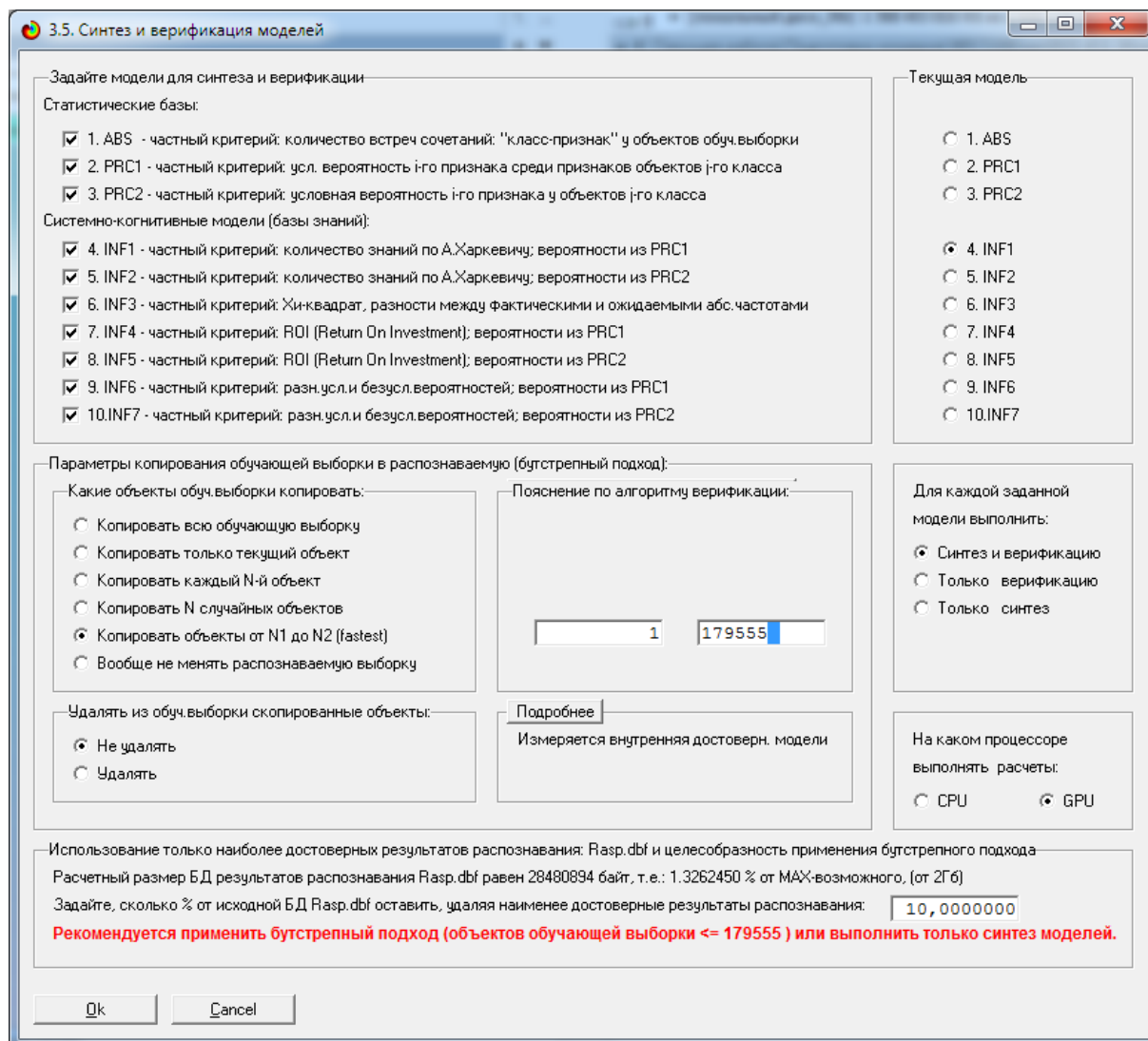


Рисунок 10

Обращаем внимание на то, что для расчетов на экранной форме **выбираем** графический процессор (GPU), т.к. расчеты на нем идут в несколько тысяч раз быстрее, чем на центральном процессоре (CPU). Конкретно это ускорение зависит от многих параметров: характеристик размерностей исходных данных и моделей, типа центрального процессор и видеокарты (числа шейдерных процессоров), размера оперативной памяти компьютера, наличия на нем накопителя SSD с интерфейсом M2 и т.п.

Как только мы запускаем режим 3.5 в нашей задаче, сразу появляется экранная форма с сообщением (рисунок 11):

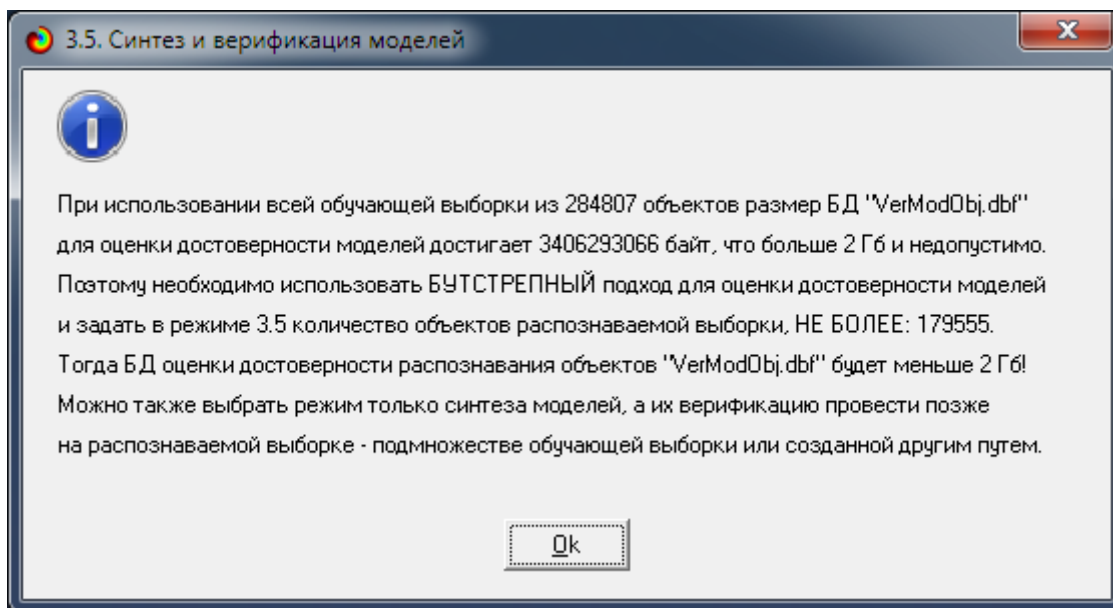


Рисунок 11

Это сообщение не требует особых комментариев. Добавим лишь, что и на экранной форме на рисунке 10 также приведена рекомендация применить бутстрепный подход, ограничив распознаваемую выборку не более чем 179555 записей. Тогда режим синтеза и верификации моделей пройдет до конца без остановки. Если же верификацию моделей осуществлять на всей обучающей выборке из 284807 записей, то при расчете *одной* из выходных форм в каждой из моделей будет появляться сообщение, что она не будет полностью рассчитана. В группе параметров бутстрепного режима необходимо выбрать опцию: «Копировать объекты от N1 до N2 (fastest)» и задать в ней все записи распознаваемой выборки от 1 до 179555-й. Так как исходные данные (таблица 2) были предварительно рассортированы по колонке классов в порядке возрастания, то все случаи мошенничества попадут в распознаваемую выборку. Для этого она и была рассортирована таким образом.

Как видно из рисунка 12, расчет и верификация всех 10 статистических и системно-когнитивных моделей на обучающей выборке из 284807 наблюдений и распознаваемой выборки 179555 наблюдений заняли 1 час 44 минуты 37 секунд. Если бы графический процессор использовался не только для синтеза моделей и распознавания, но и для верификации моделей и расчета всех выходных форм по результатам распознавания, то время расчетов сократилось бы до нескольких секунд.

Обратим внимание также на то, что данная модель имеет размерность около 1,3% от теоретически максимально-возможной в системе «Эйдос» (1-й рисунок 10).

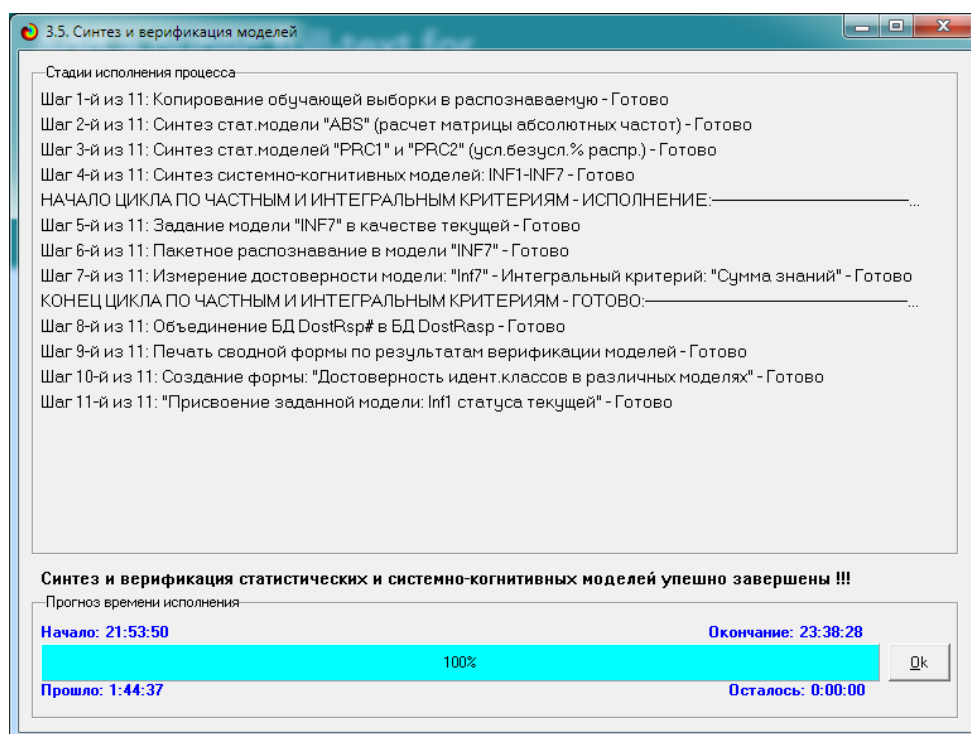


Рисунок 12

В результате выполнения режима 3.5 созданы следующие статистические и системно-когнитивные модели: ABS, PRC1, PRC2, INF1, INF2, INF3, INF4, INF5, INF6, INF7 (таблица 8). Фрагменты некоторых из этих моделей приведены ниже на рисунке 13:

Код признака	Наименование описательной шкалы и градации	1. CLASS CHEATER	2. CLASS NOT A FRAUDSTER	Сумма	Среднее	Средн. квадрат. откл.
1	TIME-1/3-{0.000000, 57597.333333}	191	78387	78578	39289.00	55292.92
2	TIME-2/3-{57597.333333, 115194.666667}	165	83779	83944	41972.00	59124.03
3	TIME-3/3-{115194.666667, 172792.000000}	136	122149	122285	61142.50	86276.22
4	V1-1/3-{-56.4075096, -36.7866964}		8	8	4.00	5.66
5	V1-2/3-{-36.7866964, -17.1658832}	37	199	236	118.00	114.55
6	V1-3/3-{-17.1658832, 2.4549300}	455	284107	284562	142281.00	200572.25
7	V2-1/3-{-72.7157276, -41.1245754}		9	9	4.50	6.36
8	V2-2/3-{-41.1245754, -9.5334232}		752	752	376.00	531.74
9	V2-3/3-{-9.5334232, 22.0577290}	491	283554	284045	142022.50	200155.77
10	V3-1/3-{-48.3255894, -29.0895401}	8	5	13	6.50	2.12
11	V3-2/3-{-29.0895401, -9.8534909}	99	174	273	136.50	53.03
12	V3-3/3-{-9.8534909, 9.3825584}	385	284135	284520	142260.00	200641.55
13	V4-1/3-{-5.6831712, 1.8363339}	88	264026	264114	132057.00	186632.35
14	V4-2/3-{-1.8363339, 9.3558389}	374	20249	20623	10311.50	14053.75
15	V4-3/3-{-9.3558389, 16.8753440}	30	40	70	35.00	7.07
16	V5-1/3-{-113.7433067, -64.2283158}		1	1	0.50	0.71
17	V5-2/3-{-64.2283158, -14.7133250}	27	67	94	47.00	28.28
18	V5-3/3-{-14.7133250, 34.8016659}	465	284246	284711	142355.50	200663.47
19	V6-1/3-{-26.1605059, 6.9935379}	492	284154	284646	142323.00	200579.32
20	V6-2/3-{-6.9935379, 40.1475817}		160	160	80.00	113.14
21	V6-3/3-{-40.1475817, 73.3016255}		1	1	0.50	0.71
22	V7-1/3-{-43.5572416, 11.1583369}	492	284177	284669	142334.50	200595.59
23	V7-2/3-{-11.1583369, 65.8739154}		137	137	68.50	96.87
24	V7-3/3-{-65.8739154, 120.5894939}					
25	V8-1/3-{-73.2167185, -42.1420762}		4	4	2.00	2.83
26	V8-2/3-{-42.1420762, -11.0674339}	19	432	451	225.50	292.04

5.5. Модель: "3. PRC2 - частный критерий: условная вероятность i-го признака у объектов j-го класс..."

Код признака	Наименование описательной шкалы и градации	1. CLASS CHEATER	2. CLASS NOT A FRAUDST...	Безусл. вероятн.	Среднее	Средн. квадр. откл.
1	TIME-1/3-{0.0000000, 57597.3333333}	38.821	27.570	27.590	33.196	7.955
2	TIME-2/3-{57597.3333333, 115194.6666667}	33.537	29.467	29.474	31.502	2.878
3	TIME-3/3-{115194.6666667, 172792.0000000}	27.642	42.963	42.936	35.302	10.833
4	V1-1/3-{-56.4075096, -36.7866964}		0.003	0.003	0.001	0.002
5	V1-2/3-{-36.7866964, -17.1658832}	7.520	0.070	0.083	3.795	5.268
6	V1-3/3-{-17.1658832, 2.4549300}	92.480	99.927	99.914	96.203	5.266
7	V2-1/3-{-72.7157276, -41.1245754}		0.003	0.003	0.002	0.002
8	V2-2/3-{-41.1245754, -9.5334232}		0.264	0.264	0.132	0.187
9	V2-3/3-{-9.5334232, 22.0577290}	99.797	99.732	99.732	99.765	0.046
10	V3-1/3-{-48.3255894, -29.0895401}	1.626	0.002	0.005	0.814	1.149
11	V3-2/3-{-29.0895401, -9.8534909}	20.122	0.061	0.096	10.092	14.185
12	V3-3/3-{-9.8534909, 9.3825584}	78.252	99.937	99.899	89.094	15.333
13	V4-1/3-{-5.6831712, 1.8363339}	17.886	92.864	92.734	55.375	53.017
14	V4-2/3-{-1.8363339, 9.3558389}	76.016	7.122	7.241	41.569	48.716
15	V4-3/3-{-9.3558389, 16.8753440}	6.098	0.014	0.025	3.056	4.302
16	V5-1/3-{-113.7433067, -64.2283158}		0.000	0.000	0.000	0.000
17	V5-2/3-{-64.2283158, -14.7133250}	5.488	0.024	0.033	2.756	3.864
18	V5-3/3-{-14.7133250, 34.8016659}	94.512	99.976	99.966	97.244	3.863
19	V6-1/3-{-26.1605059, 6.9935379}	100.000	99.943	99.943	99.972	0.040
20	V6-2/3-{-6.9935379, 40.1475817}		0.056	0.056	0.028	0.040
21	V6-3/3-{-40.1475817, 73.3016255}		0.000	0.000	0.000	0.000
22	V7-1/3-{-43.5572416, 11.1583369}	100.000	99.951	99.952	99.976	0.034
23	V7-2/3-{-11.1583369, 65.8739154}		0.048	0.048	0.024	0.034
24	V7-3/3-{-65.8739154, 120.5894939}					
25	V8-1/3-{-73.2167185, -42.1420762}		0.001	0.001	0.001	0.001
26	V8-2/3-{-42.1420762, -11.0674339}	3.862	0.152	0.158	2.007	2.623

5.5. Модель: "4. INF1 - частный критерий: количество знаний по А.Харкевичу; вероятности из PRC1"

Код признака	Наименование описательной шкалы и градации	1. CLASS CHEATER	2. CLASS NOT A FRAUDSTER	Сумма	Среднее	Средн. квадр. откл.
1	TIME-1/3-{0.0000000, 57597.3333333}	0.021	0.000	0.021	0.011	0.015
2	TIME-2/3-{57597.3333333, 115194.6666667}	0.008	0.000	0.008	0.004	0.006
3	TIME-3/3-{115194.6666667, 172792.0000000}	-0.028	0.000	-0.028	-0.014	0.020
4	V1-1/3-{-56.4075096, -36.7866964}		0.000	0.000	0.000	0.000
5	V1-2/3-{-36.7866964, -17.1658832}	0.282	-0.011	0.272	0.136	0.207
6	V1-3/3-{-17.1658832, 2.4549300}	-0.005	0.000	-0.005	-0.002	0.003
7	V2-1/3-{-72.7157276, -41.1245754}		0.000	0.000	0.000	0.000
8	V2-2/3-{-41.1245754, -9.5334232}		0.000	0.000	0.000	0.000
9	V2-3/3-{-9.5334232, 22.0577290}	0.000		0.000	0.000	0.000
10	V3-1/3-{-48.3255894, -29.0895401}	0.368	-0.060	0.308	0.154	0.303
11	V3-2/3-{-29.0895401, -9.8534909}	0.335	-0.028	0.307	0.153	0.257
12	V3-3/3-{-9.8534909, 9.3825584}	-0.015	0.000	-0.015	-0.008	0.011
13	V4-1/3-{-5.6831712, 1.8363339}	-0.103	0.000	-0.103	-0.052	0.073
14	V4-2/3-{-1.8363339, 9.3558389}	0.147	-0.001	0.146	0.073	0.105
15	V4-3/3-{-9.3558389, 16.8753440}	0.345	-0.035	0.311	0.155	0.269
16	V5-1/3-{-113.7433067, -64.2283158}		0.000	0.000	0.000	0.000
17	V5-2/3-{-64.2283158, -14.7133250}	0.320	-0.021	0.299	0.150	0.241
18	V5-3/3-{-14.7133250, 34.8016659}	-0.004	0.000	-0.004	-0.002	0.002
19	V6-1/3-{-26.1605059, 6.9935379}	0.000		0.000	0.000	0.000
20	V6-2/3-{-6.9935379, 40.1475817}		0.000	0.000	0.000	0.000
21	V6-3/3-{-40.1475817, 73.3016255}		0.000	0.000	0.000	0.000
22	V7-1/3-{-43.5572416, 11.1583369}	0.000		0.000	0.000	0.000
23	V7-2/3-{-11.1583369, 65.8739154}		0.000	0.000	0.000	0.000
24	V7-3/3-{-65.8739154, 120.5894939}					
25	V8-1/3-{-73.2167185, -42.1420762}		0.000	0.000	0.000	0.000
26	V8-2/3-{-42.1420762, -11.0674339}	0.200	-0.003	0.198	0.099	0.143

Код признака	Наименование описательной шкалы и градации	1. CLASS CHEATER	2. CLASS NOT A FRAUDSTER	Сумма	Среднее	Средн. квадр. откл.
1	TIME-1/3-{0.0000000, 57597.3333333}	55.276	-55.278	-0.002	-0.001	78.174
2	TIME-2/3-{57597.3333333, 115194.6666667}	20.007	-20.008	0.000	0.000	28.295
3	TIME-3/3-{115194.6666667, 172792.0000000}	-75.217	75.218	0.001	0.000	106.374
4	V1-1/3-{-56.4075096, -36.7866964}	-0.014	0.014			0.020
5	V1-2/3-{-36.7866964, -17.1658832}	36.592	-36.592	0.000	0.000	51.749
6	V1-3/3-{-17.1658832, 2.4549300}	-36.511	36.505	-0.005	-0.003	51.630
7	V2-1/3-{-72.7157276, -41.1245754}	-0.016	0.016	0.000	0.000	0.022
8	V2-2/3-{-41.1245754, -9.5334232}	-1.299	1.299	0.000	0.000	1.837
9	V2-3/3-{-9.5334232, 22.0577290}	0.382	-0.380	0.002	0.001	0.539
10	V3-1/3-{-48.3255894, -29.0895401}	7.978	-7.978	0.000	0.000	11.282
11	V3-2/3-{-29.0895401, -9.8534909}	98.528	-98.528	0.000	0.000	139.340
12	V3-3/3-{-9.8534909, 9.3825584}	-106.438	106.446	0.008	0.004	150.532
13	V4-1/3-{-5.6831712, 1.8363339}	-368.192	368.203	0.011	0.006	520.710
14	V4-2/3-{-1.8363339, 9.3558389}	338.379	-338.379	-0.001	0.000	478.540
15	V4-3/3-{-9.3558389, 16.8753440}	29.879	-29.879	0.000	0.000	42.255
16	V5-1/3-{-113.7433067, -64.2283158}	-0.002	0.002			0.002
17	V5-2/3-{-64.2283158, -14.7133250}	26.838	-26.838			37.954
18	V5-3/3-{-14.7133250, 34.8016659}	-26.768	26.764	-0.004	-0.002	37.853
19	V6-1/3-{-26.1605059, 6.9935379}	0.344	-0.346	-0.001	-0.001	0.488
20	V6-2/3-{-6.9935379, 40.1475817}	-0.276	0.276	0.000	0.000	0.391
21	V6-3/3-{-40.1475817, 73.3016255}	-0.002	0.002			0.002
22	V7-1/3-{-43.5572416, 11.1583369}	0.304	-0.295	0.010	0.005	0.424
23	V7-2/3-{-11.1583369, 65.8739154}	-0.237	0.237	0.000	0.000	0.335
24	V7-3/3-{-65.8739154, 120.5894939}					
25	V8-1/3-{-73.2167185, -42.1420762}	-0.007	0.007			0.010
26	V8-2/3-{-42.1420762, -11.0674339}	18.221	-18.221	0.000	0.000	25.768

Рисунок 13

2.2. Верификация моделей

В системе «Эйдос» есть возможность оценки достоверности модели с помощью различных критериев достоверности в трех аспектах:

1. Рассмотрение частотных распределений истинных и ложных положительных и отрицательных решений в различных моделях.
2. Сравнение достоверности положительных и отрицательных решений в модели с вероятностью случайного угадывания.
3. Оценка информационной мощности модели.

2.2.1. Критерии достоверности моделей

Для оценки достоверности созданных статистических и системно-когнитивных моделей служит режим 3.4 (рисунок 14).

Оценка достоверности моделей в системе «Эйдос» осуществляется в соответствии с F-мерой Ван Ризбергена и двумя ее обобщениями (L1 и L2), разработанными проф. Е.В.Луценко [10].

Из рисунка 14 мы видим, что по F-критерию Ван Ризбергена и L1-критерию проф.Е.В.Луценко достоверность модели INF3 по обоим интегральным критериям составляет 0,999 при максимуме 1,000, что является очень хорошим результатом, учитывая сложность задачи и существенную несбалансированность исходных данных.

Пояснения к данному режиму даны в его хелпе (рисунок 15) и в работе [10].

3.4. Обобщенная форма по достоям моделей при различит. Текущая модель: "INF1"

Наименование модели и частного критерия	Интегральный критерий	о ложно-отрицательных (FN)	Число ложно-отрицательных (FN)	Точность модели	Полнота модели	F-мера Ван Ризбергена	Сумма модулей уровней сходства истинно-положительных (ST)	Сумма модулей уровней сходства истинно-отрицательных (STN)	Сумма модулей уровней сходства ложно-положительных (SFP)	Сумма модулей уровней сходства ложно-отрицательных (SFN)	S-Точность модели	S-Полнота модели	L1-мера проф. Е.В.Луценко
1. ABS - частный критерий: количество строк совпавших "хлеб"	Корреляция abs частот с обобщенной моделью	17955	2	0,909	1,000	0,952	147776,607	0,021	14463,098	0,011	0,921	1,000	0,959
1. ABS - частный критерий: количество строк совпавших "хлеб"	Сумма abs частот по признаку	17955	2	0,909	1,000	0,952	171683,702	0,021	418,502	0,011	0,998	1,000	0,959
2. PR1 - частный критерий: усл. вероятность i-го признака сред.	Корреляция усл.отн.частот с обобщенной моделью	17955	2	0,909	1,000	0,952	147776,602	0,021	14463,093	0,011	0,921	1,000	0,959
2. PR1 - частный критерий: усл. вероятность i-го признака сред.	Сумма усл.отн.частот по признаку	17955	2	0,909	1,000	0,952	172082,791	0,021	14847,788	0,011	0,922	1,000	0,959
3. PR2 - частный критерий: условная вероятность i-го признака.	Корреляция усл.отн.частот с обобщенной моделью	17955	2	0,909	1,000	0,952	147776,596	0,021	14463,090	0,011	0,921	1,000	0,959
3. PR2 - частный критерий: условная вероятность i-го признака.	Сумма усл.отн.частот по признаку	17955	2	0,909	1,000	0,952	172082,728	0,021	14845,883	0,011	0,922	1,000	0,959
4. INF1 - частный критерий: количество знаний по А.Харизвану: в.	Семантический резонанс: зна.	391	404	0,998	0,998	0,999	99812,468	15310,111	157,268	228,970	0,998	0,998	0,998
4. INF1 - частный критерий: количество знаний по А.Харизвану: в.	Сумма знаний	452	36159	0,997	0,999	0,999	151,255	1404,940	22,422	18,423	0,970	0,990	0,998
5. INF2 - частный критерий: количество знаний по А.Харизвану: в.	Семантический резонанс: зна.	391	404	0,998	0,998	0,999	99820,829	15310,152	157,264	225,973	0,998	0,998	0,998
5. INF2 - частный критерий: количество знаний по А.Харизвану: в.	Сумма знаний	452	36157	0,997	0,999	0,999	151,255	1405,051	22,600	18,626	0,970	0,990	0,998
6. INF3 - частный критерий: "Минимальная разница между фактом и прогнозом"	Семантический резонанс: зна.	205	231	0,998	0,999	0,999	94280,653	11162,810	85,276	91,691	0,999	0,999	0,999
6. INF3 - частный критерий: "Минимальная разница между фактом и прогнозом"	Сумма знаний	205	231	0,998	0,999	0,999	141281,027	11162,665	85,277	91,686	0,999	0,999	0,999
7. INF4 - частный критерий: ROI (Return On Investment): вероисп.	Семантический резонанс: зна.	395	399	0,998	0,998	0,998	137549,694	13504,068	217,789	219,096	0,998	0,998	0,998
7. INF4 - частный критерий: ROI (Return On Investment): вероисп.	Сумма знаний	8246	35667	0,946	0,801	0,868	140,800	6,736	30,119	0,197	0,824	0,999	0,903
8. INF5 - частный критерий: ROI (Return On Investment): вероисп.	Семантический резонанс: зна.	395	399	0,998	0,998	0,998	137549,712	13504,082	217,789	219,096	0,998	0,998	0,998
8. INF5 - частный критерий: ROI (Return On Investment): вероисп.	Сумма знаний	8245	35632	0,946	0,802	0,868	140,800	6,746	30,111	0,197	0,824	0,999	0,903
9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей: вер.	Семантический резонанс: зна.	206	231	0,999	0,999	0,999	141280,041	11175,914	85,255	92,030	0,999	0,999	0,998
9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей: вер.	Сумма знаний	206	231	0,999	0,999	0,999	350,946	11214,365	11,369	75,254	0,969	0,823	0,830
10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей: ве.	Семантический резонанс: зна.	206	231	0,999	0,999	0,999	141297,231	11178,233	85,281	92,046	0,999	0,999	0,999
10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей: ве.	Сумма знаний	206	231	0,999	0,999	0,999	350,787	11218,222	11,347	75,316	0,969	0,823	0,830

Рисунок 14

Помощь по режимам: 3.4. 4.1.3.6. 4.1.3.7. 4.1.3.8. 4.1.3.10. Виды прогнозов и меры достоверности моделей в системе "Эйдос-X++"

ПОЛОЖИТЕЛЬНЫЙ ПСЕВДОПРОГНОЗ.
 Предполагим, модель дает такой прогноз, что выпадет все: и 1, и 2, и 3, и 4, и 5, и 6. Понятно, что из всего этого выпадет лишь что-то одно. В этом случае модель не предскажет, что не выпадет, но зато она обязательно предскажет, что выпадет. Однако при этом очень много объектов будет отнесено к классам, к которым они не относятся. Тогда вероятность истинно-положительных решений у модели будет 1/6, а вероятность ложно-положительных решений - 5/6. Ясно, что такой прогноз бесполезен, поэтому он и назван мной псевдопрогнозом.

ОТРИЦАТЕЛЬНЫЙ ПСЕВДОПРОГНОЗ.
 Представим себе, что мы выбрасываем кубик с 6 гранями, и модель предсказывает, что ничего не выпадет, т.е. не выпадет ни 1, ни 2, ни 3, ни 4, ни 5, ни 6, но что-то из этого, естественно, выпадет. Конечно, модель не предсказала, что выпадет, зато она очень хорошо предсказала, что не выпадет. Вероятность истинно-отрицательных решений у модели будет 5/6, а вероятность ложно-отрицательных решений - 1/6. Такой прогноз гораздо достовернее, чем положительный псевдопрогноз, но тоже бесполезен.

ИДЕАЛЬНЫЙ ПРОГНОЗ.
 Если в случае с кубиком мы прогнозируем, что выпадет, например 1, и соответственно прогнозируем, что не выпадет 2, 3, 4, 5, и 6, то это идеальный прогноз, имеющий, если он осуществляется, 100% достоверность идентификации и не идентификации. Идеальный прогноз, который полностью снимает неопределенность о будущем состоянии объекта прогнозирования, на практике удается получить крайне редко и обычно мы имеем дело с реальным прогнозом.

РЕАЛЬНЫЙ ПРОГНОЗ.
 На практике мы чаще всего сталкиваемся именно с этим видом прогноза. Реальный прогноз уменьшает неопределенность о будущем состоянии объекта прогнозирования, но не полностью, как идеальный прогноз, а оставляет некоторую неопределенность не снятой. Например, для игрального кубика делается такой прогноз: выпадет 1 или 2, и, соответственно, не выпадет 3, 4, 5 или 6. Понятно, что полностью на практике такой прогноз не может осуществляться, т.к. варианты выпадения кубика альтернативны, т.е. не может выпадать одновременно и 1, и 2. Поэтому у реального прогноза всегда будет определенная ошибка идентификации. Соответственно, если не осуществляется один или несколько из прогнозируемых вариантов, то возникнет и ошибка не идентификации, т.к. это не прогнозировалось моделью. Теперь представьте себе, что у Вас не 1 кубик и прогноз его поведения, а тысячи. Тогда можно посчитать средневзвешенные характеристики всех этих видов прогнозов.

Таким образом, если просуммировать число верно идентифицированных и не идентифицированных объектов и вычесть число ошибочно идентифицированных и не идентифицированных объектов, а затем разделить на число всех объектов то это и будет критерий качества модели (классификатора), учитывающий как ее способность верно относить объекты к классам, к которым они относятся, так и ее способность верно не относить объекты к тем классам, к которым они не относятся. Этот критерий предложен и реализован в системе "Эйдос" проф. Е.В.Луценко в 1994 году. Эта мера достоверности модели предполагает два варианта нормировки: $\{ -1, +1 \}$ и $\{ 0, 1 \}$.

$$L_a = \frac{TP + TN - FP - FN}{TP + TN + FP + FN} \quad \text{(нормировка: } \{ -1, +1 \} \text{)}$$

$$L_b = \frac{1 + (TP + TN - FP - FN) / (TP + TN + FP + FN)}{2} \quad \text{(нормировка: } \{ 0, 1 \} \text{)}$$

где количество: TP - истинно-положительных решений; TN - истинно-отрицательных решений; FP - ложно-положительных решений; FN - ложно-отрицательных решений;

Классическая F-мера достоверности моделей Ван Ризбергена (колонка выделена ярко-голубым фоном):
 $F\text{-мера} = 2 * \text{Precision} * \text{Recall} / (\text{Precision} + \text{Recall})$ - достоверность модели
 Precision = TP / (TP + FP) - точность модели;
 Recall = TP / (TP + FN) - полнота модели;

L1-мера проф. Е.В.Луценко - нечеткое мультиклассовое обобщение классической F-меры с учетом СММ уровней сходства (колонка выделена ярко-зеленым фоном):
 $L1\text{-мера} = 2 * (\text{SPrecision} * \text{SRecall}) / (\text{SPrecision} + \text{SRecall})$
 $\text{SPrecision} = \text{STP} / (\text{STP} + \text{SFP})$ - точность с учетом сумм уровней сходства;
 $\text{SRecall} = \text{STP} / (\text{STP} + \text{SFN})$ - полнота с учетом сумм уровней сходства;
 STP - Сумма модулей сходства истинно-положительных решений; STN - Сумма модулей сходства истинно-отрицательных решений;
 SFP - Сумма модулей сходства ложно-положительных решений; SFN - Сумма модулей сходства ложно-отрицательных решений.

L2-мера проф. Е.В.Луценко - нечеткое мультиклассовое обобщение классической F-меры с учетом СРЕДНИХ уровней сходства (колонка выделена желтым фоном):
 $L2\text{-мера} = 2 * (\text{APrecision} * \text{ARecall}) / (\text{APrecision} + \text{ARecall})$
 $\text{APrecision} = \text{ATP} / (\text{ATP} + \text{AFP})$ - точность с учетом средних уровней сходства;
 $\text{ARecall} = \text{ATP} / (\text{ATP} + \text{AFN})$ - полнота с учетом средних уровней сходства;
 ATP=STP/TP - Среднее модулей сходства истинно-положительных решений; AFN=SFN/FN - Среднее модулей сходства истинно-отрицательных решений;
 AFP=SFP/FP - Среднее модулей сходства ложно-положительных решений; AFN=SFN/FN - Среднее модулей сходства ложно-отрицательных решений.

Строки с максимальными значениями F-меры, L1-меры и L2-меры выделены фоном цвета, соответствующего колонке.

Из графиков частотных распределений истинно-положительных, истинно-отрицательных, ложно-положительных и ложно-отрицательных решений видно, что чем выше модуль уровня сходства, тем больше доля истинных решений. Это значит, что модуль уровня сходства является адекватной мерой степени истинности решения и степени уверенности системы в этом решении. Поэтому система "Эйдос" имеет адекватный критерий достоверности собственных решений, с помощью которого она может отфильтровать заведомо ложные решения.

Луценко Е.В. Инвариантное относительное обобщение нечеткого мультиклассового обобщения F-меры достоверности моделей Ван Ризбергена в АСК-анализе и системе "Эйдос" / Е.В. Луценко // Политетический сетевой и электронный научный журнал Кубанского государственного аграрного университета [Научный журнал КубГАУ] [Электронный ресурс]. - Краснодар: КубГАУ, 2017. - №02(126). С. 1 - 32. - IDA [article ID]: 1261702001. - Режим доступа: <http://ej.kubagro.ru/2017/02/pdf/01.pdf>, 2 у.п.л.

Рисунок 15

2.2.2. Частотные распределения истинных и ложных положительных и отрицательных решений в различных моделях

В режиме 3.4, кликнув по кнопке [TP,TN,FP,FN] и расположенным правее нее, получаем частотные распределения числа положительных и отрицательных истинных и ложных решений, их разностей и их нормированных разностей (рисунок 16):

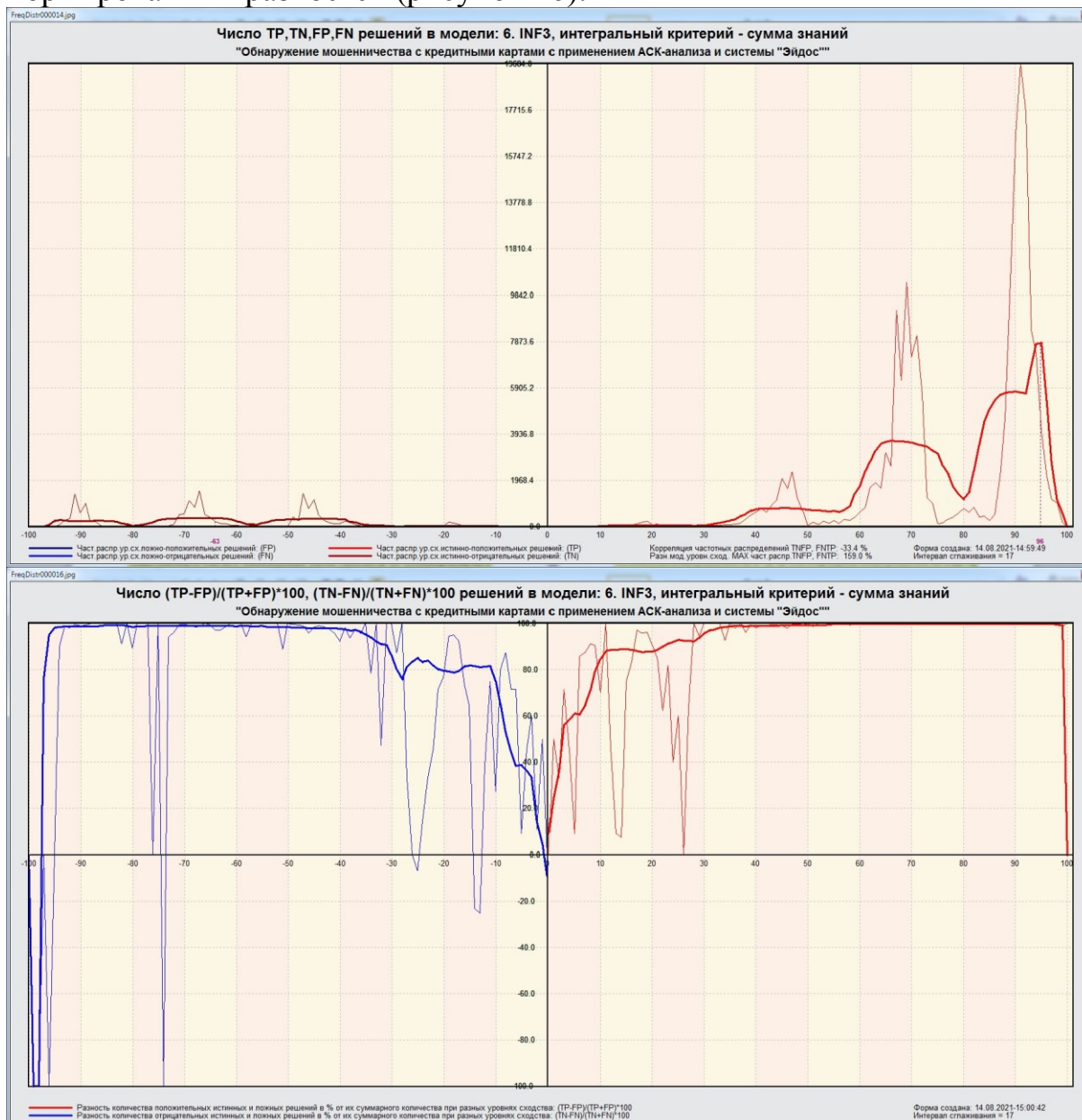


Рисунок 16

Главное, что мы видим из рисунка 16 (1-е изображение), это то, что истинных решений при всех уровнях сходства и различия, т.е. и при положительных, и при отрицательных решениях, всегда больше, чем ложных.

Из второго рисунка 11 видно, что при увеличении уровня сходства объекта с классом закономерно и монотонно растет и доля истинных решений среди всех решений, а доля ложных решений также закономерно уменьшается. Из этого можно обоснованно сделать очень важный вывод: **уровень сходства объекта с классом, т.е. значение интегрального критерия, является адекватной мерой степени истинности решения.** Этот фундаментальный вывод подтверждается на огромном количестве решенных в системе «Эйдос» задач из самых различных предметных областей.

Это означает, что *в системе «Эйдос» есть достоверный внутренний критерий степени истинности решений задач, предлагаемых системой на основе созданных в ней моделей. Таким образом, система «Эйдос» не просто идентифицирует, но и оценивает достоверность идентификации, не просто прогнозирует, но и оценивает достоверность прогнозирования, не просто предлагает решение, но и оценивает эффективность этого решения, и т.д.*

Таким образом, система Эйдос не только прогнозирует значения будущих параметров, но и адекватно оценивает достоверность их прогнозирования. Наличие в системе «Эйдос» внутреннего достоверного критерия достоверности прогнозирования позволяет прогнозировать наступление точки *бифуркации*, точки неопределенности. *В точках бифуркации резко уменьшается достоверность прогнозирования и возрастает разброс точечных прогнозов с различных позиций во времени. Фактически это означает, что можно либо достоверно прогнозировать, что произойдет, либо достоверно прогнозировать, что мы не можем достоверно прогнозировать, т.е. достоверно прогнозировать точку бифуркации.* Об этом есть в работе [10]: <http://lc.kubagro.ru/aidos/aidos02/7.4.htm>. В этой монографии 2002 года описаны результаты, полученные в 1994 году.

Все это и позволяет решать с использованием модели задачу оценки риска мошенничества при операциях с кредитными картами.

Если бы никакой зависимости риска мошенничества от различных параметров транзакций не существовало бы, т.е. выдвинутая нами гипотеза не подтверждалась бы, то приведенные на рисунках 16 зависимости не наблюдались бы.

Вместо них на 1-м изображении рисунка 16 число истинных и ложных решений практически бы совпадало и для положительных, и для отрицательных решений. В результате на 2-м изображении рисунка 16 мы бы видели прямую линию, близкую к нулю.

Однако все это наблюдается, что и подтверждает предложенную автором гипотезу.

2.2.3. Сравнение достоверности положительных и отрицательных решений в модели с вероятностью случайного угадывания

Рассмотрим экранную форму режима 4.1.3.7 (это более подробный вариант режима 3.4), представленную на рисунке 17.

Как видно из рисунка 17, использование модели наиболее высокой достоверности INF3 с интегральным критерием «Сумма знаний» позволяет повысить достоверность прогнозирования по сравнению со случайным угадыванием для положительных решений в 1,638 раза, для отрицательных решений в 232,277 раза, в среднем для всех решений в 116,957 раз.

Это очень хороший результат.

Наименование задачи	Наименование интегрального критерия	интегральный критерий	Средняя доля верных ответов	Результат при модели	Результат при модели с вероятностью случайного угадывания	Средняя достоверность	Доля верных ответов при модели	Доля верных ответов при модели с вероятностью случайного угадывания	Средняя достоверность для положительных решений	Средняя достоверность для отрицательных решений	Средняя достоверность для всех решений	Значит ли это, что модель лучше?	Значит ли это, что модель хуже?	Средняя достоверность при модели	Средняя достоверность при модели с вероятностью случайного угадывания	Средняя достоверность для всех решений	Дата запуска	Время запуска
1. ABS - частный критерий: количество верных ответов на вопрос.	Корреляция абсолютных частот с.о.	0.894	0.485	0.248	99.999	0.203	99.950	0.001	99.493	0.547	2.000	0.740	1.970	11.08.2021	22:24:01			
2. PRC1 - частный критерий: доля верности отгадки по признаку.	Корреляция доли частот с.о.	0.954	0.248	0.248	99.999	0.203	99.950	0.001	99.493	0.547	2.000	0.740	1.970	11.08.2021	22:24:01			
3. PRC1 - частный критерий: доля верности отгадки по признаку.	Средняя доля верных ответов	0.892	0.485	0.248	99.999	0.203	99.950	0.001	99.493	0.547	2.000	0.740	1.970	11.08.2021	22:24:01			
4. INF3 - частный критерий: сумма верности отгадки по признаку.	Корреляция доли частот с.о.	0.934	0.485	0.248	99.999	0.203	99.950	0.001	99.493	0.547	2.000	0.740	1.970	11.08.2021	22:24:01			
4. INF3 - частный критерий: сумма верности отгадки по признаку.	Средняя доля верных ответов	0.950	0.485	0.248	99.999	0.203	99.950	0.001	99.493	0.547	2.000	0.740	1.970	11.08.2021	22:24:01			
4. INF1 - частный критерий: количество знаний по А.Харкевичу, в.	Сумма знаний	0.000	0.000	0.000	79.862	82.050	7.703	20.138	99.493	0.547	1.638	335.840	168.728	11.08.2021	22:14:58			
9. INF2 - частный критерий: количество знаний по А.Харкевичу, в.	Сумма знаний	0.000	0.000	79.862	82.050	7.703	20.137	99.493	0.547	1.638	335.840	168.728	11.08.2021	22:16:57				
9. INF3 - частный критерий: количество знаний по А.Харкевичу, в.	Сумма знаний	0.000	0.000	79.862	82.050	7.703	20.137	99.493	0.547	1.638	335.840	168.728	11.08.2021	22:16:57				
9. INF3 - частный критерий: количество знаний по А.Харкевичу, в.	Интегральный критерий: сумма знаний	0.957	0.248	0.248	99.971	82.050	7.703	20.137	99.493	0.547	1.638	335.840	168.728	11.08.2021	22:16:57			
9. INF3 - частный критерий: количество знаний по А.Харкевичу, в.	Интегральный критерий: сумма знаний	0.957	0.248	0.248	99.971	82.050	7.703	20.137	99.493	0.547	1.638	335.840	168.728	11.08.2021	22:16:57			
7. INF4 - частный критерий: ROI (Return On Investment) верности.	Сумма знаний	0.000	0.000	0.000	80.136	82.038	7.715	19.564	99.493	0.547	1.724	335.839	165.702	11.08.2021	23:11:07			
8. INF3 - частный критерий: ROI (Return On Investment) верности.	Сумма знаний	0.748	0.544	0.222	99.778	24.147	75.605	0.222	99.493	0.547	1.242	81.029	44.435	11.08.2021	23:17:29			
8. INF3 - частный критерий: ROI (Return On Investment) верности.	Сумма знаний	0.000	0.000	0.000	80.155	82.038	7.715	19.564	99.493	0.547	1.724	335.839	165.702	11.08.2021	23:18:52			
9. INF6 - частный критерий: разность и безразмерности, вер.	Сумма знаний	0.000	0.000	0.000	80.136	82.038	7.715	19.564	99.493	0.547	1.724	335.839	165.702	11.08.2021	23:18:52			
9. INF6 - частный критерий: разность и безразмерности, вер.	Сумма знаний	0.000	0.000	0.000	80.136	82.038	7.715	19.564	99.493	0.547	1.724	335.839	165.702	11.08.2021	23:18:52			
9. INF6 - частный критерий: разность и безразмерности, вер.	Сумма знаний	0.000	0.000	0.000	80.136	82.038	7.715	19.564	99.493	0.547	1.724	335.839	165.702	11.08.2021	23:18:52			
10. INF7 - частный критерий: разность и безразмерности, вер.	Сумма знаний	0.000	0.000	0.000	80.136	82.038	7.715	19.564	99.493	0.547	1.724	335.839	165.702	11.08.2021	23:18:52			
10. INF7 - частный критерий: разность и безразмерности, вер.	Сумма знаний	0.000	0.000	0.000	80.136	82.038	7.715	19.564	99.493	0.547	1.724	335.839	165.702	11.08.2021	23:18:52			

Рисунок 17

2.2.4. Оценка информационно-мощности модели

Рассмотрим экранную форму на рисунке 18 (отображается в режиме 5.5), на которой приведен фрагмент модели INF1, содержащий **информационную мощность модели** (число в самой правой колонке в последней строке, на которой стоит курсор). В данном случае информационная мощность модели составляет 0,104 бита.

Код признака	Наименование описательной шкалы и градации	1. CLASS CHEATER	2. CLASS NOT A FRAUDSTER	Сумма	Среднее	Средн. квадрат. откл.
69	V22-3/3-(3.3576788, 10.5030901)	0.261	-0.007	0.254	0.127	0.190
70	V23-1/3-(4.48077352, -22.3623529)	0.000	0.000	0.000	0.000	0.000
71	V23-2/3-(22.3623529, 0.0830294)	-0.002	0.000	-0.002	-0.001	0.002
72	V23-3/3-(0.0830294, 22.5284117)	0.004	0.000	0.004	0.002	0.003
73	V24-1/3-(2.8366269, -0.3629016)	0.012	0.000	0.012	0.006	0.009
74	V24-2/3-(0.3629016, 2.1108238)	-0.005	0.000	-0.005	-0.002	0.003
75	V24-3/3-(2.1108238, 4.5845491)	0.000	0.000	0.000	0.000	0.000
76	V25-1/3-(10.2953971, -4.3570685)	0.221	-0.004	0.217	0.109	0.159
77	V25-2/3-(4.3570685, 1.5812601)	-0.002	0.000	-0.002	-0.001	0.001
78	V25-3/3-(1.5812601, 7.5195887)	0.147	-0.001	0.146	0.073	0.105
79	V26-1/3-(2.6045906, -0.5639185)	-0.019	0.000	-0.019	-0.010	0.013
80	V26-2/3-(0.5639185, 4.4767135)	0.002	0.000	0.002	0.001	0.001
81	V26-3/3-(1.4767135, 3.5173456)	0.016	0.000	0.016	0.008	0.011
82	V27-1/3-(2.5656793, -4.5063868)	0.201	-0.003	0.198	0.099	0.144
83	V27-2/3-(4.5063868, 13.5529056)	-0.001	0.000	-0.001	0.000	0.001
84	V27-3/3-(13.5529056, 31.6121981)	0.000	0.000	0.000	0.000	0.000
85	V28-1/3-(15.4300839, 0.9958800)	-0.001	0.000	-0.001	-0.001	0.001
86	V28-2/3-(0.9958800, 17.4218439)	0.120	-0.001	0.120	0.060	0.085
87	V28-3/3-(17.4218439, 33.8478078)	0.000	0.000	0.000	0.000	0.000
88	AMOUNT-1/3-(0.0000000, 8563.7200000)	0.000	0.000	0.000	0.000	0.000
89	AMOUNT-2/3-(8563.7200000, 17127.4400000)	0.000	0.000	0.000	0.000	0.000
90	AMOUNT-3/3-(17127.4400000, 25591.1600000)	0.000	0.000	0.000	0.000	0.000
	Сумма	6.910	-1.013	5.897		
	Среднее	0.077	-0.111		0.033	
	Среднеквадратичное отклонение	0.145	0.029			0.104

Рисунок 18

2.2.5. Выводы

Все это на взгляд автора представляет собой совершенно ясное и чрезвычайно убедительное подтверждение предложенной гипотезы, подтверждающее, что значения характеристик транзакций тесно связаны с величиной риска мошенничества и могут быть успешно применены для количественной оценки этого риска.

И все же, чтобы дать более убедительную оценку полученным характеристикам модели, нужно с чем-то его сравнивать, т.е. иметь базу для сравнения. В качестве базы сравнения для сравнения характеристик созданной модели предлагается использовать модель, аналогичную по размерностям всех баз данных и по их заполнению числами, но в которой числа получены с помощью цифрового генератора псевдослучайных чисел. Это сравнение реальной и случайной моделей мы проведем в следующем разделе.

2.3. Оценка долей влияния характеристик транзакций в системе детерминации степени риска мошенничества

2.3.1. Синтез случайной модели на основе реальной

В любом наборе данных есть полезный сигнал и есть шум. Соответственно есть соотношение сигнал/шум. Чем выше это соотношение, тем выше *качество набора данных*. Для оценки качества реального набора данных необходимо сгенерировать случайный набор данных, близкий по размерам (количество и тип шкал и градаций, обучающая выборка) и сравнить модели, созданные на основе этих наборов. Различие между этими моделями связано с наличием паттернов, закономерностей, полезного сигнала в реальном наборе данных.

Итак, предположим, что исходные данные зашумлены, т.е. в них содержится не только полезный сигнал, отражающий влияние значений характеристик транзакций на степень риска мошенничества, но и шум, причем отношение сигнал/шум неизвестно, и хотелось бы его определить.

С целью оценки силы влияния полезного сигнала и шума на полученные статистические и системно-когнитивные модели предлагается сравнить модели и результаты их использования, полученные на реальных данных, с моделями той же размерности всех баз данных, но заполненными случайными данными.

Для создания случайной модели, совпадающей с моделью, построенной на реальных данных по размерностям и степени заполнения всех баз данных, в диспетчере приложений (режим 1.3) сделаем текущей реальную модель и установим лабораторную работу 2.01 (рисунки 19, 20 и 21).

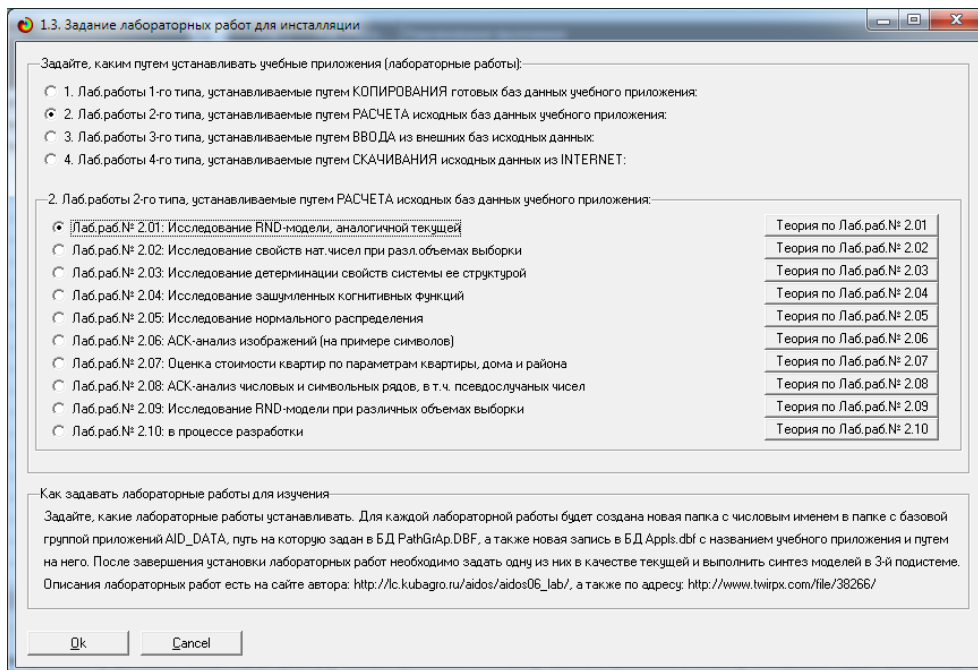


Рисунок 19

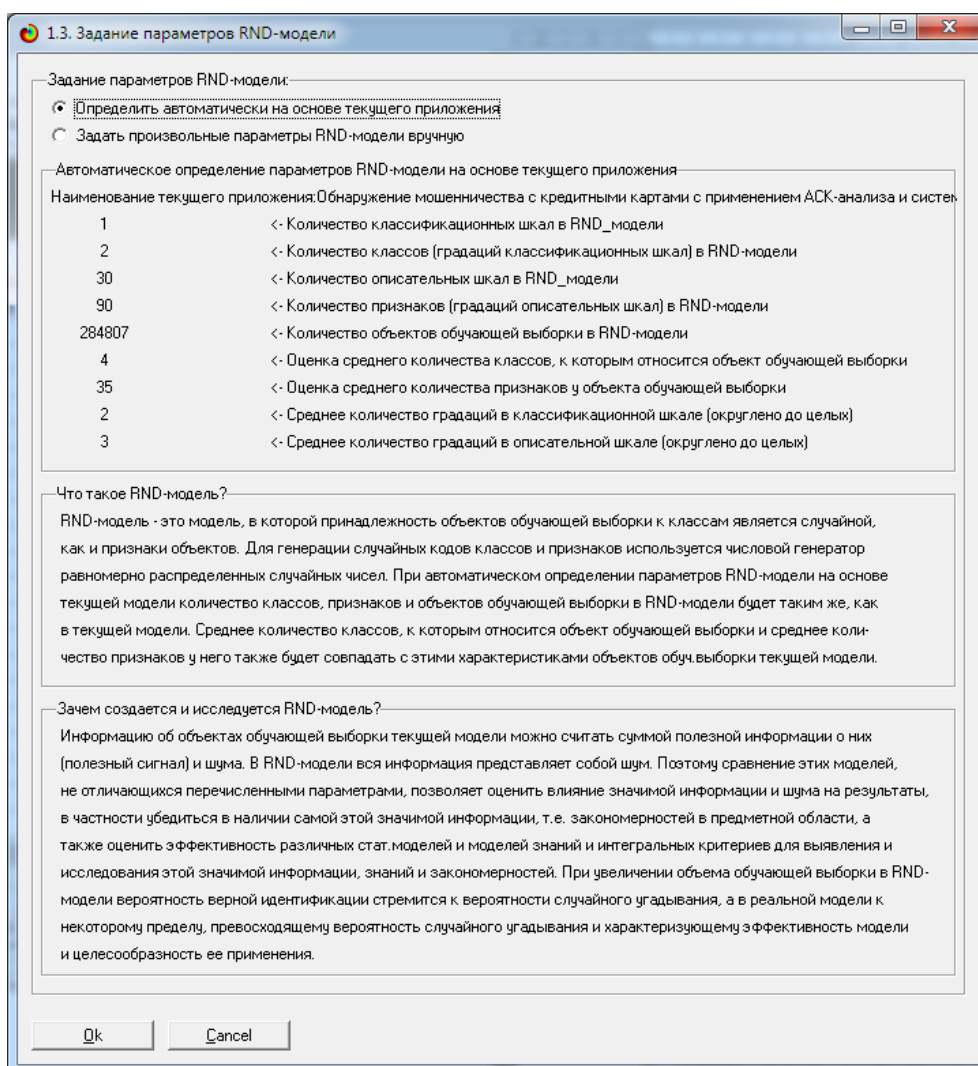


Рисунок 20

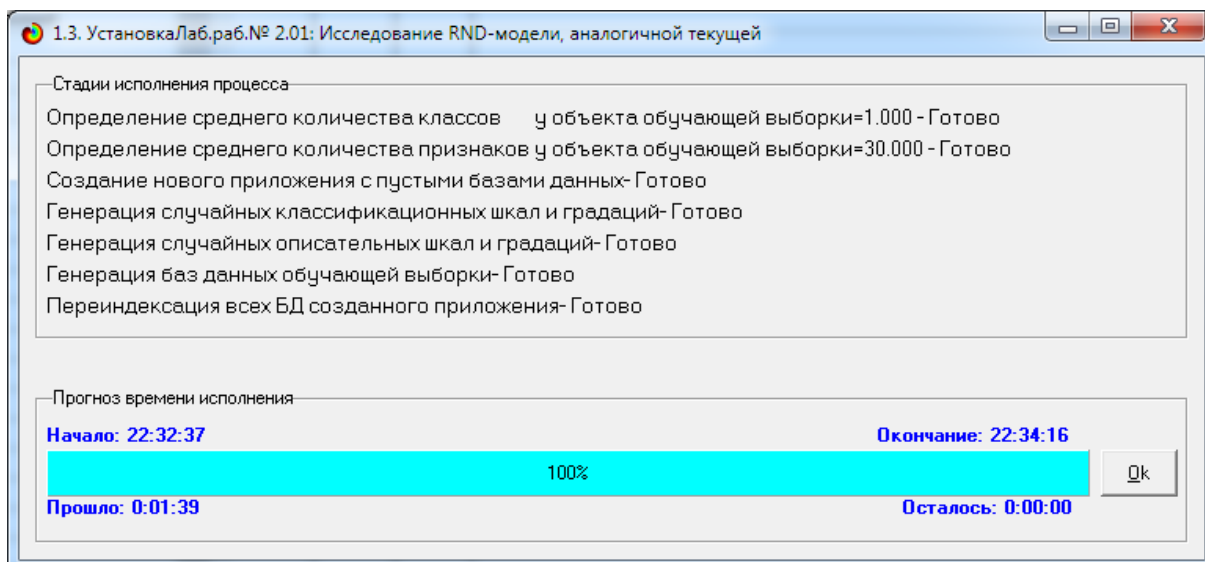


Рисунок 21

В результате в диспетчере приложений (режим 1.3) появляется еще одно приложение (рисунок 22) со случайной моделью, совпадающее с реальной моделью по размерностям всех баз данных и их заполнению данными, но заполненной псевдослучайными данными.

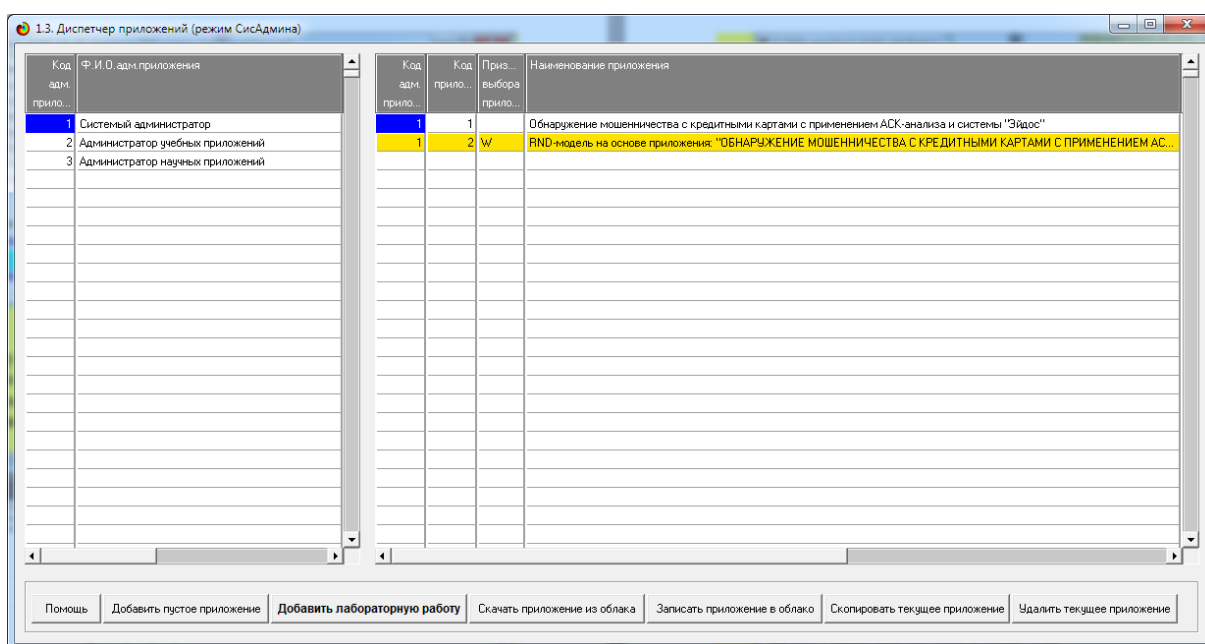


Рисунок 22

Затем запускаем синтез и верификацию моделей (режим 3.5) с теми же параметрами, что и в реальной модели (рисунки 23 и 24):

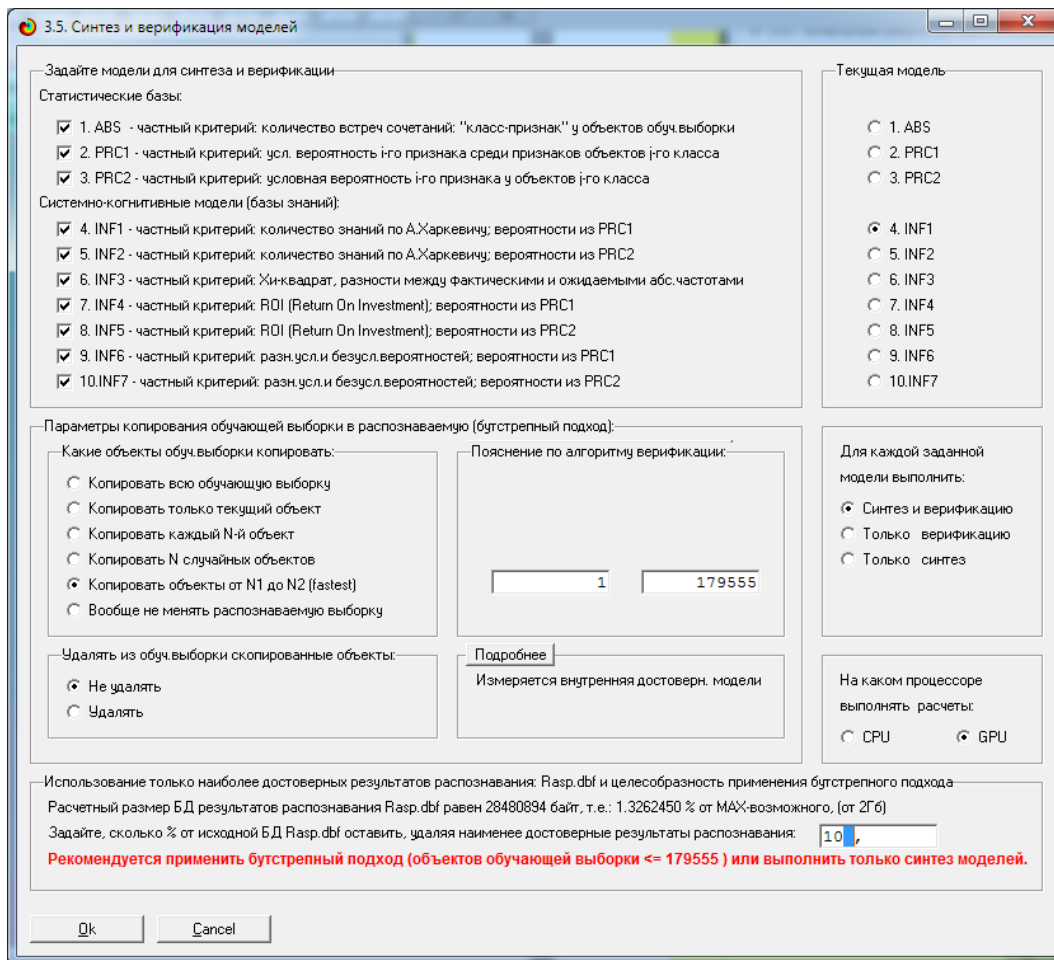


Рисунок 23

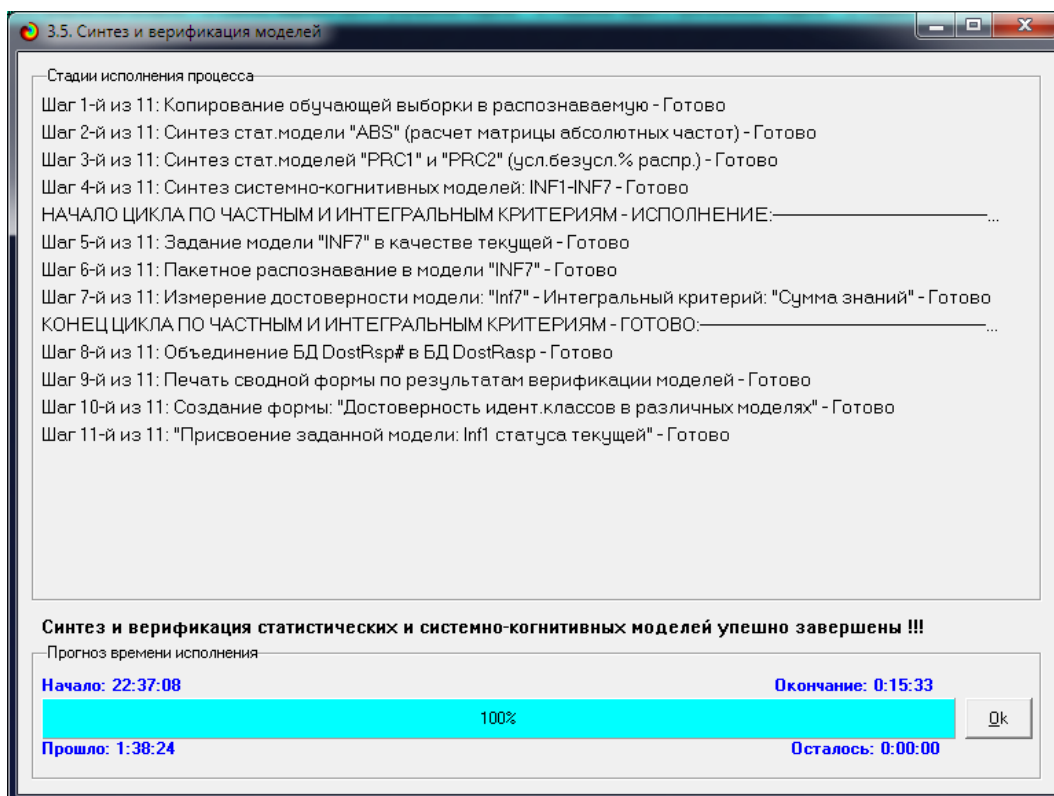


Рисунок 24

Для генерации псевдослучайных чисел в системе «Эйдос» используется улучшенный проф. Е.В.Луценко генератор шума Фон Неймана⁷. Слабой стороной классического генератора Фон Неймана является наличие повторов псевдослучайных последовательностей, зависящих от первого числа и основания системы счисления. Поэтому автором предложено при запуске генератора Фон Неймана для определения первого числа использовать внешний источник энтропии в качестве которого применить значение компьютерного таймера, практически никогда не повторяющееся. Кроме того, предложено вместо возведения числа в квадрат использовать другие математические функции, теоретически дающие *иррациональные* значения.

В системе «Эйдос» одновременно применяется несколько различных *методов подавления шума* в исходных данных и в основанных на них моделях. Примененный метод формирования матрицы абсолютных частот является методом выделения полезного сигнала из шума. Используется неметрический интегральный критерий (межвекторное расстояние), который представляет собой фильтр, подавляющий белый шум. Применяется разделение классов на типичные и нетипичные части, оценка значимости шкал и градаций и удаление из модели незначимых шкал, удаление несформированных классов, удаление из дальнейшего анализа результатов распознавания с низкой достоверностью и другие методы.

Существует три аспекта исследования влияния характеристик транзакций на объект моделирования, в качестве которого выступают операции с кредитными картами:

1. Сравнение частотных распределений истинных и ложных положительных и отрицательных решений в реальной и случайной моделях.
2. Сравнение достоверностей положительных и отрицательных решений в реальной и случайной моделях.
3. Сравнение информационной мощности реальной и случайной моделей и оценка отношения полезного сигнала к шуму.

Рассмотрим все эти три аспекта.

2.3.2. Сравнение частотных распределений истинных и ложных положительных и отрицательных решений в реальной и случайной моделях

После этого в режиме 3.4 смотрим и сравниваем достоверность системно-когнитивной модели INF3 и строим графики частотных распределений количества истинных и ложных, положительных и отрицательных решений в зависимости от уровней сходства для реальной и случайной моделей (рисунки 14 и 25).

⁷ См., например: <https://math.bobrodobro.ru/7894>

Наименование модели и частного критерия	Интегральный критерий	Всего логических объектов выборки	Число истинно-положительных решений (TP)	Число истинно-отрицательных решений (TN)	Число ложно-положительных решений (FP)	Число ложно-отрицательных решений (FN)	Точность модели	Полнота модели	F-мера Ван Ризбергера	Сходство уровней сходства истинно-положительных решений (ST)	Сходство уровней сходства истинно-отрицательных решений (SF)	Сходство уровней сходства ложно-положительных решений (SP)	Сходство уровней ложно-отрицательных решений (SF)
1.ABS - частный критерий: количество встреч сонетаний "клас..."	Корреляция абс. частот с обр...	179555	91542	8977	8979	88013	0,911	0,510	0,654	16977,118	1681,201	1655,548	159
1.ABS - частный критерий: количество встреч сонетаний "клас..."	Сумма абс. частот по признаку...	179555	179555		17956		0,909	1,000	0,952	178907,582		17890,764	159
2.PRC1 - частный критерий: усл. вероятность иго признака сред...	Корреляция усл.отн. частот с о...	179555	91540	8978	8978	88015	0,911	0,510	0,654	16977,126	1681,202	1655,550	159
2.PRC1 - частный критерий: усл. вероятность иго признака сред...	Сумма усл.отн. частот по признаку...	179555	179555		17956		0,909	1,000	0,952	178856,325		17885,691	159
3.PRC2 - частный критерий: условная вероятность иго признака...	Корреляция усл.отн. частот с о...	179555	91547	8979	8977	88008	0,911	0,510	0,654	16977,126	1681,202	1655,549	159
3.PRC2 - частный критерий: условная вероятность иго признака...	Сумма усл.отн. частот по признаку...	179555	179555		17956		0,909	1,000	0,952	178856,324		17885,691	159
4.INF1 - частный критерий: количество знаний по А.Харкевичу; в...	Семантический резонанс: зна...	179555	91177	9157	8799	88378	0,912	0,508	0,652	17366,128	1757,334	1658,478	163
4.INF1 - частный критерий: количество знаний по А.Харкевичу; в...	Сумма знаний	179555	90557	9208	8748	88998	0,912	0,504	0,649	16771,785	1730,456	1598,749	163
5.INF2 - частный критерий: количество знаний по А.Харкевичу; в...	Семантический резонанс: зна...	179555	91177	9157	8799	88378	0,912	0,508	0,652	17366,085	1757,330	1658,474	163
5.INF2 - частный критерий: количество знаний по А.Харкевичу; в...	Сумма знаний	179555	90557	9208	8748	88998	0,912	0,504	0,649	16771,897	1730,455	1598,760	163
6.INF3 - частный критерий: Хинквадрат, разности между фактик...	Семантический резонанс: зна...	179555	91148	9157	8799	88407	0,912	0,508	0,652	17357,268	1754,287	1655,742	163
6.INF3 - частный критерий: Хинквадрат, разности между фактик...	Сумма знаний	179555	91147	9157	8799	88408	0,912	0,508	0,652	16987,020	1715,622	1619,505	161
7.INF4 - частный критерий: ROI (Return On Investment); вероя...	Семантический резонанс: зна...	179555	91134	9155	8801	88421	0,912	0,508	0,652	17364,266	1754,973	1656,444	163
7.INF4 - частный критерий: ROI (Return On Investment); вероя...	Сумма знаний	179555	91143	9155	8801	88412	0,912	0,508	0,652	16996,303	1716,538	1620,426	161
8.INF5 - частный критерий: ROI (Return On Investment); вероя...	Семантический резонанс: зна...	179555	91135	9155	8801	88420	0,912	0,508	0,652	17364,224	1754,969	1656,440	163
8.INF5 - частный критерий: ROI (Return On Investment); вероя...	Сумма знаний	179555	91143	9155	8801	88412	0,912	0,508	0,652	16996,416	1716,537	1620,437	161
9.INF6 - частный критерий: разн.усли. без усл. вероятностей; вер...	Семантический резонанс: зна...	179555	91148	9157	8799	88407	0,912	0,508	0,652	17357,254	1754,285	1655,740	163
9.INF6 - частный критерий: разн.усли. без усл. вероятностей; вер...	Сумма знаний	179555	91148	9157	8799	88407	0,912	0,508	0,652	16992,570	1716,172	1620,036	161
10.INF7 - частный критерий: разн.усли. без усл. вероятностей; ве...	Семантический резонанс: зна...	179555	91148	9157	8799	88407	0,912	0,508	0,652	17357,257	1754,286	1655,741	163
10.INF7 - частный критерий: разн.усли. без усл. вероятностей; ве...	Сумма знаний	179555	91148	9157	8799	88407	0,912	0,508	0,652	16992,570	1716,172	1620,036	161

Рисунок 25

Из рисунка 14 мы видим, что по F-критерию Ван Ризбергера и L1-критерию проф.Е.В.Луценко достоверность модели INF3 по обоим интегральным критериям составляет 0,999 при максимуме 1,000, что является очень хорошим результатом, учитывая сложность задачи и существенную несбалансированность исходных данных.

Для случайной модели с теми же размерностями, что и у реальной, достоверность модели INF3 по обоим интегральным критериям составляет по F-критерию Ван Ризбергера 0,652, а по L1-критерию проф.Е.В.Луценко 0,656, т.е. существенно ниже, чем в реальной модели, как и следовало ожидать.

Это различие связано с наличием в реальной модели ярко выраженных закономерных взаимосвязей между значениями характеристик транзакций и степенью риска мошенничества и полным отсутствием этих взаимосвязей в случайной модели.

Рассмотрим графики частотных распределений количества истинных и ложных, положительных и отрицательных решений в зависимости от уровня сходства для реальной и случайной моделей (рисунок 26).

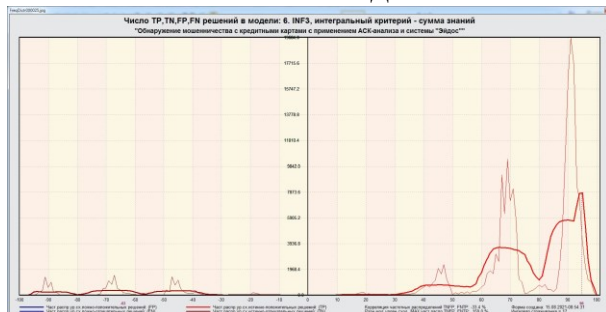
Из сравнения графиков частотных распределений истинных и ложных, положительных и отрицательных решений различных уровней сходства для реальной и случайной моделей, приведенных на рисунке 26, видно, что **степень различия частотных распределений истинных и ложных решений в реальной модели значительно превосходит степень их различия в случайной модели.**

Частотные распределения количества истинных и ложных, положительных и отрицательных решений в зависимости от уровней сходства в случайной модели образуют два распределения, очень сходных с нормальными распределениями, максимумы которых полностью совпадают друг с другом и с началом отсчета (0) по шкале уровня сходства. В реальной модели этот сдвиг значительный, что численно отражено в значениях коэффициентов корреляции между этими

распределениями и расстоянием между максимумами этих распределений. Это различие в реальной модели и обеспечивает адекватную оценку риска мошенничества по значениям характеристик транзакций.

Реальная модель

Случайная модель



Корреляция = -33,4%
Разность ур.сх.максимумов = 159,0%

Корреляция = 99,4%
Разность ур.сх.максимумов = 0,0%

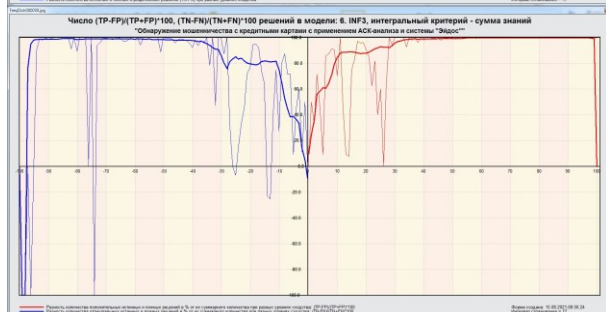
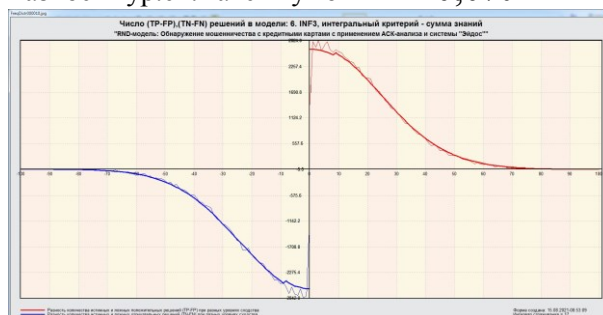
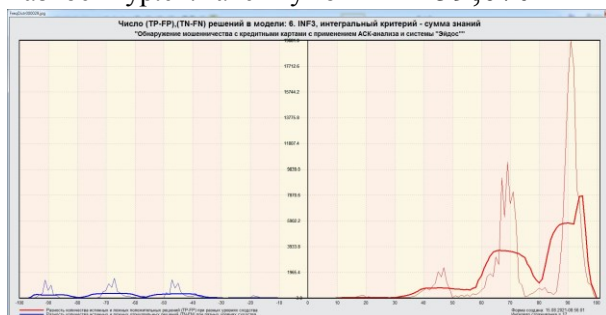


Рисунок 26

По 3-му графику для случайной модели мы видим, что чем больше количество наблюдений с определенным уровнем сходства, тем меньше различия в доле истинных и ложных положительных и отрицательных решений. По-видимому, это проявление действия закона больших чисел.

По 1-му и 2-му графикам для реальной модели мы видим, что они практически не отличаются друг от друга. Это связано с тем, что доля ложных решений в реальной модели близка к нулю.

В этих графиках наблюдается ярко выраженная периодическая зависимость количества истинных положительных и отрицательных решений от модуля уровня сходства с периодом около 20%. У автора нет гипотезы, объясняющей наличие этой периодической зависимости.

2.3.3. Сравнение достоверностей положительных и отрицательных решений в реальной и случайной моделях по сравнению со случайным угадыванием

Рассмотрим экранную форму режима 4.1.3.7 (более подробный вариант режима 3.4), представленную для реальной модели на рисунках 17 и 27, а для случайной на рисунке 28:

Наименование модели и частного критерия	Наименование интегрального критерия	Эффект. модели при идентиф.: =C13/C17	Эффект. модели при неидент.: =C14/C18	Средняя эффект. модели: =(C19+C20)/2	Дата получения результата	Время получения результ...
1. ABS - частный критерий: количество встреч сочетаний: "клас...	Корреляция абс частот с обр...	2.000	0.740	1.370	11.08.2021	22:24:01
1. ABS - частный критерий: количество встреч сочетаний: "клас...	Сумма абс частот по признак...	2.000	0.740	1.000	11.08.2021	22:25:25
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Корреляция усл.отн. частот с о...	2.000	0.740	1.370	11.08.2021	22:31:42
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Сумма усл.отн. частот по приз...	2.000	0.740	1.000	11.08.2021	22:33:07
3. PRC2 - частный критерий: условная вероятность i-го признака...	Корреляция усл.отн. частот с о...	2.000	0.740	1.000	11.08.2021	22:39:10
3. PRC2 - частный критерий: условная вероятность i-го признака...	Сумма усл.отн. частот по приз...	2.000	0.740	1.000	11.08.2021	22:40:31
4. INF1 - частный критерий: количество знаний по А.Харкевичу; в...	Семантический резонанс зна...	1.225	91.727	46.476	11.08.2021	22:46:34
4. INF1 - частный критерий: количество знаний по А.Харкевичу; в...	Сумма знаний	1.618	335.840	168.729	11.08.2021	22:47:58
5. INF2 - частный критерий: количество знаний по А.Харкевичу; в...	Семантический резонанс зна...	1.225	91.727	46.476	11.08.2021	22:54:14
5. INF2 - частный критерий: количество знаний по А.Харкевичу; в...	Сумма знаний	1.618	335.840	168.729	11.08.2021	22:55:37
6. INF3 - частный критерий: Хи-квадрат, разности между фактич...	Семантический резонанс зна...	1.638	232.277	116.957	11.08.2021	23:02:09
6. INF3 - частный критерий: Хи-квадрат, разности между фактич...	Сумма знаний	1.638	232.277	116.957	11.08.2021	23:03:34
7. INF4 - частный критерий: ROI (Return On Investment); вероятн...	Семантический резонанс зна...	1.242	88.029	44.635	11.08.2021	23:09:43
7. INF4 - частный критерий: ROI (Return On Investment); вероятн...	Сумма знаний	1.724	335.839	168.782	11.08.2021	23:11:07
8. INF5 - частный критерий: ROI (Return On Investment); вероятн...	Семантический резонанс зна...	1.242	88.029	44.635	11.08.2021	23:17:29
8. INF5 - частный критерий: ROI (Return On Investment); вероятн...	Сумма знаний	1.724	335.839	168.782	11.08.2021	23:18:52
9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей; вер...	Семантический резонанс зна...	1.638	232.277	116.957	11.08.2021	23:25:20
9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей; вер...	Сумма знаний	1.638	232.277	116.957	11.08.2021	23:26:43
10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей; ве...	Семантический резонанс зна...	1.638	232.277	116.957	11.08.2021	23:33:09
10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей; ве...	Сумма знаний	1.638	232.277	116.957	11.08.2021	23:34:37

Рисунок 27

Наименование модели и частного критерия	Наименование интегрального критерия	Эффект. модели при идентиф.: =C13/C17	Эффект. модели при неидент.: =C14/C18	Средняя эффект. модели: =(C19+C20)/2	Дата получения результата	Время получения результ...
1. ABS - частный критерий: количество встреч сочетаний: "клас...	Корреляция абс частот с обр...	1.020	0.100	0.560	14.08.2021	22:44:01
1. ABS - частный критерий: количество встреч сочетаний: "клас...	Сумма абс частот по признак...	2.000	0.100	1.000	14.08.2021	22:46:09
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Корреляция усл.отн. частот с о...	1.020	0.100	0.560	14.08.2021	22:52:56
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Сумма усл.отн. частот по приз...	2.000	0.100	1.000	14.08.2021	22:54:58
3. PRC2 - частный критерий: условная вероятность i-го признака...	Корреляция усл.отн. частот с о...	1.020	0.100	0.560	14.08.2021	23:01:06
3. PRC2 - частный критерий: условная вероятность i-го признака...	Сумма усл.отн. частот по приз...	2.000	0.100	1.000	14.08.2021	23:03:19
4. INF1 - частный критерий: количество знаний по А.Харкевичу; в...	Семантический резонанс зна...	1.016	0.102	0.559	14.08.2021	23:10:51
4. INF1 - частный критерий: количество знаний по А.Харкевичу; в...	Сумма знаний	1.009	0.103	0.556	14.08.2021	23:13:01
5. INF2 - частный критерий: количество знаний по А.Харкевичу; в...	Семантический резонанс зна...	1.016	0.102	0.559	14.08.2021	23:20:51
5. INF2 - частный критерий: количество знаний по А.Харкевичу; в...	Сумма знаний	1.009	0.103	0.556	14.08.2021	23:22:51
6. INF3 - частный критерий: Хи-квадрат, разности между фактич...	Семантический резонанс зна...	1.015	0.102	0.559	14.08.2021	23:29:48
6. INF3 - частный критерий: Хи-квадрат, разности между фактич...	Сумма знаний	1.019	0.102	0.559	14.08.2021	23:31:48
7. INF4 - частный критерий: ROI (Return On Investment); вероятн...	Семантический резонанс зна...	1.015	0.102	0.559	14.08.2021	23:39:33
7. INF4 - частный критерий: ROI (Return On Investment); вероятн...	Сумма знаний	1.015	0.102	0.559	14.08.2021	23:41:40
8. INF5 - частный критерий: ROI (Return On Investment); вероятн...	Семантический резонанс зна...	1.015	0.102	0.559	14.08.2021	23:49:55
8. INF5 - частный критерий: ROI (Return On Investment); вероятн...	Сумма знаний	1.015	0.102	0.559	14.08.2021	23:52:01
9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей; вер...	Семантический резонанс зна...	1.015	0.102	0.559	14.08.2021	23:59:41
9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей; вер...	Сумма знаний	1.015	0.102	0.559	15.08.2021	00:01:41
10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей; ве...	Семантический резонанс зна...	1.015	0.102	0.559	15.08.2021	00:09:21
10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей; ве...	Сумма знаний	1.015	0.102	0.559	15.08.2021	00:11:27

Рисунок 28

Как видно из рисунков 27 и 28, использование модели наиболее высокой достоверности INF3 с интегральным критерием «Сумма знаний»

позволяет очень существенно повысить достоверность прогнозирования по сравнению со случайным угадыванием:

	Реальная модель	Случайная модель	Разность
– для положительных решений (в раз)	1,638	1,015	0,623
– для отрицательных решений (в раз)	232,277	0,102	232,125
– в среднем для всех решений (в раз)	116,957	0,559	116,398

Из приведенных данных по достоверности оценки риска мошенничества в реальной и случайной моделях по сравнению со случайным угадыванием можно сделать обоснованный вывод о том, что использование информации о значениях характеристик в реальной модели позволяет повысить достоверность прогнозирования степени риска мошенничества **в сотни раз**, по сравнению со случайным угадыванием и случайной моделью.

Это говорит о том, что информация о значениях характеристик транзакций является весьма существенной для оценки степени риска мошенничества, что и подтверждает правильность предложенной гипотезы.

2.3.4. Сравнение информационной мощности реальной и случайной моделей и оценка отношения полезного сигнала к шуму

Итак, будем считать, что исходные данные зашумлены, т.е. в них содержится как полезный сигнал, отражающий влияние характеристик транзакций на степень риска мошенничества, так и шум, отражающий влияние скрытых параметров, неучтенных в модели, причем отношение сигнал/шум неизвестно и необходимо его оценить.

Понятие корня информационной мощности модели мы определили ранее в таблице 10.

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

где:

$$\bar{I} = \frac{1}{W \cdot M} \sum_{j=1}^W \sum_{i=1}^M I_{ij} \tag{1}$$

Информационная мощность модели – это квадрат среднеквадратичного значения, т.е. дисперсия амплитуды частных критериев в данной модели. Но поскольку сам термин прямо указывает на информацию, то предпочтительно использовать системно-когнитивные модели INF1 и INF2.

Введем следующие обозначения:

- информационная мощность реальной модели: P_{real} ;
- информационная мощность случайной модели: P_{noise} ;
- среднеквадратичное значение амплитуды полезного сигнала: A_{signal} ;
- среднеквадратичное значение амплитуды шума: A_{noise} .

Будем считать, что:

1. Случайная модель состоит только из шума, т.е. в ней полностью отсутствует влияние характеристик транзакций:

$$P_{noise} = (A_{noise})^2 \dots\dots \dots (2)$$

2. Мощность реальной модели является суммой мощностей полезного сигнала и шума:

$$P_{real} = (A_{real})^2 = P_{signal} + P_{noise} \dots\dots\dots (3)$$

3. Отношение сигнал/шум равно SNR (*англ. signal-to-noise ratio*), отношению мощности полезного сигнала к мощности шума:

$$SNR = \frac{P_{signal}}{P_{noise}} = \frac{(A_{signal})^2}{(A_{noise})^2},$$

где: $\dots\dots\dots (4)$

$$P_{signal} = (A_{signal})^2$$

Из выражения (3) получаем:

$$P_{signal} = P_{real} - P_{noise} \dots\dots\dots (5)$$

Подставляем (5) в (4) и получаем (6):

$$SNR = \frac{P_{real} - P_{noise}}{P_{noise}} = \frac{(A_{real})^2 - (A_{noise})^2}{(A_{noise})^2} \dots\dots\dots (6)$$

Или окончательно (7):

$$SNR = \frac{P_{real}}{P_{noise}} - 1 = \frac{(A_{real})^2}{(A_{noise})^2} - 1 \dots\dots\dots (7)$$

Нельзя не заметить высокое формальное и смысловое сходство выражения (7) с частным критерием знаний ROI (таблица 8).

Подставим в выражение (7) информационную мощность реальной и случайной моделей из таблиц 11 и 12 и, соответственно, и учтем выражения (1) – (7).

Таблица 11 – Системно-когнитивная модель INF1 на реальных данных

KOD PR	NAME	CLS1	CLS2	SUMMA	SREDN	DISP
1	TIME-1/3-{0.000000, 57597.333333}	0.0214056	-0.0000441	0.0213614	0.0106807	0.0151672
2	TIME-2/3-{57597.333333, 115194.666667}	0.0080988	-0.0000149	0.0080838	0.0040419	0.0057373
3	TIME-3/3-{115194.666667, 172792.000000}	-0.0275820	0.0000386	-0.0275435	-0.0137717	0.0195308
4	V1-1/3-{-56.4075096, -36.7866964}	0.0000000	0.001083	0.001083	0.0000542	0.0000766
5	V1-2/3-{-36.7866964, -17.1658832}	0.2824620	-0.0105757	0.2718861	0.1359431	0.2072090
6	V1-3/3-{-17.1658832, 2.4549300}	-0.0048359	0.0000081	-0.0048279	-0.0024139	0.0034253
7	V2-1/3-{-72.7157276, -41.1245754}	0.0000000	0.001083	0.001083	0.0000542	0.0000766
8	V2-2/3-{-41.1245754, -9.5334232}	0.0000000	0.001083	0.001083	0.0000542	0.0000766
9	V2-3/3-{-9.5334232, 22.0577290}	0.0000488	0.0000000	0.0000487	0.0000244	0.0000346
10	V3-1/3-{-48.3255894, -29.0895401}	0.3681349	-0.0597579	0.3083770	0.1541885	0.3025660
11	V3-2/3-{-29.0895401, -9.8534909}	0.3350009	-0.0281118	0.3068890	0.1534445	0.2567596
12	V3-3/3-{-9.8534909, 9.3825584}	-0.0152932	0.0000235	-0.0152697	-0.0076348	0.0108306
13	V4-1/3-{-5.6831712, 1.8363339}	-0.1031013	0.0000874	-0.1030138	-0.0515069	0.0729655
14	V4-2/3-{-1.8363339, 9.3558389}	0.1473185	-0.0010382	0.1462802	0.0731401	0.1049042
15	V4-3/3-{-9.3558389, 16.8753440}	0.3454675	-0.0349536	0.3105138	0.1552569	0.2689984
16	V5-1/3-{-113.7433067, -64.2283158}	0.0000000	0.001083	0.001083	0.0000542	0.0000766
17	V5-2/3-{-64.2283158, -14.7133250}	0.3203960	-0.0211062	0.2992897	0.1496448	0.2414786
18	V5-3/3-{-14.7133250, 34.8016659}	-0.0035066	0.0000059	-0.0035007	-0.0017503	0.0024838
19	V6-1/3-{-26.1605059, 6.9935379}	0.0000438	0.0000000	0.0000438	0.0000219	0.0000311
20	V6-2/3-{-6.9935379, 40.1475817}	0.0000000	0.001083	0.001083	0.0000542	0.0000766
21	V6-3/3-{-40.1475817, 73.3016255}	0.0000000	0.001083	0.001083	0.0000542	0.0000766
22	V7-1/3-{-43.5572416, -11.1583369}	0.0000388	0.0000000	0.0000387	0.0000194	0.0000275
23	V7-2/3-{-11.1583369, 65.8739154}	0.0000000	0.001083	0.001083	0.0000542	0.0000766
24	V7-3/3-{-65.8739154, 120.5894939}	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
25	V8-1/3-{-73.2187185, -42.1420762}	0.0000000	0.001083	0.001083	0.0000542	0.0000766
26	V8-2/3-{-42.1420762, -11.0674339}	0.2001279	-0.0025883	0.1975395	0.0987698	0.1433421
27	V8-3/3-{-11.0674339, 20.0072084}	-0.0024912	0.0000042	-0.0024870	-0.0012434	0.0017646
28	V9-1/3-{-13.4340663, -3.7577127}	0.3832773	-0.0957984	0.2874788	0.1437394	0.3387577
29	V9-2/3-{-3.7577127, 5.9186410}	-0.0198098	0.0000294	-0.0197804	-0.0098902	0.0140285
30	V9-3/3-{-5.9186410, 15.5949946}	0.0000000	0.001083	0.001083	0.0000542	0.0000766
31	V10-1/3-{-24.5882624, -8.4771296}	0.3823999	-0.0927195	0.2896802	0.1448401	0.3359602
32	V10-2/3-{-8.4771296, 7.6340033}	-0.0143990	0.0000223	-0.0143768	-0.0071883	0.0101975
33	V10-3/3-{-7.6340033, 23.7451361}	0.0000000	0.001083	0.001083	0.0000542	0.0000766
34	V11-1/3-{-4.7974735, 0.8079887}	-0.1144440	0.0000909	-0.1143531	-0.0571765	0.0809884
35	V11-2/3-{-0.8079887, 6.4134510}	0.0718430	-0.0002332	0.0716098	0.0358049	0.0509656
36	V11-3/3-{-6.4134510, 12.0189132}	0.3819890	-0.0913425	0.2906464	0.1453232	0.3346960
37	V12-1/3-{-18.6837146, -9.8396790}	0.3866107	-0.1096149	0.2769957	0.1384978	0.3508846
38	V12-2/3-{-9.8396790, -0.9956435}	0.1064174	-0.0004859	0.1059314	0.0529657	0.0755921
39	V12-3/3-{-0.9956435, 7.8483921}	-0.1083565	0.0000891	-0.1082674	-0.0541336	0.0766827
40	V13-1/3-{-5.7918812, -1.4856265}	0.0361343	-0.0000845	0.0360497	0.0180248	0.0256107
41	V13-2/3-{-1.4856265, 2.8206283}	-0.0033630	0.0000057	-0.0033573	-0.0016786	0.0023821
42	V13-3/3-{-2.8206283, 7.1268830}	0.0000000	0.001083	0.001083	0.0000542	0.0000766
43	V14-1/3-{-19.2143255, -9.3006283}	0.3866648	-0.1098733	0.2767915	0.1383957	0.3511055
44	V14-2/3-{-9.3006283, 0.6130689}	-0.0083403	0.0000135	-0.0083268	-0.0041633	0.0059071
45	V14-3/3-{-0.6130689, 10.5267661}	-0.1262598	0.0000939	-0.1261659	-0.0630829	0.0893456
46	V15-1/3-{-4.4989447, -0.000493}	0.0065243	-0.0000118	0.0065124	0.0032562	0.0046218
47	V15-2/3-{-0.000493, 4.4188462}	-0.0061680	0.0000102	-0.0061578	-0.0030788	0.0043687
48	V15-3/3-{-4.4188462, 8.8777416}	0.0000000	0.001083	0.001083	0.0000542	0.0000766
49	V16-1/3-{-14.1298545, -3.6481992}	0.3675403	-0.0588180	0.3087222	0.1543611	0.3014810
50	V16-2/3-{-3.6481992, 6.8334562}	-0.0423211	0.0000532	-0.0422678	-0.0211339	0.0299632
51	V16-3/3-{-6.8334562, 17.3151115}	0.0000000	0.001083	0.001083	0.0000542	0.0000766
52	V17-1/3-{-25.1627994, -13.6906908}	0.3829868	-0.0947572	0.2882295	0.1441147	0.3378161
53	V17-2/3-{-13.6906908, -2.2185823}	0.3868720	-0.0577887	0.3090832	0.1545416	0.3002805
54	V17-3/3-{-2.2185823, 9.2535263}	-0.0777268	0.0000770	-0.0776498	-0.0388248	0.0550157
55	V18-1/3-{-9.4987459, -4.6521409}	0.3870465	-0.1117328	0.2753137	0.1376568	0.3528903
56	V18-2/3-{-4.6521409, 0.1944642}	-0.0090547	0.0000146	-0.0090401	-0.0045200	0.0064130
57	V18-3/3-{-0.1944642, 5.0410692}	-0.0354058	0.0000468	-0.0353590	-0.0176794	0.0250688
58	V19-1/3-{-7.2135274, -2.9450278}	0.1059795	-0.0004817	0.1054977	0.0527488	0.0752795
59	V19-2/3-{-2.9450278, 13.234718}	-0.0202783	0.0000300	-0.0202483	-0.0101241	0.0143602
60	V19-3/3-{-13.234718, 5.5919714}	0.1179659	-0.0006069	0.1173589	0.0586795	0.0838437
61	V20-1/3-{-54.4977205, -23.1915123}	0.0000000	0.001083	0.001083	0.0000542	0.0000766
62	V20-2/3-{-23.1915123, 8.1146960}	-0.0002216	0.0000004	-0.0002212	-0.0001106	0.0001570
63	V20-3/3-{-8.1146960, 39.4209042}	0.1480519	-0.0010519	0.1470000	0.0735000	0.1054324
64	V21-1/3-{-34.8303821, -14.1526417}	0.2936731	-0.0129010	0.2807720	0.1403860	0.2167806
65	V21-2/3-{-14.1526417, 6.5250988}	-0.0014452	0.0000025	-0.0014428	-0.0007213	0.0010237
66	V21-3/3-{-6.5250988, 27.2028392}	0.1595901	-0.0012891	0.1583009	0.0791504	0.1137589
67	V22-1/3-{-10.9331437, -3.7877324}	0.2012021	-0.0026360	0.1985661	0.0992830	0.1441354
68	V22-2/3-{-3.7877324, 3.3576788}	-0.0018827	0.0000032	-0.0018795	-0.0009397	0.0013336
69	V22-3/3-{-3.3576788, 10.5030901}	0.2608896	-0.0072711	0.2536184	0.1268092	0.1896184
70	V23-1/3-{-44.8077352, -22.3623529}	0.0000000	0.001083	0.001083	0.0000542	0.0000766
71	V23-2/3-{-22.3623529, 0.0830294}	-0.0022011	0.0000037	-0.0021974	-0.0010987	0.0015592
72	V23-3/3-{-0.0830294, 22.5284117}	0.0041149	-0.0000073	0.0041075	0.0020538	0.0029149
73	V24-1/3-{-2.8366269, -0.3629016}	0.0123611	-0.0000235	0.0123374	0.0061687	0.0087573
74	V24-2/3-{-0.3629016, 2.1108238}	-0.0045607	0.0000076	-0.0045531	-0.0022765	0.0032303
75	V24-3/3-{-2.1108238, 4.5845491}	0.0000000	0.001083	0.001083	0.0000542	0.0000766
76	V25-1/3-{-10.2953971, -4.3570685}	0.2210426	-0.0036899	0.2173526	0.1086763	0.1589100
77	V25-2/3-{-4.3570685, 1.5812601}	-0.0016389	0.0000028	-0.0016361	-0.0008180	0.0011609
78	V25-3/3-{-1.5812601, 7.5195887}	0.1474798	-0.0010412	0.1464384	0.0732192	0.1050203
79	V26-1/3-{-2.6045506, -0.5639185}	-0.0190450	0.0000284	-0.0190166	-0.0095082	0.0134870
80	V26-2/3-{-0.5639185, 1.4767135}	0.0016283	-0.0000028	0.0016255	0.0008127	0.0011534
81	V26-3/3-{-1.4767135, 3.5173456}	0.0157872	-0.0000310	0.0157561	0.0078780	0.0111852
82	V27-1/3-{-22.5656793, -4.5063868}	0.2007562	-0.0026161	0.1981400	0.0990700	0.1438060
83	V27-2/3-{-4.5063868, 13.5529056}	-0.0009770	0.0000017	-0.0009753	-0.0004876	0.0008921
84	V27-3/3-{-13.5529056, 31.6121981}	0.0000000	0.001083	0.001083	0.0000542	0.0000766
85	V28-1/3-{-15.4300839, 0.9958800}	-0.0013136	0.0000022	-0.0013113	-0.0006556	0.0009305
86	V28-2/3-{-0.9958800, 17.4218439}	0.1202689	-0.0006338	0.1196331	0.0598165	0.0854898
87	V28-3/3-{-17.4218439, 33.8478078}	0.0000000	0.001083	0.001083	0.0000542	0.0000766
88	AMOUNT-1/3-{-0.0000000, 8563.7200000}	0.0000106	0.0000000	0.0000106	0.0000053	0.0000075
89	AMOUNT-2/3-{-8563.7200000, 17127.4400000}	0.0000000	0.001083	0.001083	0.0000542	0.0000766
90	AMOUNT-3/3-{-17127.4400000, 25691.1600000}	0.0000000	0.001083	0.001083	0.0000542	0.0000766
	Сумма	6.9096241	-1.0127755	5.8968486	0.0000000	0.0000000
	Среднее	0.0767736	-0.0112530	0.0000000	0.0327603	0.0000000
	Среднеквадратичное отклонение	0,1445500	0,0287318	0,0000000	0,0000000	0,1039203

Таблица 12 – Системно-когнитивная модель INF1 на случайных данных

KOD_PR	NAME	CLS1	CLS2	SUMMA	SREDN	DISP
1	Описательная шкала 1/30-Градация описательной шкалы 1/3	-0,0002206	0,0002198	-0,0000008	-0,0000004	0,0003115
2	Описательная шкала 1/30-Градация описательной шкалы 2/3	-0,0000010	0,0000011	0,0000000	0,0000000	0,0000015
3	Описательная шкала 1/30-Градация описательной шкалы 3/3	-0,0000602	0,0000602	0,0000000	0,0000000	0,0000853
4	Описательная шкала 2/30-Градация описательной шкалы 1/3	-0,0001867	0,0001862	-0,0000006	-0,0000002	0,0002637
5	Описательная шкала 2/30-Градация описательной шкалы 2/3	-0,0000155	0,0000155	0,0000000	0,0000000	0,0000220
6	Описательная шкала 2/30-Градация описательной шкалы 3/3	-0,0003958	0,0003931	-0,0000026	-0,0000013	0,0005579
7	Описательная шкала 3/30-Градация описательной шкалы 1/3	0,0000121	-0,0000120	0,0000000	0,0000000	0,0000171
8	Описательная шкала 3/30-Градация описательной шкалы 2/3	-0,0002929	0,0002915	-0,0000014	-0,0000007	0,0004133
9	Описательная шкала 3/30-Градация описательной шкалы 3/3	-0,0000036	0,0000037	0,0000000	0,0000000	0,0000053
10	Описательная шкала 4/30-Градация описательной шкалы 1/3	-0,0002935	0,0002920	-0,0000015	-0,0000007	0,0004141
11	Описательная шкала 4/30-Градация описательной шкалы 2/3	-0,0000741	0,0000741	0,0000000	0,0000000	0,0001049
12	Описательная шкала 4/30-Градация описательной шкалы 3/3	0,0003112	-0,0003125	-0,0000013	-0,0000006	0,0004411
13	Описательная шкала 5/30-Градация описательной шкалы 1/3	0,0001661	-0,0001664	-0,0000002	-0,0000001	0,0002352
14	Описательная шкала 5/30-Градация описательной шкалы 2/3	0,0001026	-0,0001026	0,0000000	0,0000000	0,0001452
15	Описательная шкала 5/30-Градация описательной шкалы 3/3	-0,0001061	0,0001060	-0,0000001	0,0000000	0,0001501
16	Описательная шкала 6/30-Градация описательной шкалы 1/3	0,0002299	-0,0002305	-0,0000006	-0,0000002	0,0003256
17	Описательная шкала 6/30-Градация описательной шкалы 2/3	0,0001058	-0,0001058	0,0000000	0,0000000	0,0001497
18	Описательная шкала 6/30-Градация описательной шкалы 3/3	0,0000028	-0,0000027	0,0000000	0,0000000	0,0000040
19	Описательная шкала 7/30-Градация описательной шкалы 1/3	0,0001289	-0,0001290	-0,0000001	0,0000000	0,0001824
20	Описательная шкала 7/30-Градация описательной шкалы 2/3	0,0002118	-0,0002122	-0,0000005	-0,0000002	0,0002999
21	Описательная шкала 7/30-Градация описательной шкалы 3/3	-0,0003654	0,0003632	-0,0000022	-0,0000011	0,0005153
22	Описательная шкала 8/30-Градация описательной шкалы 1/3	-0,0001322	0,0001319	-0,0000003	-0,0000001	0,0001868
23	Описательная шкала 8/30-Градация описательной шкалы 2/3	-0,0000404	0,0000405	0,0000000	0,0000000	0,0000573
24	Описательная шкала 8/30-Градация описательной шкалы 3/3	0,0003096	-0,0003109	-0,0000012	-0,0000006	0,0004388
25	Описательная шкала 9/30-Градация описательной шкалы 1/3	0,0004855	-0,0004889	-0,0000034	-0,0000016	0,0006891
26	Описательная шкала 9/30-Градация описательной шкалы 2/3	-0,0002021	0,0002193	-0,0000008	-0,0000004	0,0003108
27	Описательная шкала 9/30-Градация описательной шкалы 3/3	0,0001717	-0,0001719	-0,0000002	-0,0000001	0,0002430
28	Описательная шкала 10/30-Градация описательной шкалы 1/3	-0,0000108	0,0000109	0,0000000	0,0000000	0,0000154
29	Описательная шкала 10/30-Градация описательной шкалы 2/3	0,0001057	-0,0001057	0,0000000	0,0000000	0,0001496
30	Описательная шкала 10/30-Градация описательной шкалы 3/3	-0,0000777	0,0000777	0,0000000	0,0000000	0,0001099
31	Описательная шкала 11/30-Градация описательной шкалы 1/3	0,0001387	-0,0001389	-0,0000001	0,0000000	0,0001964
32	Описательная шкала 11/30-Градация описательной шкалы 2/3	0,0000128	-0,0000127	0,0000000	0,0000000	0,0000180
33	Описательная шкала 11/30-Градация описательной шкалы 3/3	-0,0001812	0,0001807	-0,0000005	-0,0000002	0,0002560
34	Описательная шкала 12/30-Градация описательной шкалы 1/3	0,0000780	-0,0000779	0,0000000	0,0000000	0,0001103
35	Описательная шкала 12/30-Градация описательной шкалы 2/3	0,0000035	-0,0000034	0,0000000	0,0000000	0,0000050
36	Описательная шкала 12/30-Градация описательной шкалы 3/3	-0,0004426	0,0004393	-0,0000033	-0,0000016	0,0006237
37	Описательная шкала 13/30-Градация описательной шкалы 1/3	-0,0000725	0,0000724	0,0000000	0,0000000	0,0001025
38	Описательная шкала 13/30-Градация описательной шкалы 2/3	-0,0002157	0,0002149	-0,0000008	-0,0000003	0,0003045
39	Описательная шкала 13/30-Градация описательной шкалы 3/3	-0,0000393	0,0000393	0,0000000	0,0000000	0,0000557
40	Описательная шкала 14/30-Градация описательной шкалы 1/3	-0,0002152	0,0002145	-0,0000008	-0,0000003	0,0003039
41	Описательная шкала 14/30-Градация описательной шкалы 2/3	0,0000372	-0,0000371	0,0000000	0,0000000	0,0000525
42	Описательная шкала 14/30-Градация описательной шкалы 3/3	0,0001076	-0,0001076	0,0000000	0,0000000	0,0001522
43	Описательная шкала 15/30-Градация описательной шкалы 1/3	-0,0001114	0,0001112	-0,0000002	0,0000000	0,0001575
44	Описательная шкала 15/30-Градация описательной шкалы 2/3	0,0002082	-0,0002087	-0,0000005	-0,0000002	0,0002949
45	Описательная шкала 15/30-Градация описательной шкалы 3/3	0,0001989	-0,0001993	-0,0000004	-0,0000002	0,0002816
46	Описательная шкала 16/30-Градация описательной шкалы 1/3	-0,0000488	0,0000488	0,0000000	0,0000000	0,0000691
47	Описательная шкала 16/30-Градация описательной шкалы 2/3	-0,0001536	0,0001532	-0,0000004	-0,0000001	0,0002170
48	Описательная шкала 16/30-Градация описательной шкалы 3/3	0,0000069	-0,0000068	0,0000000	0,0000000	0,0000098
49	Описательная шкала 17/30-Градация описательной шкалы 1/3	-0,0001045	0,0001044	-0,0000001	0,0000000	0,0001478
50	Описательная шкала 17/30-Градация описательной шкалы 2/3	-0,0000361	0,0000362	0,0000000	0,0000000	0,0000512
51	Описательная шкала 17/30-Градация описательной шкалы 3/3	-0,0002215	0,0002207	-0,0000008	-0,0000004	0,0003128
52	Описательная шкала 18/30-Градация описательной шкалы 1/3	-0,0000514	0,0000514	0,0000000	0,0000000	0,0000728
53	Описательная шкала 18/30-Градация описательной шкалы 2/3	0,0003941	-0,0003963	-0,0000021	-0,0000010	0,0005590
54	Описательная шкала 18/30-Градация описательной шкалы 3/3	-0,0000263	0,0000264	0,0000000	0,0000000	0,0000373
55	Описательная шкала 19/30-Градация описательной шкалы 1/3	-0,0000183	0,0000184	0,0000000	0,0000000	0,0000280
56	Описательная шкала 19/30-Градация описательной шкалы 2/3	0,0003687	-0,0003705	-0,0000018	-0,0000009	0,0005228
57	Описательная шкала 19/30-Градация описательной шкалы 3/3	-0,0000069	0,0000070	0,0000000	0,0000000	0,0000099
58	Описательная шкала 20/30-Градация описательной шкалы 1/3	-0,0002776	0,0002776	-0,0000013	-0,0000006	0,0003937
59	Описательная шкала 20/30-Градация описательной шкалы 2/3	-0,0001711	0,0001706	-0,0000005	-0,0000002	0,0002416
60	Описательная шкала 20/30-Градация описательной шкалы 3/3	-0,0001455	0,0001451	-0,0000003	-0,0000001	0,0002056
61	Описательная шкала 21/30-Градация описательной шкалы 1/3	-0,0003121	0,0003105	-0,0000016	-0,0000008	0,0004403
62	Описательная шкала 21/30-Градация описательной шкалы 2/3	0,0001403	-0,0001404	-0,0000001	0,0000000	0,0001985
63	Описательная шкала 21/30-Градация описательной шкалы 3/3	0,0001101	-0,0001101	0,0000000	0,0000000	0,0001558
64	Описательная шкала 22/30-Градация описательной шкалы 1/3	0,0001096	-0,0001096	0,0000000	0,0000000	0,0001551
65	Описательная шкала 22/30-Градация описательной шкалы 2/3	-0,0002137	0,0002129	-0,0000008	-0,0000003	0,0003017
66	Описательная шкала 22/30-Градация описательной шкалы 3/3	0,0003472	-0,0003488	-0,0000016	-0,0000008	0,0004922
67	Описательная шкала 23/30-Градация описательной шкалы 1/3	0,0001954	-0,0001958	-0,0000004	-0,0000001	0,0002767
68	Описательная шкала 23/30-Градация описательной шкалы 2/3	0,0004325	-0,0004352	-0,0000026	-0,0000013	0,0006136
69	Описательная шкала 23/30-Градация описательной шкалы 3/3	0,0001518	-0,0001520	-0,0000002	0,0000000	0,0002149
70	Описательная шкала 24/30-Градация описательной шкалы 1/3	-0,0001424	0,0001421	-0,0000003	-0,0000001	0,0002012
71	Описательная шкала 24/30-Градация описательной шкалы 2/3	0,0000745	-0,0000745	0,0000000	0,0000000	0,0001054
72	Описательная шкала 24/30-Градация описательной шкалы 3/3	0,0000312	-0,0000311	0,0000000	0,0000000	0,0000441
73	Описательная шкала 25/30-Градация описательной шкалы 1/3	-0,0005643	0,0005590	-0,0000053	-0,0000026	0,0007944
74	Описательная шкала 25/30-Градация описательной шкалы 2/3	0,0001508	-0,0001510	-0,0000002	0,0000000	0,0002135
75	Описательная шкала 25/30-Градация описательной шкалы 3/3	0,0000945	-0,0000945	0,0000000	0,0000000	0,0001338
76	Описательная шкала 26/30-Градация описательной шкалы 1/3	-0,0000961	0,0000960	-0,0000001	0,0000000	0,0001359
77	Описательная шкала 26/30-Градация описательной шкалы 2/3	-0,0000667	0,0000667	0,0000000	0,0000000	0,0000945
78	Описательная шкала 26/30-Градация описательной шкалы 3/3	0,0000589	-0,0000588	0,0000000	0,0000000	0,0000833
79	Описательная шкала 27/30-Градация описательной шкалы 1/3	-0,0002035	0,0002028	-0,0000007	-0,0000003	0,0002873
80	Описательная шкала 27/30-Градация описательной шкалы 2/3	0,0001443	-0,0001444	-0,0000001	0,0000000	0,0002042
81	Описательная шкала 27/30-Градация описательной шкалы 3/3	-0,0000005	0,0000005	0,0000000	0,0000000	0,0000007
82	Описательная шкала 28/30-Градация описательной шкалы 1/3	-0,0001193	0,0001191	-0,0000002	-0,0000001	0,0001687
83	Описательная шкала 28/30-Градация описательной шкалы 2/3	0,0001619	-0,0001622	-0,0000002	-0,0000001	0,0002292
84	Описательная шкала 28/30-Градация описательной шкалы 3/3	0,0001553	-0,0001555	-0,0000002	0,0000000	0,0002198
85	Описательная шкала 29/30-Градация описательной шкалы 1/3	0,0000372	-0,0000371	0,0000000	0,0000000	0,0000526
86	Описательная шкала 29/30-Градация описательной шкалы 2/3	0,0002797	-0,0002807	-0,0000010	-0,0000004	0,0003963
87	Описательная шкала 29/30-Градация описательной шкалы 3/3	-0,0000183	0,0000183	0,0000000	0,0000000	0,0000259
88	Описательная шкала 30/30-Градация описательной шкалы 1/3	-0,0000505	0,0000505	0,0000000	0,0000000	0,0000715
89	Описательная шкала 30/30-Градация описательной шкалы 2/3	0,0002696	-0,0002705	-0,0000009	-0,0000004	0,0003820
90	Описательная шкала 30/30-Градация описательной шкалы 3/3	-0,0000407	0,0000408	0,0000000	0,0000000	0,0000577
	Сумма	-0,0000270	-0,0000283	-0,0000554	0,0000000	0,0000000
	Среднее	-0,0000002	-0,0000002	0,0000000	-0,0000002	0,0000000
	Среднеквадратичное отклонение	0,0001978	0,0001976	0,0000000	0,0000000	0,0001971

Ниже приведем численный расчет отношения сигнал/шум для задачи, решаемой в данной работе. Из таблиц 11 и 12 видно, что информационная мощность реальной модели равна: $A_{real}=0,1039203$, а случайной модели – $A_{noise}=0,0001971$.

$$A_{real} = 0,1039203,$$

$$A_{noise} = 0,0001971,$$

$$P_{real} = 0,1039203^2 = 0,01079942875209,$$

$$P_{noise} = 0,0001971^2 = 0,00000003884841,$$

$$P_{signal} = 0,1039203^2 - 0,0001971^2 = 0,01079938990368$$

Тогда получим:

$$SNR = \frac{P_{real}}{P_{noise}} - 1 = \frac{(A_{real})^2}{(A_{noise})^2} - 1 = \frac{(0,1039203)^2}{(0,0001971)^2} - 1 = 277987,951210358. \quad (8)$$

Выражения (8) и (8*) имеют чрезвычайно важное значение для нашей работы и по своей математической форме, и по смыслу.

По форме выражение (8) просто совпадает с известным в экономике коэффициентом возврата инвестиций ROI, который в АСК-анализа и системе «Эйдос» используется как один из частных критериев в системно-когнитивных моделях INF4 и INF5 (таблица 8). Кроме того выражение (8*) по форме практически полностью совпадают (с точностью до единиц измерения) с семантической мерой целесообразности информации А.Харкевича, который в АСК-анализа и системе «Эйдос» используется как один из частных критериев в системно-когнитивных моделях INF1 и INF2 (таблица 8).

По смыслу выражение (8) означает, что **влияние характеристик транзакций более чем в 277987 раз выше влияния шума, что с чрезвычайно высокой убедительностью подтверждает предложенную автором гипотезу.**

Если измерять отношение сигнал/шум как принято в децибелах (дБ), то необходимо использовать формулу:

$$SNR(dB) = 10 \times \text{Log}_{10} \left(\frac{P_{signal}}{P_{noise}} \right) \dots \dots \dots (8^*)$$

$$SNR(dB) = 10 \times \text{Log}_{10} \left(\frac{P_{signal}}{P_{noise}} \right) = 10 \times \text{Log}_{10} \left(\frac{0,01079938990368}{0,00000003884841} \right) \approx 54,44 \quad (9)$$

Много это или мало, 54 дБ? Обратимся к таблице шумов (таблица 13):

Таблица 13 – Тбалица шумов в децибелах

Децибел дБА	Характеристика шума	Источники звука
0	Ничего не слышно	
5	Почти не слышно	
10	Почти не слышно	тихий шелест листьев
15	Едва слышно	шорох листвы
20	Едва слышно	шепот человека (на расстоянии 1 метр).
25	Тихо	шепот человека (1м)
30	Тихо	шепот, тиканье настенных часов.
		Допустимый максимум по нормам для жилых помещений ночью, с 23 до 7 ч. (СНиП 23-03-2003 «Защита от шума»).
35	Довольно слышно	приглушенный разговор
40	Довольно слышно	обычная человеческая речь. Норма для жилых помещений днём, с 7 до 23 ч. Подробнее читать в "Российской газете".
		Сорока-децибелльный уровень звука установлен, как допустимый, нормативами (СП 51.13330... ЗАЩИТА ОТ ШУМА, СНиП 23-03...) для аудиторий образовательных организаций, учебных кабинетов, классных помещений, библиотек и конференц-залов.
45	Довольно слышно	обычный, спокойный разговор. Норма для спортивных залов.
50	Отчётливо слышно	разговор, пишущая машинка. Норма для кабинетов административных зданий, офисных и рабочих помещений.
55	Отчётливо слышно	Верхняя норма для офисных помещений класса А (по европейским нормам)
60	Шумно	Норма для контор
65	Шумно	громкий разговор (1м)
70	Шумно	громкие разговоры (1м)
75	Шумно	крик, смех (1м)
80	Очень шумно	крик, мотоцикл с глушителем, шум пылесоса (с большой мощностью двигателя - 2 киловатта).
85	Очень шумно	громкий человеческий крик, мотоцикл с глушителем
90	Очень шумно	громкие крики, грузовой железнодорожный вагон (в семи метрах)
95	Очень шумно	вагон метро (в 7 метрах снаружи или внутри движущегося вагона)
100	Крайне шумно	оркестр, вагон метро (прерывисто), раскаты грома, визг работающей бензопилы, шум от бензиновой газонокосилки
		Максимально допустимое звуковое давление для наушников плеера (по европейским нормам)
105	Крайне шумно	шум в салоне самолёта (до 80-х годов XX столетия) и предельный уровень звукового давления (в дБА) для бензодвигательных цепных пил малой мощности, на их максимальных режимах работы.
110	Крайне шумно	вертолёт
115	Крайне шумно	пескоструйный аппарат (1м) и подобные механизмы
120	Почти невыносимо	отбойный молоток (1м)
125	Почти невыносимо	
130	Болевой порог	самолёт на старте
135	Контузия	
140	Контузия	звук взлетающего реактивного самолета, у края взлётно-посадочной полосы
145	Контузия	старт ракеты.
150	Контузия, травмы	звук, в момент выстрела из ружья "unsuppressed"(без глушителя), среднего калибра, вблизи дульной части ствола.
155	Контузия, травмы	
160	Шок, травмы	ударная волна от сверхзвукового самолёта

Источник: <https://www.kakras.ru/doc/shum-decibel.html>

Как видно из таблицы 13 громкость 55 децибел соответствует отчетливой слышимости. Это верхняя норма для офисных помещений класса А (по европейским нормам).

Это означает, что зависимость степени риска мошенничества от значений характеристик транзакций при операциях по кредитным картам в созданной модели выражена громко и совершенно отчетливо.

2.3.5. Выводы

Все это и есть совершенно ясное и убедительное подтверждение предложенной гипотезы, подтверждающее существование зависимости степени риска мошенничества от значений характеристик транзакций при операциях по кредитным картам.

Задача-3. Решение задачи прогнозирования (системная идентификация), интегральные критерии

3.1. Что такое интегральный критерий и для чего он нужен?

Как влияет на поведение объекта моделирования одно значение фактора, отражено в системно-когнитивных моделях. Как влияет система значений факторов, определяется с помощью интегральных критериев. В интегральном критерии используется система из большого количества частных критериев и значения всех их сводятся к одному значению интегрального критерия. Поэтому вычисление значений интегрального критерия сходства объекта распознаваемой (ее еще называют тестовой) выборки с обобщенными образами всех классов называется *системной идентификацией*.

В настоящее время в системе «Эйдос» используется два *аддитивных* интегральных критерия:

- сумма знаний;
- резонанс знаний.

3.2. 1-й интегральный критерий «Сумма знаний»

Интегральный критерий «Сумма знаний» представляет собой суммарное количество знаний, содержащееся в системе факторов различной природы, характеризующих сам объект управления, управляющие факторы и окружающую среду, о переходе объекта в будущие целевые или нежелательные состояния.

Интегральный критерий представляет собой аддитивную функцию от частных критериев знаний, представленных в help режима 3.3:

$$I_j = (\vec{I}_{ij}, \vec{L}_i).$$

В выражении круглыми скобками обозначено скалярное произведение. В координатной форме это выражение имеет вид:

$$I_j = \sum_{i=1}^M I_{ij} L_i,$$

где: М – количество градаций описательных шкал (признаков);

$\vec{I}_{ij} = \{I_{ij}\}$ – вектор состояния j-го класса;

$\vec{L}_i = \{L_i\}$ – вектор состояния распознаваемого объекта, включающий все виды факторов, характеризующих сам объект, управляющие воздействия и окружающую среду (массив–локатор), т.е.:

$$\vec{L}_i = \begin{cases} 1, & \text{если } i - \text{й фактор действует;} \\ n, & \text{где: } n > 0, \text{ если } i - \text{й фактор действует с истинностью } n; \\ 0, & \text{если } i - \text{й фактор не действует.} \end{cases}$$

В текущей версии системы «Эйдос-Х++» значения координат вектора состояния распознаваемого объекта принимались равными либо 0, если признака нет, или n, если он присутствует у объекта с интенсивностью n, т.е. представлен n раз (например, буква «о» в слове «молоко» представлена 3 раза, а буква «м» - один раз).

Если представить информацию распознаваемой выборки в виде матрицы, в которой каждая строка будет описывать один объект распознаваемой выборки, то *операцию распознавания этой выборки с помощью 1-го интегрального критерия можно представить себе как операцию умножения матрицы распознаваемой выборки на матрицу статистической или системно-когнитивной модели.* Результатом является матрица произведения, в которой каждый элемент является суммой произведений элементов соответствующих строки распознаваемой матрицы и столбца модели.

3.3. 2-й интегральный критерий «Семантический резонанс знаний»

Этот интегральный критерий представляет собой нормированное суммарное количество знаний, содержащееся в системе факторов различной природы, характеризующих сам объект управления, управляющие факторы и окружающую среду, о переходе объекта в будущие целевые или нежелательные состояния.

Интегральный критерий представляет собой аддитивную функцию от частных критериев знаний и имеет вид:

$$I_j = \frac{1}{\sigma_j \sigma_l M} \sum_{i=1}^M (I_{ij} - \bar{I}_j) (L_i - \bar{L}),$$

где:

M – количество градаций описательных шкал (признаков);

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

\bar{L} – среднее по вектору объекта;

σ_j – среднеквадратичное отклонение частных критериев знаний вектора класса;

σ_1 – среднеквадратичное отклонение по вектору распознаваемого объекта.

$\vec{I}_{ij} = \{I_{ij}\}$ – вектор состояния j -го класса;

$\vec{L}_i = \{L_i\}$ – вектор состояния распознаваемого объекта, включающий все виды факторов, характеризующих сам объект, управляющие воздействия и окружающую среду (массив–локатор), т.е.:

$$\vec{L}_i = \begin{cases} 1, & \text{если } i - \text{й фактор действует;} \\ n, & \text{где: } n > 0, \text{ если } i - \text{й фактор действует с истинностью } n; \\ 0, & \text{если } i - \text{й фактор не действует.} \end{cases}$$

В текущей версии системы «Эйдос-Х++» значения координат вектора состояния распознаваемого объекта принимались равными либо 0, если признака нет, или n , если он присутствует у объекта с интенсивностью n , т.е. представлен n раз (например, буква «о» в слове «молоко» представлена 3 раза, а буква «м» – один раз).

3.4. Некоторые математические свойства интегральных критериев

Свое наименование интегральный критерий сходства «Семантический резонанс знаний» получил потому, что по своей математической форме является корреляцией двух векторов: состояния j -го класса и состояния распознаваемого объекта.

Приведенное выражение для интегрального критерия «Семантический резонанс знаний» получается непосредственно из выражения для критерия «Сумма знаний» после замены координат перемножаемых векторов их **стандартизированными** значениями:

$$I_{ij} \rightarrow \frac{I_{ij} - \bar{I}_j}{\sigma_j}, \quad L_i \rightarrow \frac{L_i - \bar{L}}{\sigma_1}.$$

Свое наименование интегральный критерий сходства «Семантический резонанс знаний» получил потому, что по своей математической форме является корреляцией двух векторов: состояния j -го класса и состояния распознаваемого объекта.

Существует и много других способов **нормировки** векторов, например, по формуле линейной интерполяции:

$$I_{ij} \rightarrow \frac{I_{ij} - I_j^{\min}}{I_j^{\max} - I_j^{\min}}, \quad L_i \rightarrow \frac{L_i - L^{\min}}{L^{\max} - L^{\min}},$$

но в системе «Эйдос» они не используются.

Вообще говоря, разные интегральные критерии приводят к различным результатам идентификации и прогнозирования

Оба интегральных критерия, применяемые в системе «Эйдос», обладают очень интересными *математическими свойствами*, которые обеспечивают им важные достоинства.

Во-первых, интегральные критерии имеют **неметрическую** природу, т.е. они являются мерой сходства векторов класса и объекта, но не расстоянием между ними, а косинусом угла между ними, т.е. это межвекторное или информационное расстояние. Поэтому его применение является корректным в **неортонормированных** пространствах, которые, как правило, и встречаются на практике и в которых применение Евклидова расстояния (теоремы Пифагора) является некорректным.

Во-вторых, данные интегральные критерии являются **фильтром**, подавляющим белый шум, который всегда представлен в эмпирических исходных данных и в моделях, созданных на их основе. Это свойство подавлять белый шум проявляется у данного критерия тем ярче, чем больше в модели градаций описательных шкал.

В-третьих, интегральные критерии сходства представляют собой количественную меру сходства/различия конкретного объекта с обобщенным образом класса и имеют тот же смысл, что и **функция принадлежности** элемента множеству в нечеткой логике Лотфи Заде. **Однако** в нечеткой логике эта функция задается исследователем априорно путем выбора из нескольких возможных вариантов, а в АСК-анализе и его программном инструментарии – интеллектуальной системе «Эйдос» она рассчитывается в соответствии с хорошо обоснованной математической моделью непосредственно на основе эмпирических данных.

В-четвертых, кроме того значение интегрального критерия сходства представляет собой адекватную самооценку **степени уверенности** системы в положительном или отрицательном решении о принадлежности/непринадлежности объекта к классу или **риска ошибки** при таком решении.

В-пятых, по сути, при распознавании происходит расчет коэффициентов I_j разложения функции объекта L_i в ряд по функциям классов I_{ij} , т.е. определяется **вес** каждого обобщенного образа класса в образе объекта, что подробнее описано в монографии: [12].

3.5. Решение задачи идентификации и прогнозирования

Для решения задачи оценки степени риска мошенничества при операциях с кредитными картами в наиболее достоверной модели INF3, необходимо присвоить ей статус текущей в режиме 5.6 (рисунок 29):

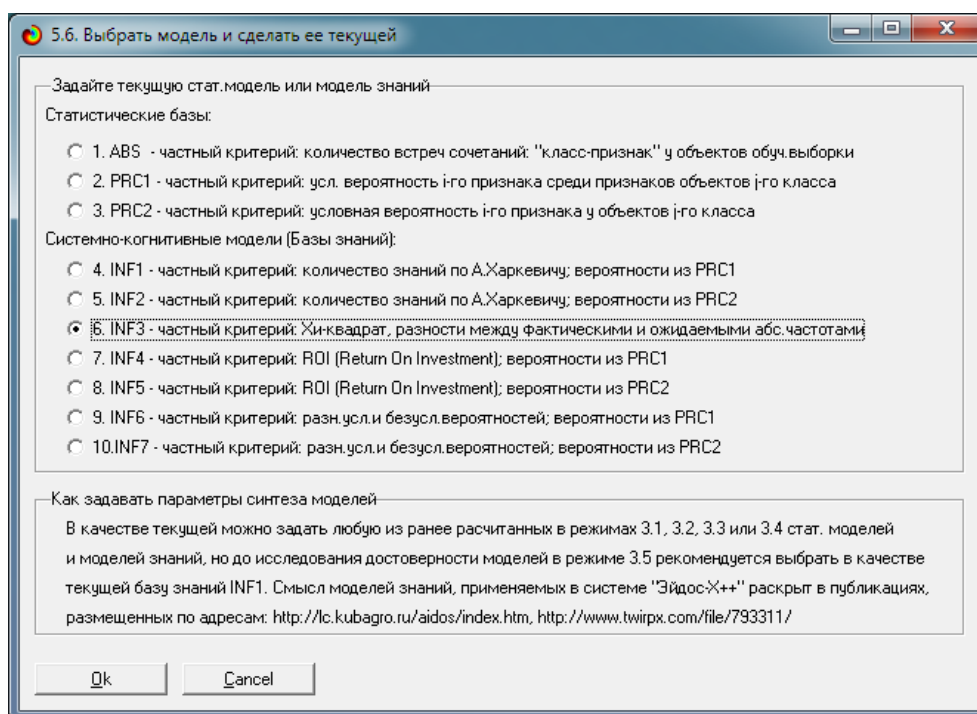


Рисунок 29

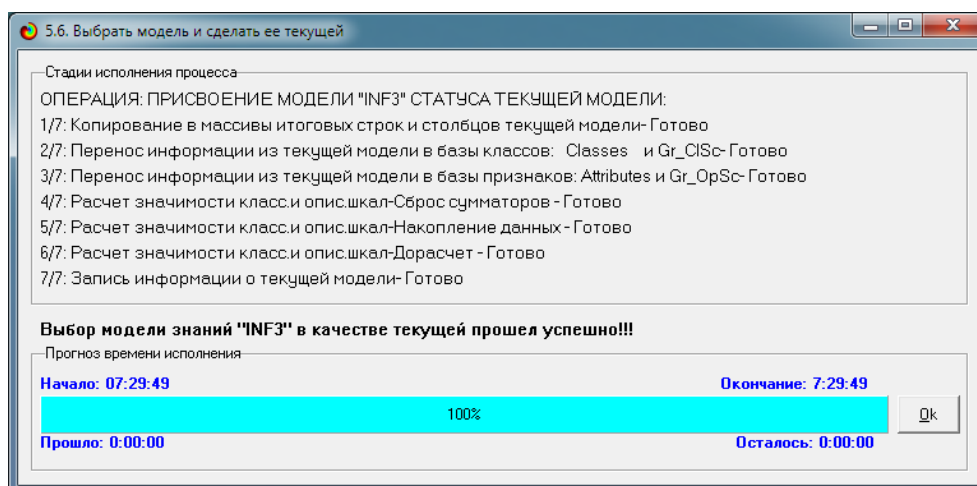


Рисунок 30

Из рисунка 30 видно, что в данном интеллектуальном приложении эта процедура заняла менее секунды.

Затем необходимо ввести *распознаваемую* выборку в систему.

Файл распознаваемой выборки должен иметь абсолютно такую же структуру, как файл исходных данных, но колонки классификационных шкал могут быть пустыми.

Этот файл должен иметь имя: Inp_rasp.xls(x) и должен находиться в той же папке, что и файл исходных данных, например по пути: c:\Aidos-X\AID_DATA\Inp_data\Inp_rasp.xlsx.

Сам ввод распознаваемой выборки осуществляется в уже ранее рассмотренном программном интерфейсе 2.3.2.2 со следующими параметрами (рисунок 31). Эти *параметры те же самые*, что при вводе

обучающей выборки (см. рисунок 7), за исключением одного параметра: задана опция генерации не обучающей, а распознаваемой выборки.

В нашем случае в качестве распознаваемой выборки взята обучающая выборка. Поэтому мы просто скопируем обучающую выборку в распознаваемую в режиме 2.3.1 (рисунок 32). Этот процесс длится на данной задаче около 3 секунд. Посмотреть распознаваемую выборку можно в режиме 4.1.1.

Распознавание (прогнозирование) проводится в режиме 4.1.2 на графическом процессоре (GPU) (рисунок 33).

Из рисунка 34 видно, что весь процесс решения задачи оценки степени риска мошенничества при операциях с кредитными картами в наиболее достоверной модели INF3 на данных из 284807 транзакций занял 10 минут 7 секунд. Отметим, что собственно само распознавание на GPU заняло несколько секунд, а 99% этого времени занял расчет 11 выходных форм по результатам распознавания, которые в текущей версии системы «Эйдос» рассчитываются на центральном процессоре (CPU). На центральном процессоре этот процесс распознавания занял бы значительно больше времени.

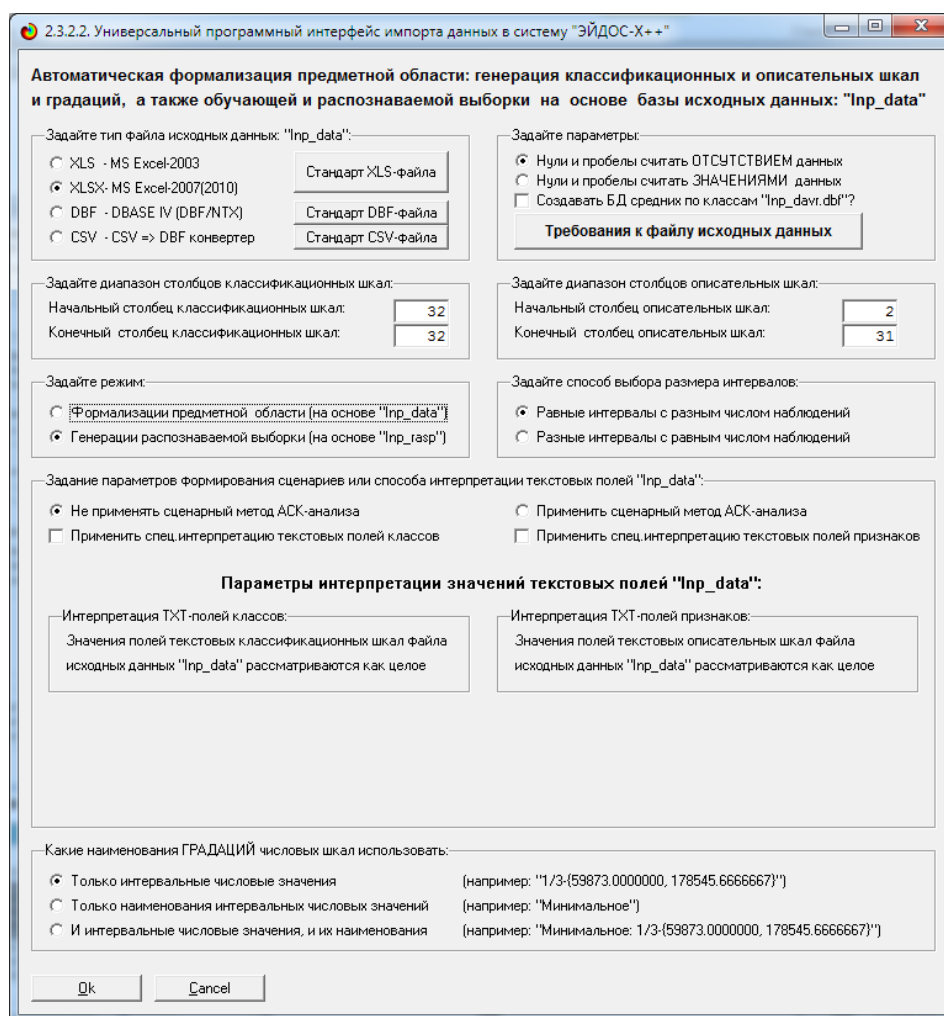


Рисунок 31

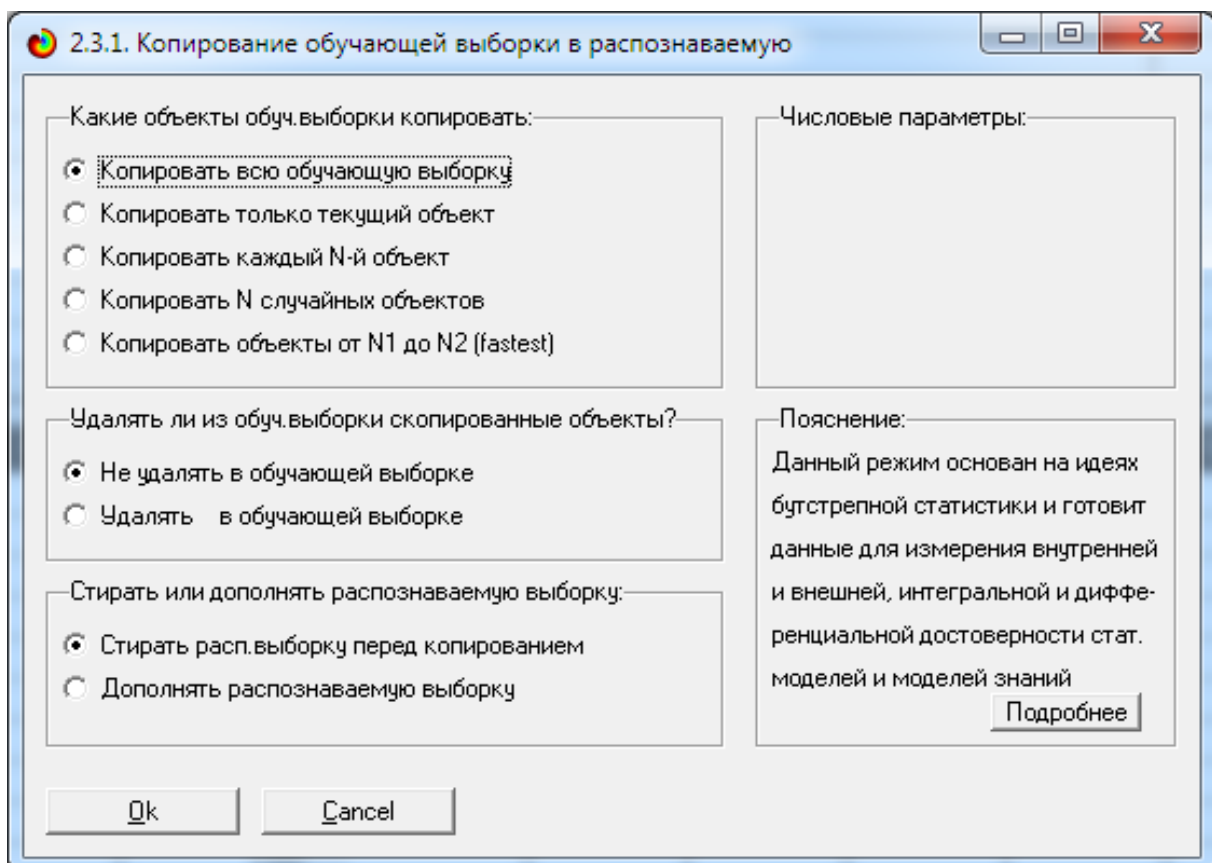
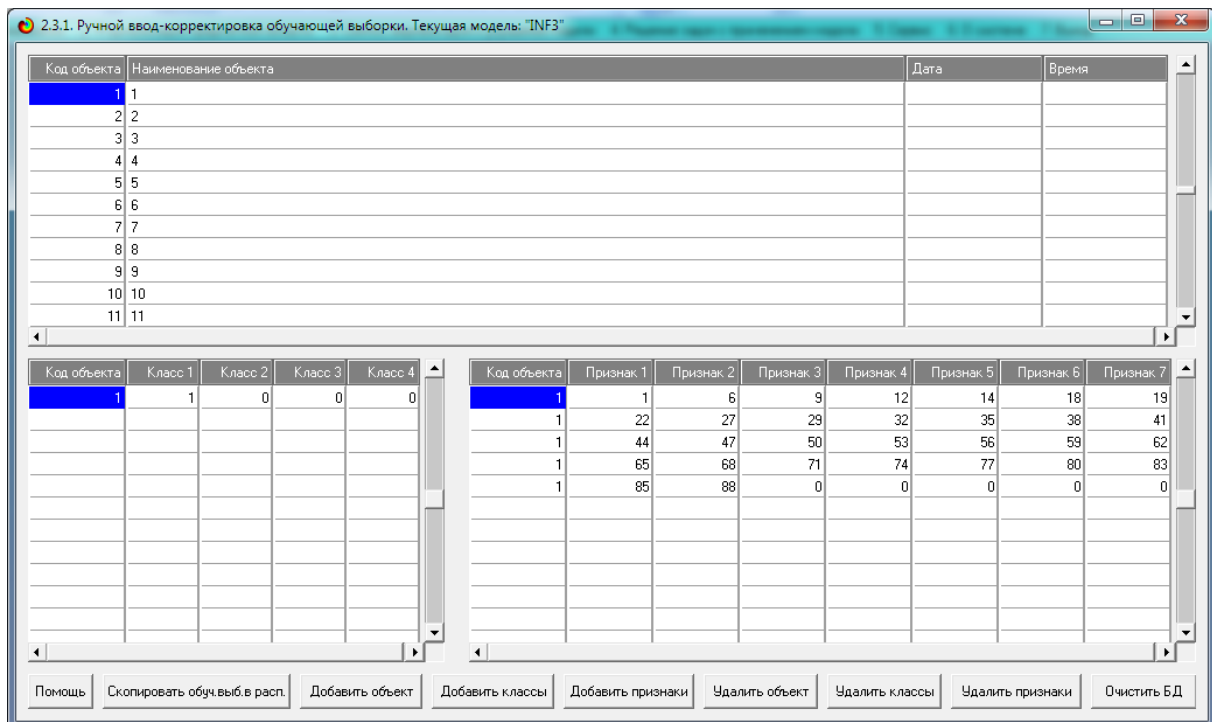


Рисунок 32

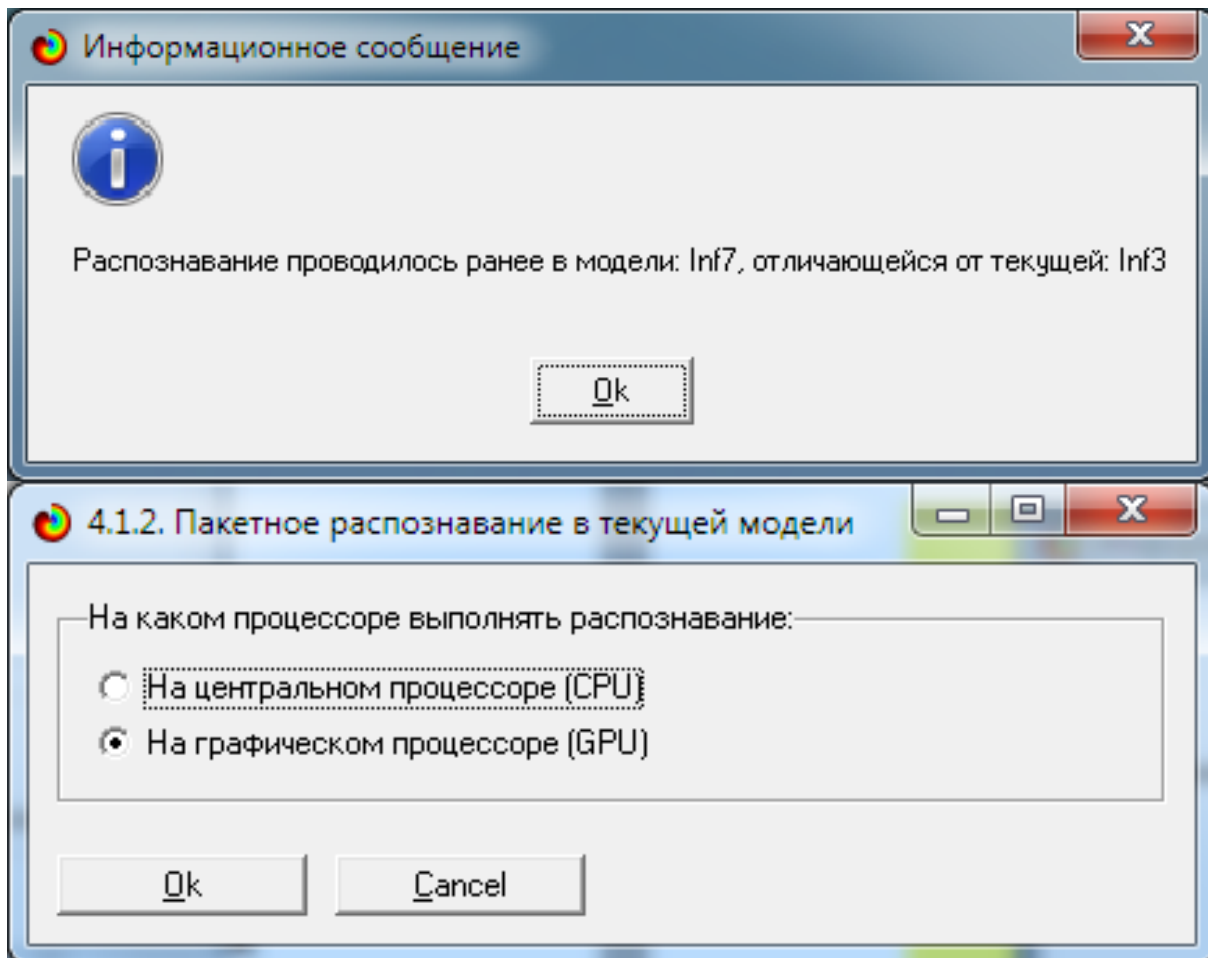


Рисунок 33

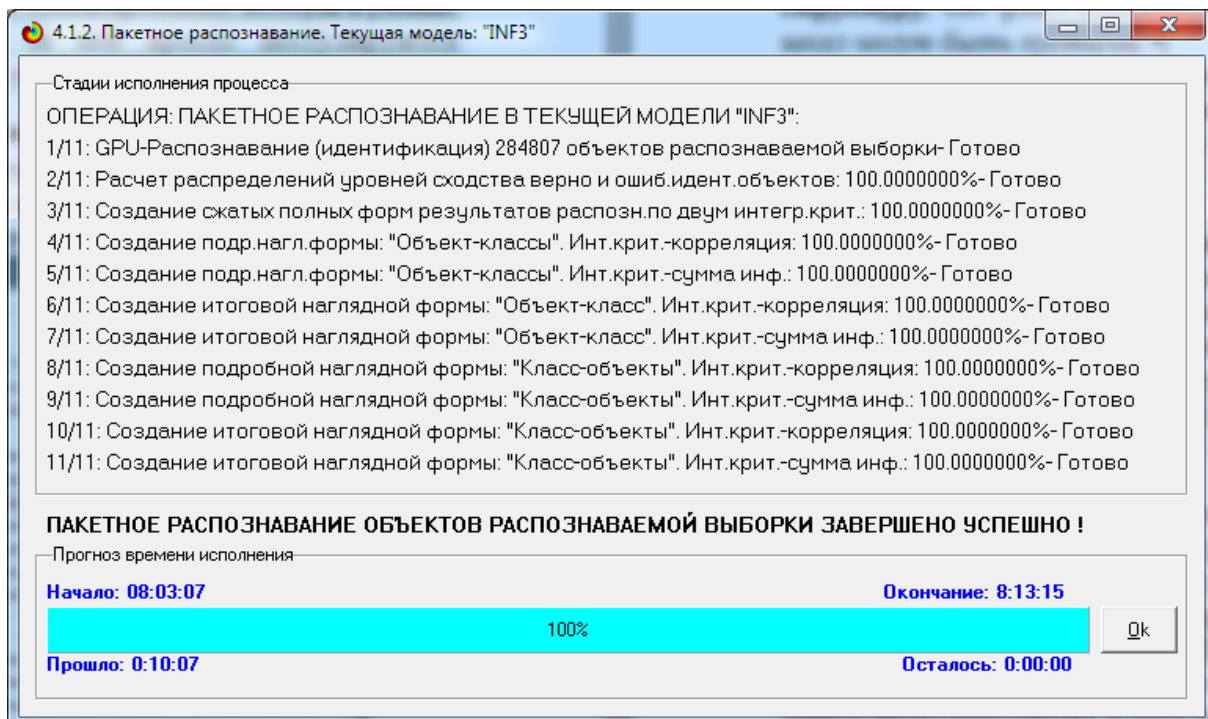


Рисунок 34

Ниже приведены некоторые экранные формы с результатами решения задачи оценки степени риска мошенничества при операциях с кредитными картами в наиболее достоверной модели INF3 (рисунки 35 и 36).

В экранных формах на рисунке 35 показаны **риски мошенничества** по транзакциям (правое окно) для разных классов (левое окно, текущий класс выделен курсором) в порядке убывания сходства с обобщенным образом данного класса. Верхнее и нижнее окна справа содержат результаты идентификации по разным интегральным критериям.

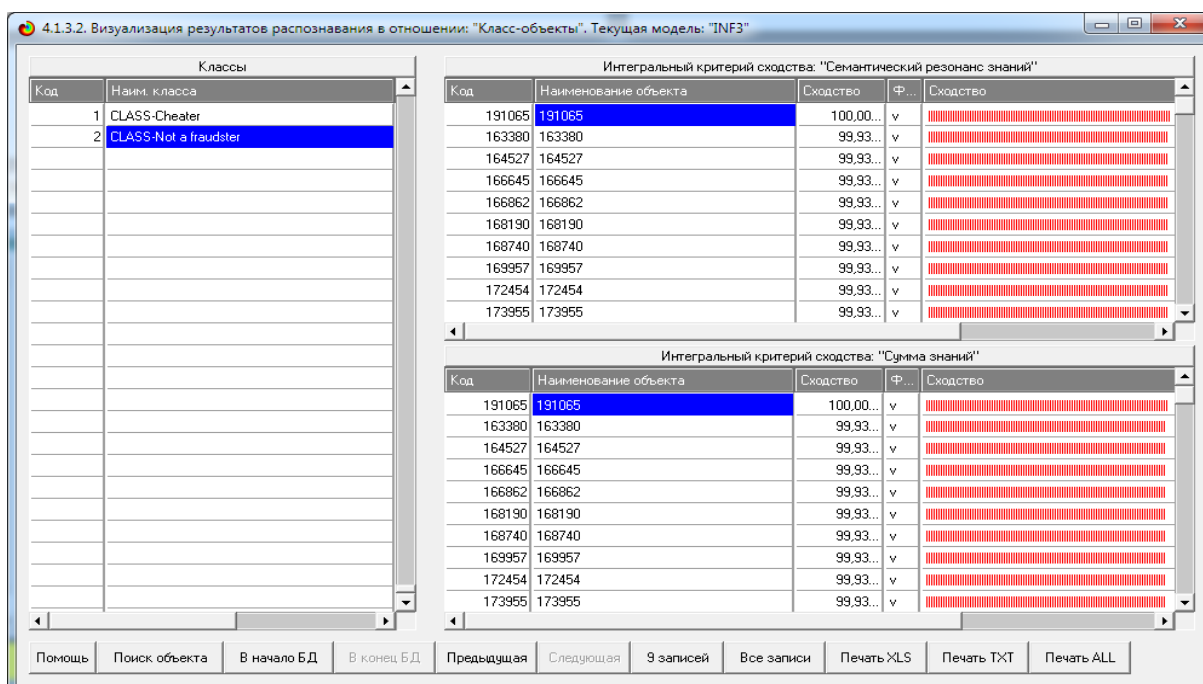
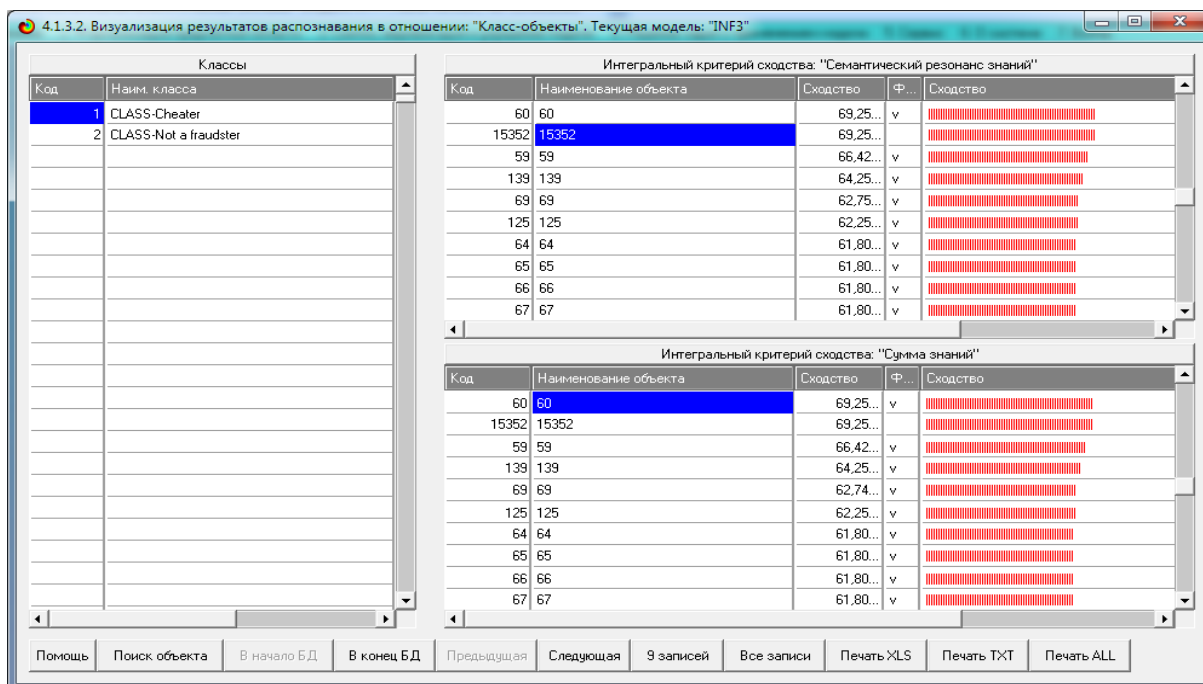


Рисунок 35

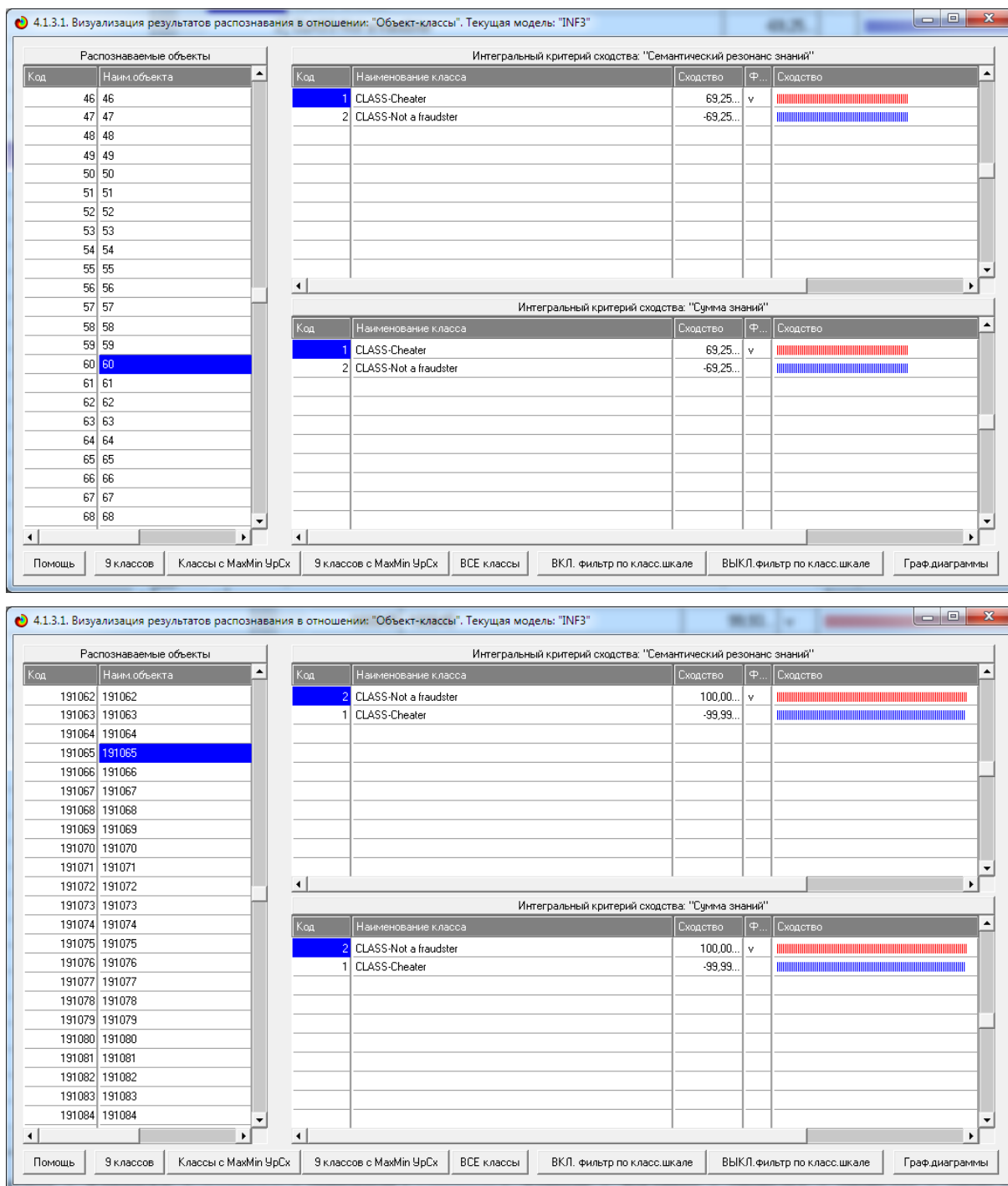


Рисунок 36

В экранных формах на рисунке 36 показаны *риски мошенничества* (сходство) конкретной транзакции (левое окно, выделена курсором) с разными классами (правое окно) в порядке убывания сходства с обобщенным образом данного класса.

В системе «Эйдос» есть и соответствующие табличные формы, представляющие собой базы данных (dbf-файлы) открывающиеся в MS Excel и содержащие значительно больший объем информации.

Задача-4. Решение задачи принятия решений (результаты многопараметрической типизации классов)

Мы привели системе «Эйдос» в качестве примеров 284807 транзакций, описанных рядом параметров. Для каждой транзакции было указано, является ли она мошеннической или нет. На основе этих примеров система провела многопараметрическую типизацию и сформировала обобщенные образы классов. Эти образы количественно отражают, какие значения факторов и в какой степени характерны, а какие и в какой степени нехарактерны для каждого класса, т.е. каждого результата влияния этих значений факторов [10]. По сути, обобщенный образ класса количественно отражает систему детерминации состояния объекта моделирования, соответствующего этому классу значениями факторов. Эта система детерминации классов сама по себе представляет собой большой научный интерес.

На рисунках 37 приведены фрагменты системы детерминации значениями факторов всех классов, отраженных в исследуемой модели. На экранных формах показаны только наиболее сильно влияющие значения факторов. Эти экранные формы можно получить в режиме 4.4.8 системы «Эйдос»:

Выбор класса, соответствующего будущему состоянию объекта управления

Код	Наименование класса	Редукция класс...	N объектов (абс.)	N объектов (%)
1	CLASS-Cheater	112,1322540	14758	0,0000000
2	CLASS-Not a fraudster	112,1333787	8529442	0,0000000

SWOT-анализ класса:1 "CLASS-Cheater" в модели:6 "INF3"

Способствующие факторы и сила их влияния

Код	Наименование фактора и его интервального значения	Сила влияния
14	V4-2/3-{1.8363339, 9.3558389}	338.379
53	V17-2/3-{13.6906908, -2.2185823}	271.221
38	V12-2/3-{9.8396790, -0.9956435}	257.368
35	V11-2/3-{0.8079887, 6.4134510}	242.219
49	V16-1/3-{14.1298545, -3.6481992}	241.314
43	V14-1/3-{19.2143255, -9.3006283}	133.720
28	V9-1/3-{13.4340663, -3.7577127}	133.705
60	V19-3/3-{1.3234718, 5.5919714}	127.176
55	V18-1/3-{9.4987459, -4.6521409}	123.743
31	V10-1/3-{24.5882624, -8.4771296}	101.772
37	V12-1/3-{18.6837146, -9.8396790}	99.791
11	V3-2/3-{29.0895401, -9.8534909}	98.528
52	V17-1/3-{25.1627994, -13.6906908}	77.827
36	V11-3/3-{6.4134510, 12.0189132}	75.829

Препятствующие факторы и сила их влияния

Код	Наименование фактора и его интервального значения	Сила влияния
13	V4-1/3-{5.6831712, 1.8363339}	-368.192
39	V12-3/3-{0.9956435, 7.8483921}	-357.093
54	V17-3/3-{2.2185823, 9.2535263}	-348.982
34	V11-1/3-{4.7974735, 0.8079887}	-317.982
50	V16-2/3-{3.6481992, 6.8334562}	-241.243
29	V9-2/3-{3.7577127, 5.9186410}	-133.132
59	V19-2/3-{2.9450278, 1.3234718}	-129.557
12	V3-3/3-{9.8534909, 9.3825584}	-106.438
32	V10-2/3-{8.4771296, 7.6340033}	-100.766
45	V14-3/3-{0.6130689, 10.5267661}	-84.529
57	V18-3/3-{0.1944642, 5.0410692}	-83.559
3	TIME-3/3-{115194.6666667, 172792.0000000}	-75.217
44	V14-2/3-{9.3006283, 0.6130689}	-49.123
56	V18-2/3-{4.6521409, 1.9444642}	-40.115

ВКЛЮЧИТЬ фильтр по фактору ВЫКЛЮЧИТЬ фильтр по фактору

Помощь Abs Pr1 Pr2 Inf1 Inf2 Inf3 Inf4 Inf5 Inf6 Inf7

SWOT-диаграмма

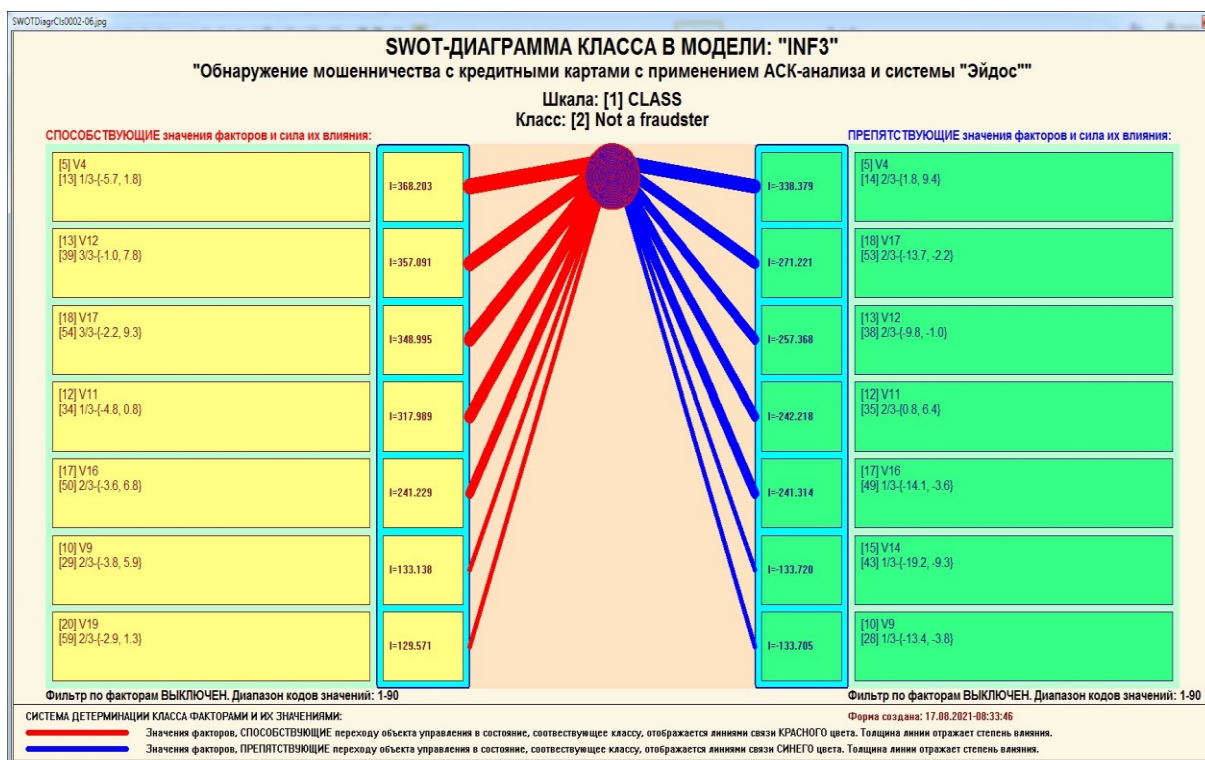
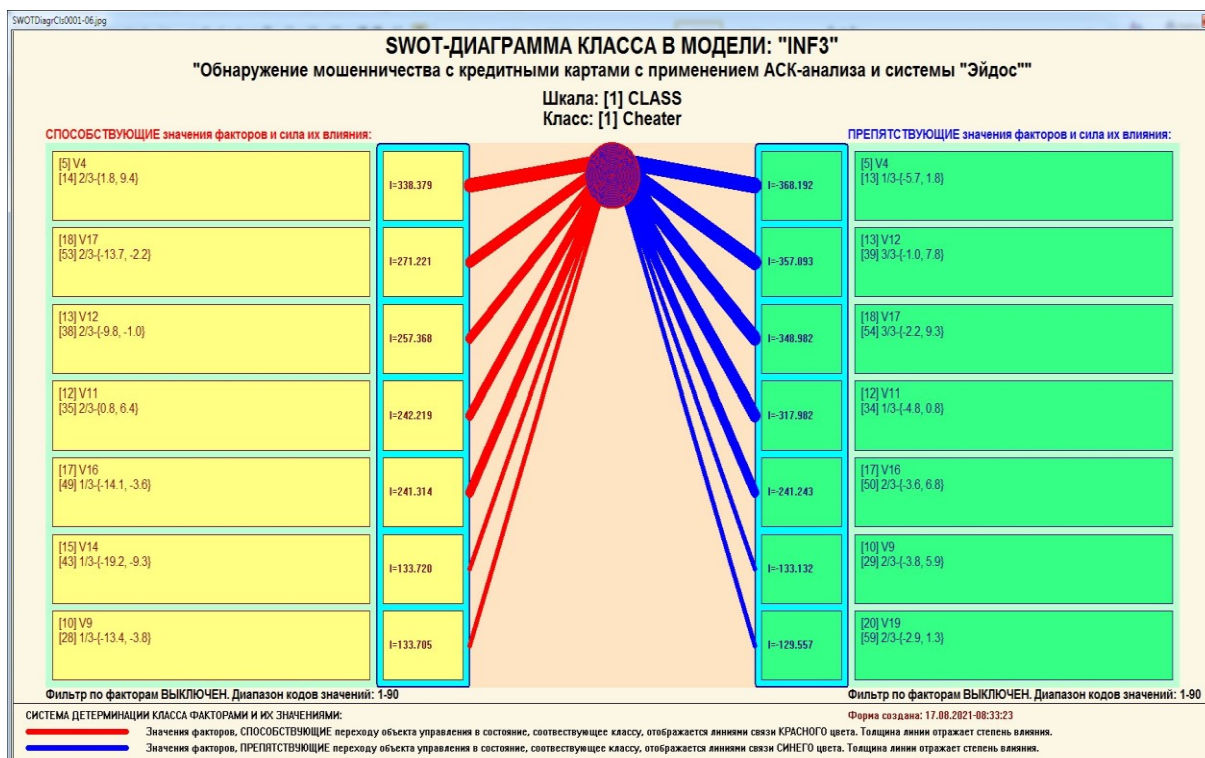


Рисунок 37

На приведенных SWOT-диаграммах слева показаны значения характеристик транзакций, наиболее сильно положительно (слева) и отрицательно (справа) влияющие на принадлежность транзакций к классу, выбранному курсором в верхнем окне [10].

Задача-5. Решение задачи исследования моделируемой предметной области путем исследования ее модели

В системе «Эйдос» есть довольно развитые возможности исследования объекта моделирования путем исследования его модели. Результаты такого исследования можно считать корректными и относящимися к объекту исследования только в том случае, если модель имеет достаточно высокую достоверность, т.е. верно отражает объект моделирования. В нашем случае это именно так.

Исследуем объект моделирования, используя следующие предназначенные для этого инструменты системы «Эйдос»:

1. Инвертированные SWOT-диаграммы значений характеристик транзакций (семантические потенциалы).
2. Кластерно-конструктивный анализ классов.
3. Кластерно-конструктивный анализ значений характеристик транзакций.
4. Нелокальные нейроны.
5. Нелокальная нейронная сеть.
6. 3D-интегральные когнитивные карты.
7. 2D-интегральные когнитивные карты содержательного сравнения классов.
8. 2D-интегральные когнитивные карты содержательного сравнения значений факторов.
9. Когнитивные функции.
10. Значимость характеристик транзакций и их значений.
11. Степень детерминированности классов и классификационных шкал.

5.1. Инвертированные SWOT-диаграммы значений характеристик транзакций (семантические потенциалы)

Эти диаграммы впервые предложены Луценко Е.В. в работе [12]. Их можно получить в режиме 4.4.9. Всего таких инвертированных SWOT-диаграмм в данной модели 90, поэтому все они не могут быть приведены из-за ограничений на объем данной работы. Пример такой диаграммы приведен на рисунках 38.

Когда в задаче классов больше, чем 2, эти диаграммы выглядят более содержательно.

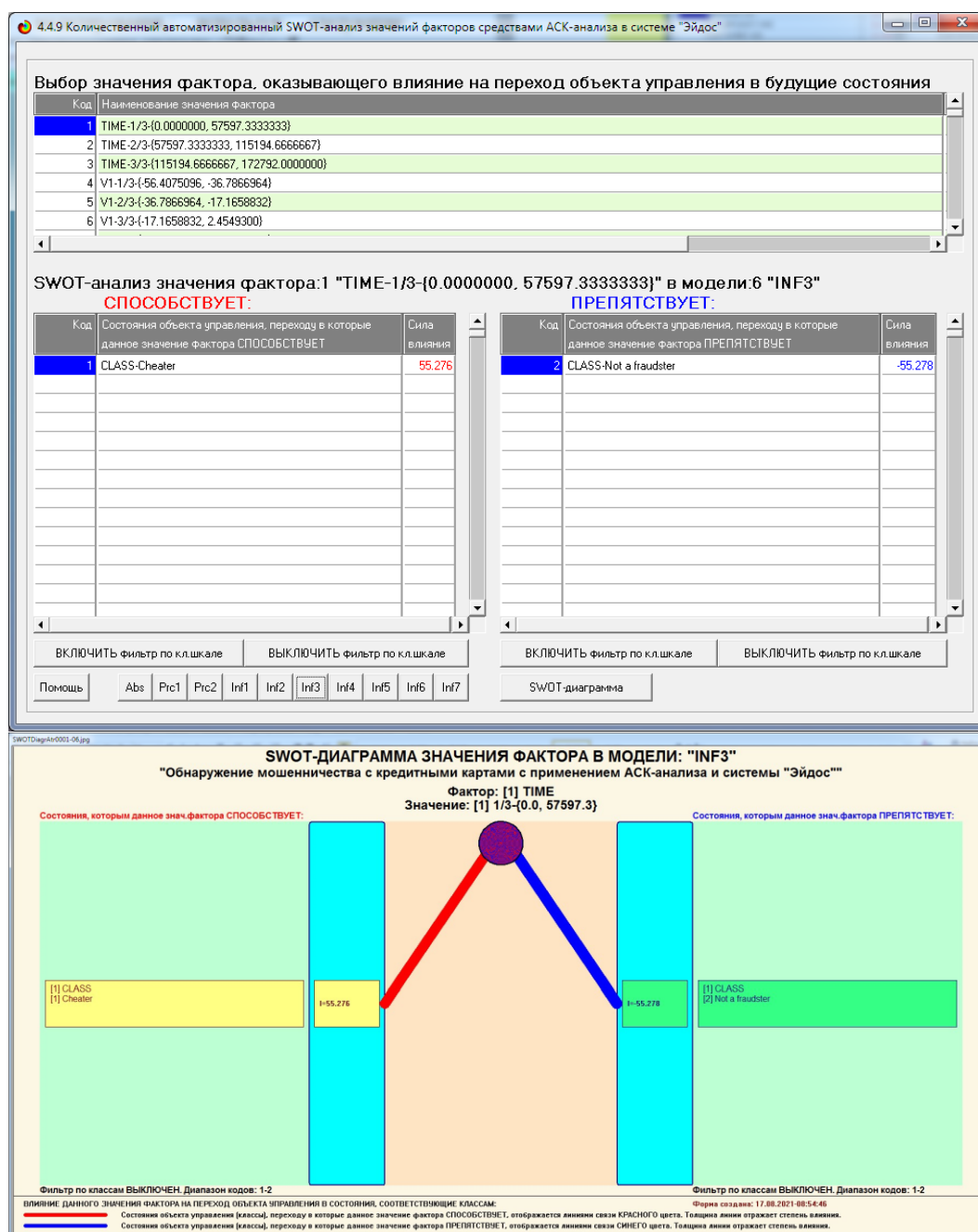


Рисунок 38

5.2. Кластерно-конструктивный анализ классов

Ранее мы решили задачу идентификации и прогнозирования степени риска мошенничества, при которой образ конкретной транзакции, описанной конкретными значениями ее характеристик, сравнивается системой со всеми обобщенными образами классов.

Теперь же мы решим задачу сравнения самих обобщенных образов классов друг с другом по системе их детерминации значениями характеристик транзакций.

Для этого сначала необходимо в режиме 4.2.2.1 рассчитать матрицу сходства обобщенных образов классов (рисунок 39). Эта матрица в нашей модели имеет вид, представленный в таблице 14.

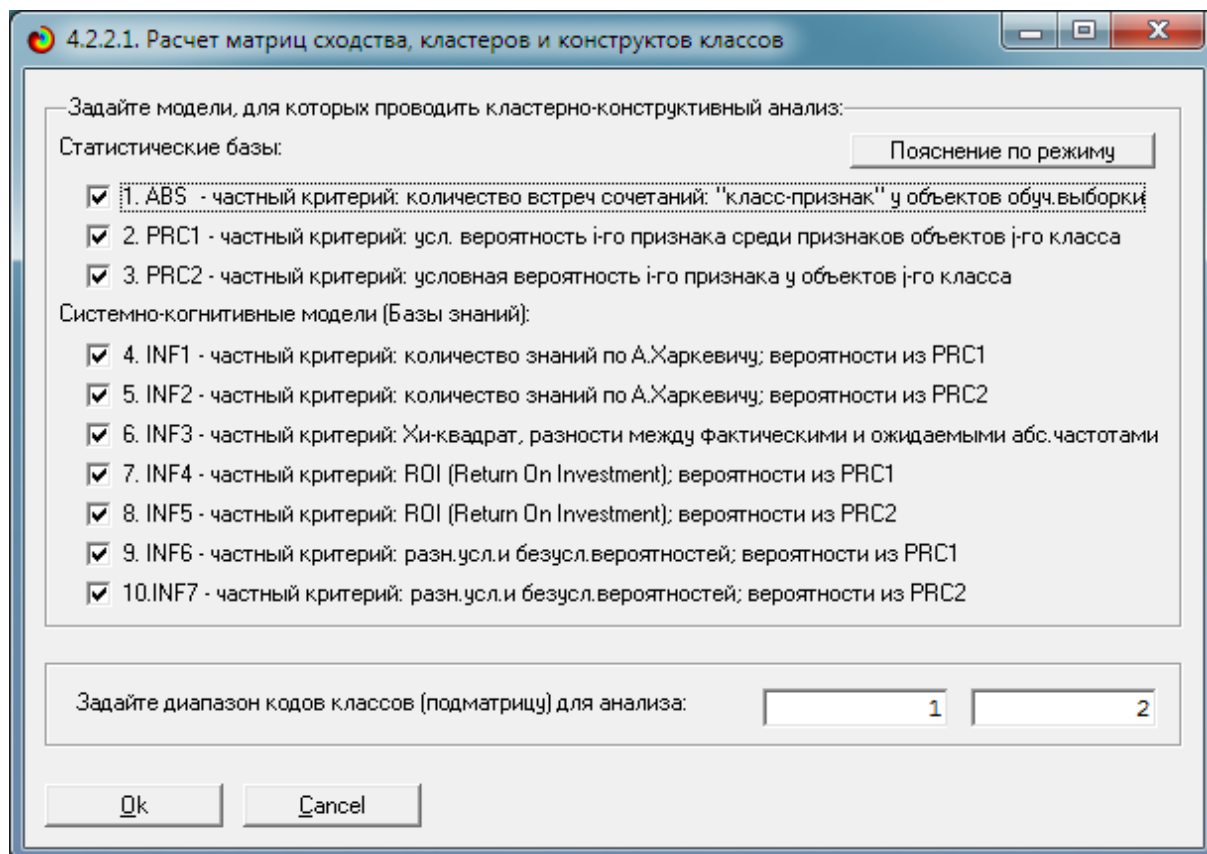


Рисунок 39

Таблица 14 – Матрица сходства классов в модели INF3

			CLASS-Cheater	CLASS-Not a fraudster
KOD_CLS	KOD_CLSC	NAME_CLS		
1	1	CLASS-Cheater	100,0000000	-99,9999999
2	1	CLASS-Not a fraudster	-99,9999999	100,0000000

Информация из матрицы сходства классов может быть отображена в системе «Эйдос» в форме круговой 2d-когнитивной диаграммы в режиме 2.3.2.2 (рисунок 40), а также в форме когнитивной агломеративной дендрограммы [10] в режиме 4.3.2.3. В данной задаче агломеративная кластеризация не проводится, т.к. в модели всего два класса, о чем выводится соответствующее сообщение (рисунок 41).

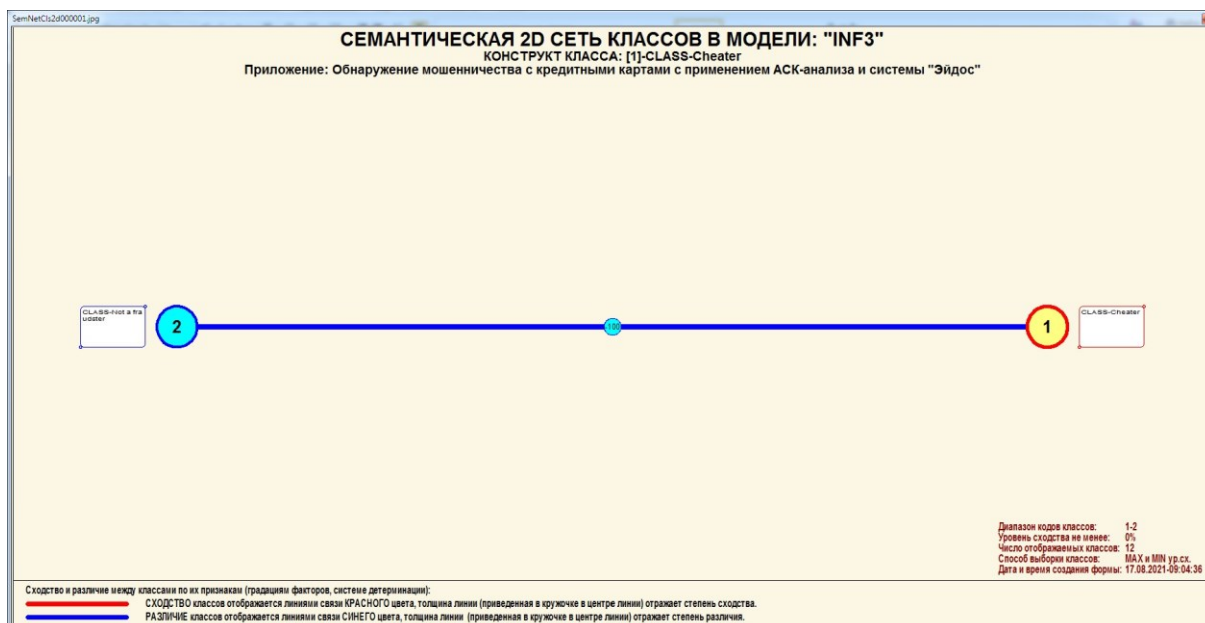


Рисунок 40

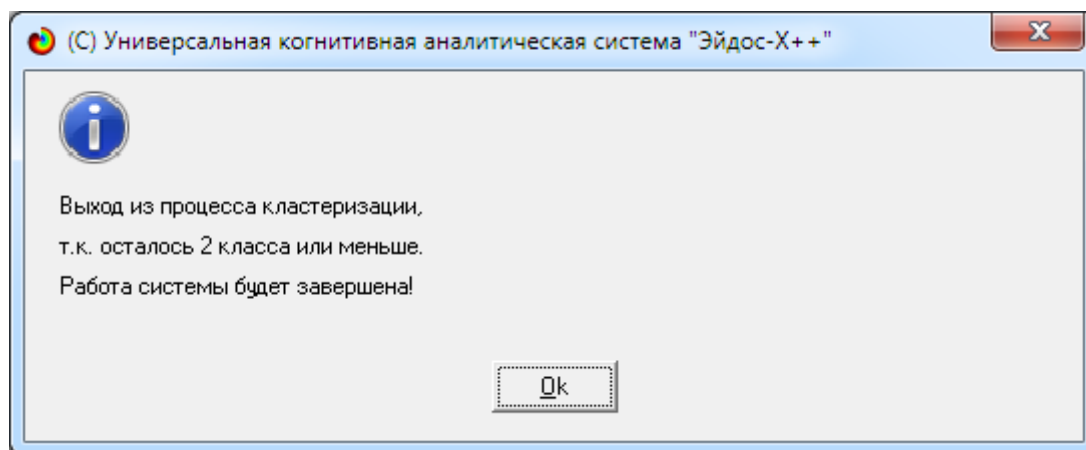


Рисунок 41

5.3. Кластерно-конструктивный анализ значений характеристик транзакций

Совершенно аналогично тому, как классы можно сравнивать по их системам детерминации, так и значения характеристик транзакций можно сравнивать по их влиянию на величину риска мошенничества.

Для этого в режиме 2.3.3.1 нужно рассчитать матрицу сходства значений характеристик транзакций (рисунок 42 и таблица 15).

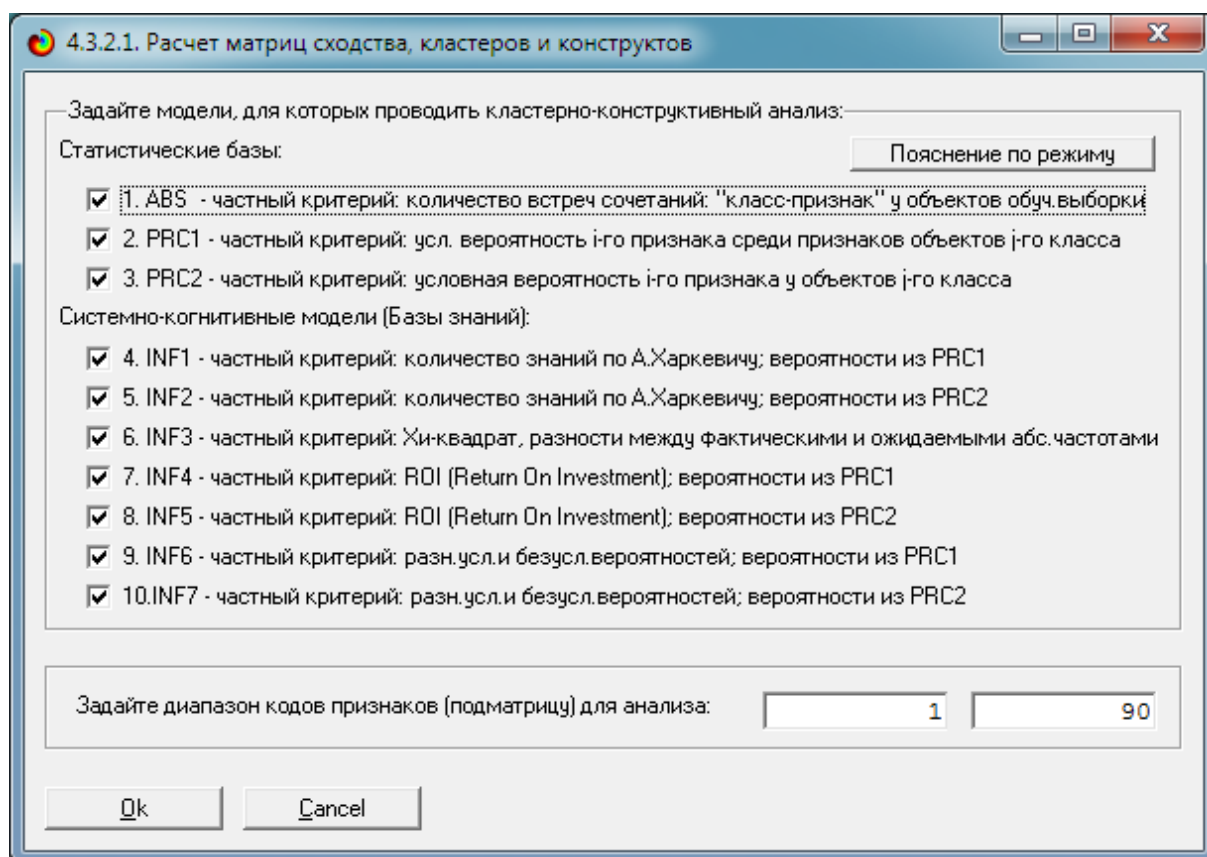


Рисунок 42

Таблица 15 – Матрица сходства значений характеристик транзакций в системно-когнитивной модели INF3 (фрагмент)

KOD_ATR	KOD_OPSC	NAME_ATR	TIME-1/3-{0.0000000, 57597.3333333}	TIME-2/3-{57597.3333333, 115194.6666667}	TIME-3/3-{115194.6666667, 172792.0000000}	V1-1/3-{56.4075096, -36.7866964}	V1-2/3-{36.7866964, -17.1658832}	V1-3/3-{17.1658832, 2.4549300}	V2-1/3-{72.7157276, -41.1245754}	V2-2/3-{41.1245754, -9.5334232}	V2-3/3-{9.5334232, 22.0577290}	V3-1/3-{48.3255894, -29.0895401}
1	1	TIME-1/3-{0.0000000, 57597.3333333}	100	100	-100	-100	100	-100	-100	-100	100	100
2	1	TIME-2/3-{57597.3333333, 115194.6666667}	100	100	-100	-100	100	-100	-100	-100	100	100
3	1	TIME-3/3-{115194.6666667, 172792.0000000}	-100	-100	100	100	-100	100	100	100	-100	-100
4	2	V1-1/3-{56.4075096, -36.7866964}	-100	-100	100	100	-100	100	100	100	-100	-100
5	2	V1-2/3-{36.7866964, -17.1658832}	100	100	-100	-100	100	-100	-100	-100	100	100
6	2	V1-3/3-{17.1658832, 2.4549300}	-100	-100	100	100	-100	100	100	100	-100	-100
7	3	V2-1/3-{72.7157276, -41.1245754}	-100	-100	100	100	-100	100	100	100	-100	-100
8	3	V2-2/3-{41.1245754, -9.5334232}	-100	-100	100	100	-100	100	100	100	-100	-100
9	3	V2-3/3-{9.5334232, 22.0577290}	100	100	-100	-100	100	-100	-100	-100	100	100
10	4	V3-1/3-{48.3255894, -29.0895401}	100	100	-100	-100	100	-100	-100	-100	100	100
11	4	V3-2/3-{29.0895401, -9.8534909}	100	100	-100	-100	100	-100	-100	-100	100	100
12	4	V3-3/3-{9.8534909, 9.3825584}	-100	-100	100	100	-100	100	100	100	-100	-100
13	5	V4-1/3-{5.6831712, 1.8363339}	-100	-100	100	100	-100	100	100	100	-100	-100
14	5	V4-2/3-{1.8363339, 9.3558389}	100	100	-100	-100	100	-100	-100	-100	100	100
15	5	V4-3/3-{9.3558389, 16.8753440}	100	100	-100	-100	100	-100	-100	-100	100	100
16	6	V5-1/3-{113.7433067, -64.2283158}	-100	-100	100	100	-100	100	100	100	-100	-100
17	6	V5-2/3-{64.2283158, -14.7133250}	100	100	-100	-100	100	-100	-100	-100	100	100
18	6	V5-3/3-{14.7133250, 34.8016659}	-100	-100	100	100	-100	100	100	100	-100	-100
19	7	V6-1/3-{26.1605059, 6.9935379}	100	100	-100	-100	100	-100	-100	-100	100	100

20	7	V6-2/3-{-6.9935379, 40.1475817}	-100	-100	100	100	-100	100	100	100	-100	-100
21	7	V6-3/3-{-40.1475817, 73.3016255}	-100	-100	100	100	-100	100	100	100	-100	-100
22	8	V7-1/3-{-43.5572416, 11.1583369}	100	100	-100	-100	100	-100	-100	-100	100	100
23	8	V7-2/3-{-11.1583369, 65.8739154}	-100	-100	100	100	-100	100	100	100	-100	-100
24	8	V7-3/3-{-65.8739154, 120.5894939}	0	0	0	0	0	0	0	0	0	0
25	9	V8-1/3-{-73.2167185, -42.1420762}	-100	-100	100	100	-100	100	100	100	-100	-100
26	9	V8-2/3-{-42.1420762, -11.0674339}	100	100	-100	-100	100	-100	-100	-100	100	100
27	9	V8-3/3-{-11.0674339, 20.0072084}	-100	-100	100	100	-100	100	100	100	-100	-100
28	10	V9-1/3-{-13.4340663, -3.7577127}	100	100	-100	-100	100	-100	-100	-100	100	100
29	10	V9-2/3-{-3.7577127, 5.9186410}	-100	-100	100	100	-100	100	100	100	-100	-100
30	10	V9-3/3-{-5.9186410, 15.5949946}	-100	-100	100	100	-100	100	100	100	-100	-100
31	11	V10-1/3-{-24.5882624, -8.4771296}	100	100	-100	-100	100	-100	-100	-100	100	100
32	11	V10-2/3-{-8.4771296, 7.6340033}	-100	-100	100	100	-100	100	100	100	-100	-100
33	11	V10-3/3-{-7.6340033, 23.7451361}	-100	-100	100	100	-100	100	100	100	-100	-100
34	12	V11-1/3-{-4.7974735, 0.8079887}	-100	-100	100	100	-100	100	100	100	-100	-100
35	12	V11-2/3-{-0.8079887, 6.4134510}	100	100	-100	-100	100	-100	-100	-100	100	100
36	12	V11-3/3-{-6.4134510, 12.0189132}	100	100	-100	-100	100	-100	-100	-100	100	100
37	13	V12-1/3-{-18.6837146, -9.8396790}	100	100	-100	-100	100	-100	-100	-100	100	100
38	13	V12-2/3-{-9.8396790, -0.9956435}	100	100	-100	-100	100	-100	-100	-100	100	100
39	13	V12-3/3-{-0.9956435, 7.8483921}	-100	-100	100	100	-100	100	100	100	-100	-100
40	14	V13-1/3-{-5.7918812, -1.4856265}	100	100	-100	-100	100	-100	-100	-100	100	100
41	14	V13-2/3-{-1.4856265, 2.8206283}	-100	-100	100	100	-100	100	100	100	-100	-100
42	14	V13-3/3-{-2.8206283, 7.1268830}	-100	-100	100	100	-100	100	100	100	-100	-100
43	15	V14-1/3-{-19.2143255, -9.3006283}	100	100	-100	-100	100	-100	-100	-100	100	100
44	15	V14-2/3-{-9.3006283, 0.6130689}	-100	-100	100	100	-100	100	100	100	-100	-100
45	15	V14-3/3-{-0.6130689, 10.5267661}	-100	-100	100	100	-100	100	100	100	-100	-100
46	16	V15-1/3-{-4.4989447, -0.0400493}	100	100	-100	-100	100	-100	-100	-100	100	100
47	16	V15-2/3-{-0.0400493, 4.4188462}	-100	-100	100	100	-100	100	100	100	-100	-100
48	16	V15-3/3-{-4.4188462, 8.8777416}	-100	-100	100	100	-100	100	100	100	-100	-100
49	17	V16-1/3-{-14.1298545, -3.6481992}	100	100	-100	-100	100	-100	-100	-100	100	100
50	17	V16-2/3-{-3.6481992, 6.8334562}	-100	-100	100	100	-100	100	100	100	-100	-100
51	17	V16-3/3-{-6.8334562, 17.3151115}	-100	-100	100	100	-100	100	100	100	-100	-100
52	18	V17-1/3-{-25.1627994, -13.6906908}	100	100	-100	-100	100	-100	-100	-100	100	100
53	18	V17-2/3-{-13.6906908, -2.2185823}	100	100	-100	-100	100	-100	-100	-100	100	100
54	18	V17-3/3-{-2.2185823, 9.2535263}	-100	-100	100	100	-100	100	100	100	-100	-100
55	19	V18-1/3-{-9.4987459, -4.6521409}	100	100	-100	-100	100	-100	-100	-100	100	100
56	19	V18-2/3-{-4.6521409, 0.1944642}	-100	-100	100	100	-100	100	100	100	-100	-100
57	19	V18-3/3-{-0.1944642, 5.0410692}	100	100	-100	-100	100	100	100	100	-100	-100
58	20	V19-1/3-{-7.2135274, -2.9450278}	100	100	-100	-100	100	-100	-100	-100	100	100
59	20	V19-2/3-{-2.9450278, 1.3234718}	-100	-100	100	100	-100	100	100	100	-100	-100
60	20	V19-3/3-{-1.3234718, 5.5919714}	100	100	-100	-100	100	-100	-100	-100	100	100
61	21	V20-1/3-{-54.4977205, -23.1915123}	-100	-100	100	100	-100	100	100	100	-100	-100
62	21	V20-2/3-{-23.1915123, 8.1146960}	-100	-100	100	100	-100	100	100	100	-100	-100
63	21	V20-3/3-{-8.1146960, 39.4209042}	100	100	-100	-100	100	-100	-100	-100	100	100
64	22	V21-1/3-{-34.8303821, -14.1526417}	100	100	-100	-100	100	-100	-100	-100	100	100
65	22	V21-2/3-{-14.1526417, 6.5250988}	-100	-100	100	100	-100	100	100	100	-100	-100
66	22	V21-3/3-{-6.5250988, 27.2028392}	100	100	-100	-100	100	-100	-100	-100	100	100
67	23	V22-1/3-{-10.9331437, -3.7877324}	100	100	-100	-100	100	-100	-100	-100	100	100
68	23	V22-2/3-{-3.7877324, 3.3576788}	-100	-100	100	100	-100	100	100	100	-100	-100
69	23	V22-3/3-{-3.3576788, 10.5030901}	100	100	-100	-100	100	-100	-100	-100	100	100
70	24	V23-1/3-{-44.8077352, -22.3623529}	-100	-100	100	100	-100	100	100	100	-100	-100
71	24	V23-2/3-{-22.3623529, 0.0830294}	-100	-100	100	100	-100	100	100	100	-100	-100
72	24	V23-3/3-{-0.0830294, 22.5284117}	100	100	-100	-100	100	-100	-100	-100	100	100
73	25	V24-1/3-{-2.8366269, -0.3629016}	100	100	-100	-100	100	-100	-100	-100	100	100
74	25	V24-2/3-{-0.3629016, 2.1108238}	-100	-100	100	100	-100	100	100	100	-100	-100
75	25	V24-3/3-{-2.1108238, 4.5845491}	-100	-100	100	100	-100	100	100	100	-100	-100
76	26	V25-1/3-{-10.2953971, -4.3570685}	100	100	-100	-100	100	-100	-100	-100	100	100
77	26	V25-2/3-{-4.3570685, 1.5812601}	-100	-100	100	100	-100	100	100	100	-100	-100
78	26	V25-3/3-{-1.5812601, 7.5195887}	100	100	-100	-100	100	-100	-100	-100	100	100
79	27	V26-1/3-{-2.6045506, -0.5639185}	-100	-100	100	100	-100	100	100	100	-100	-100
80	27	V26-2/3-{-0.5639185, 1.4767135}	100	100	-100	-100	100	-100	-100	-100	100	100
81	27	V26-3/3-{-1.4767135, 3.5173456}	100	100	-100	-100	100	-100	-100	-100	100	100
82	28	V27-1/3-{-22.5656793, -4.5063868}	100	100	-100	-100	100	-100	-100	-100	100	100
83	28	V27-2/3-{-4.5063868, 13.5529056}	-100	-100	100	100	-100	100	100	100	-100	-100
84	28	V27-3/3-{-13.5529056, 31.6121981}	-100	-100	100	100	-100	100	100	100	-100	-100
85	29	V28-1/3-{-15.4300839, 0.9958800}	-100	-100	100	100	-100	100	100	100	-100	-100
86	29	V28-2/3-{-0.9958800, 17.4218439}	100	100	-100	-100	100	-100	-100	-100	100	100
87	29	V28-3/3-{-17.4218439, 33.8478078}	-100	-100	100	100	-100	100	100	100	-100	-100
88	30	AMOUNT-1/3-{-0.0000000, 8563.7200000}	100	100	-100	-100	100	-100	-100	-100	100	100
89	30	AMOUNT-2/3-{-8563.7200000, 17127.4400000}	-100	-100	100	100	-100	100	100	100	-100	-100
90	30	AMOUNT-3/3-{-17127.4400000, 25691.1600000}	-100	-100	100	100	-100	100	100	100	-100	-100

В таблице 15 встречается только два значения сходства характеристик транзакций: +100 и -100. Это связано с тем, что в модели всего 2 класса. При большем количестве классов эти уровни сходства меняются от -100 до +100.

Полностью эта матрица сходства значений характеристик транзакций приведена в файле:

c:\Aidos-X\AID_DATA\A0000003\System\SxodAtrInf3.DBF, который формируется после выполнения режима 4.3.2.1. Этот файл, как и все другие dbf-файлы, открывается в MS Excel. Правда это возможно

только в MS Excel 2003, т.к. в более поздних версиях из Экселя к сожалению исключен dbf-xls конвертер. Но можно использовать онлайн конвертеры.

Фрагмент матрицы сходства значений характеристик транзакций отображен на круговой 2d-когнитивной диаграмме, приведенной на рисунке 43:

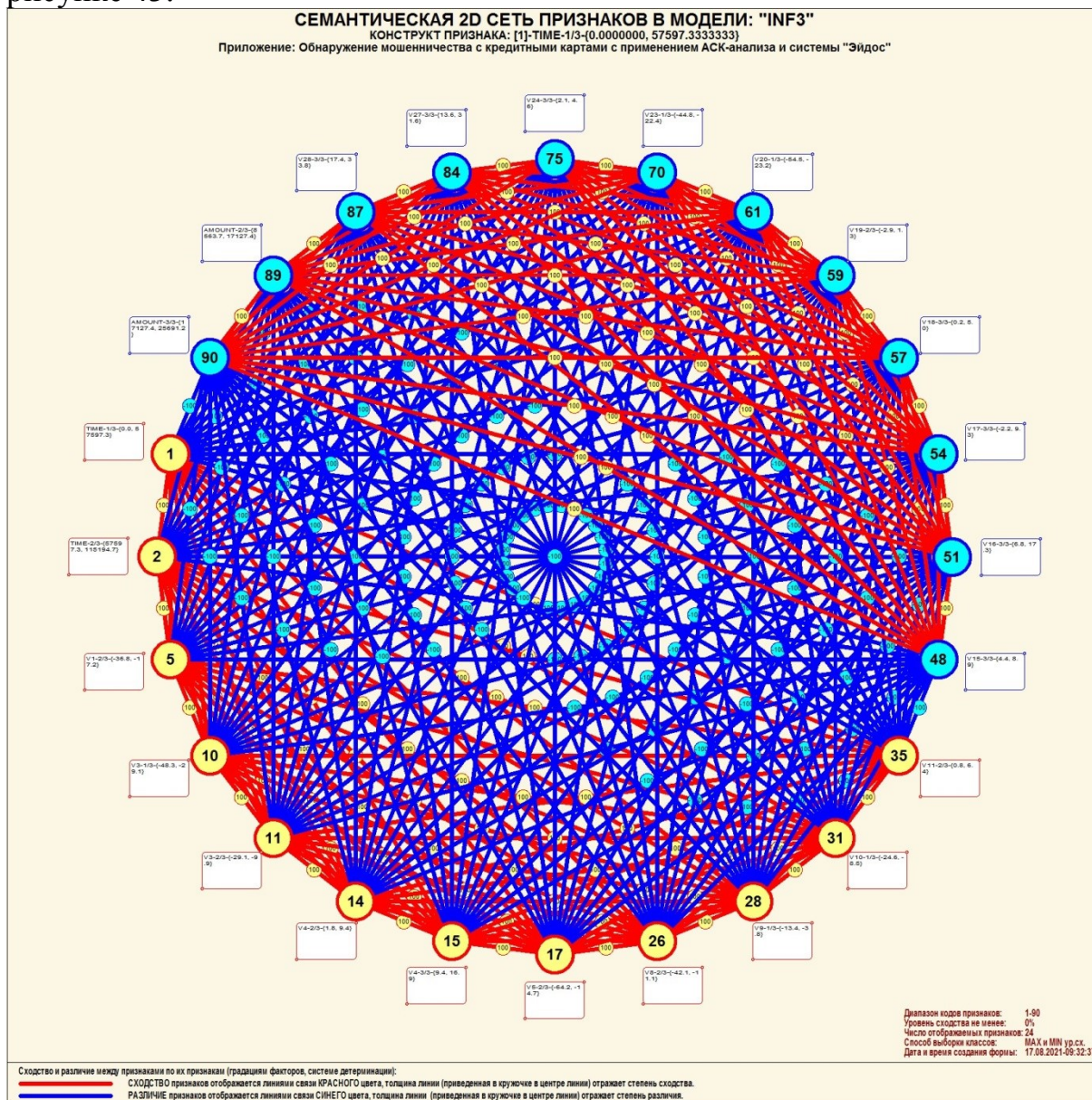


Рисунок 43

На этом рисунке справа внизу приведены значения параметров, при которых была создана данная круговая когнитивная диаграмма. Эти параметры приведены также на рисунке 44.

При отображении на этой диаграмме всех 90 значений характеристик транзакций она является *нечитабельной* даже при использовании поля рисования 4К.

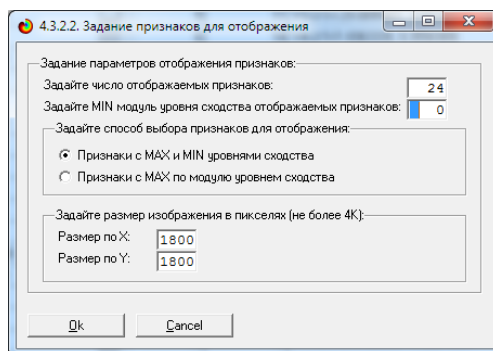


Рисунок 44

С другой стороны в этом нет особой необходимости, т.к. сходство-различие значений характеристик транзакций друг с другом по их влиянию на объект моделирования можно более наглядно и информативно отобразить также в форме когнитивной агломеративной дендрограммы [10] (рисунок 46).

ДЕНДРОГРАММА КОГНИТИВНОЙ КЛАСТЕРИЗАЦИИ ПРИЗНАКОВ В МОДЕЛИ: "INF3"
"Обнаружение мошенничества с кредитными картами с применением АСК-анализа и системы "Эйдос""

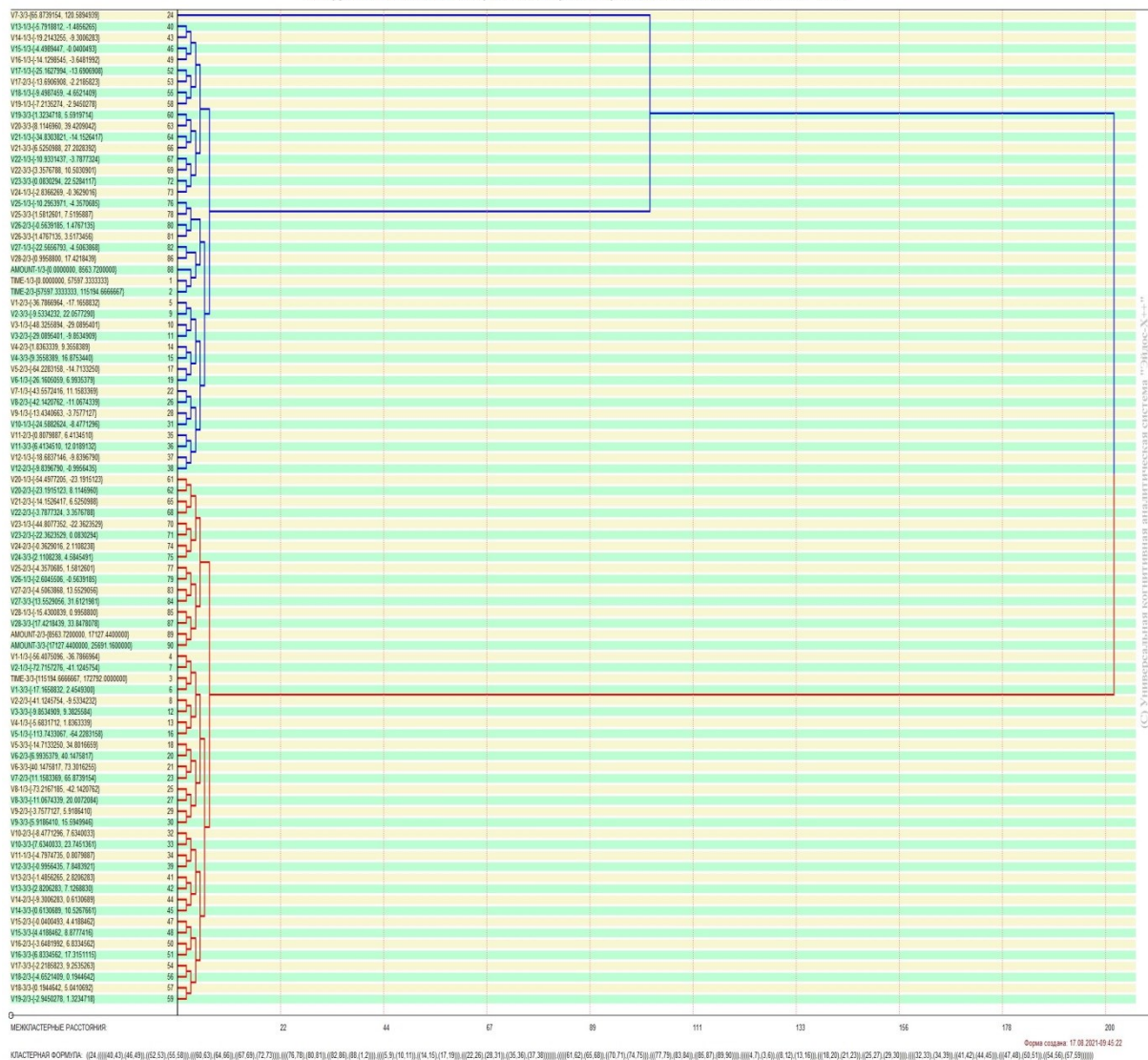


Рисунок 45

График межкластерных расстояний приведен на рисунке 47.

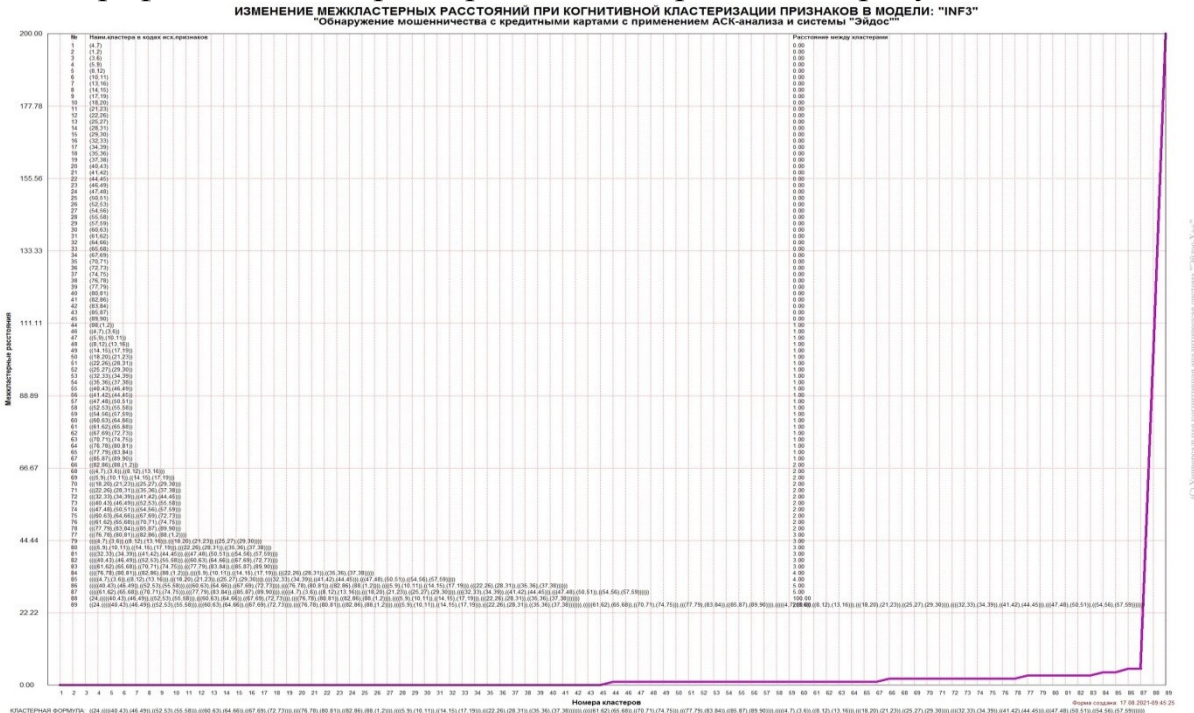


Рисунок 46

Приведенные агломеративная дендрограмма и график межкластерных расстояний созданы в режиме 4.3.2.3 при параметрах, приведенных на рисунке 47.

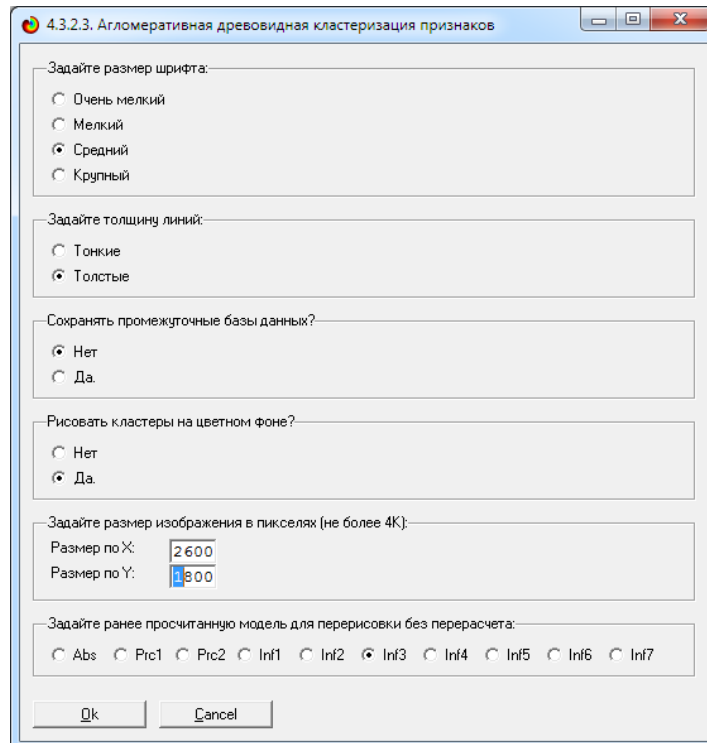


Рисунок 47

Из этой дендрограммы мы видим, что все значения характеристик транзакций образуют два ярко выраженных противоположных по смыслу полюса конструкта, которые представляют собой два противоположные

наиболее крупные кластеры. При этом каждый из кластеров, являющихся полюсами конструктора, состоит из нескольких подкластеров, образующих иерархическую структуру. Обращает на себя внимание максимальное значение характеристики $V7\ 3/3$, стоящая особняком от остальных значений характеристик верхнего кластера (обозначен синим цветом).

5.4. Нелокальные нейроны

Нелокальные нейроны и нелокальная нейронная сеть были предложены Луценко Е.В. в 2003 году [10]. В системе «Эйдос» они были реализованы всегда, с самых первых версий, т.е. еще с 1981 года.

Суть этого предложения состоит в том, что весовые коэффициенты на рецепторах нейронов было предложено не определять методов обратного распространения ошибки, а рассчитывать прямым счетом на основе системной теории информации, предложенной Луценко Е.В. в 2002 году [10].

В настоящее время весьма модными являются нейробайесовские сети, в которых делается практически тоже самое, но не на основе системного обобщения формулы А.Харкевича, а на основе формулы Байеса. При этом нужно отметить, что количество информации по А.Харкевичу по сути представляет собой логарифм формулы Байеса.

Ниже на рисунке 48 приведена экранная форма задания параметров рисования нелокальных нейронов.

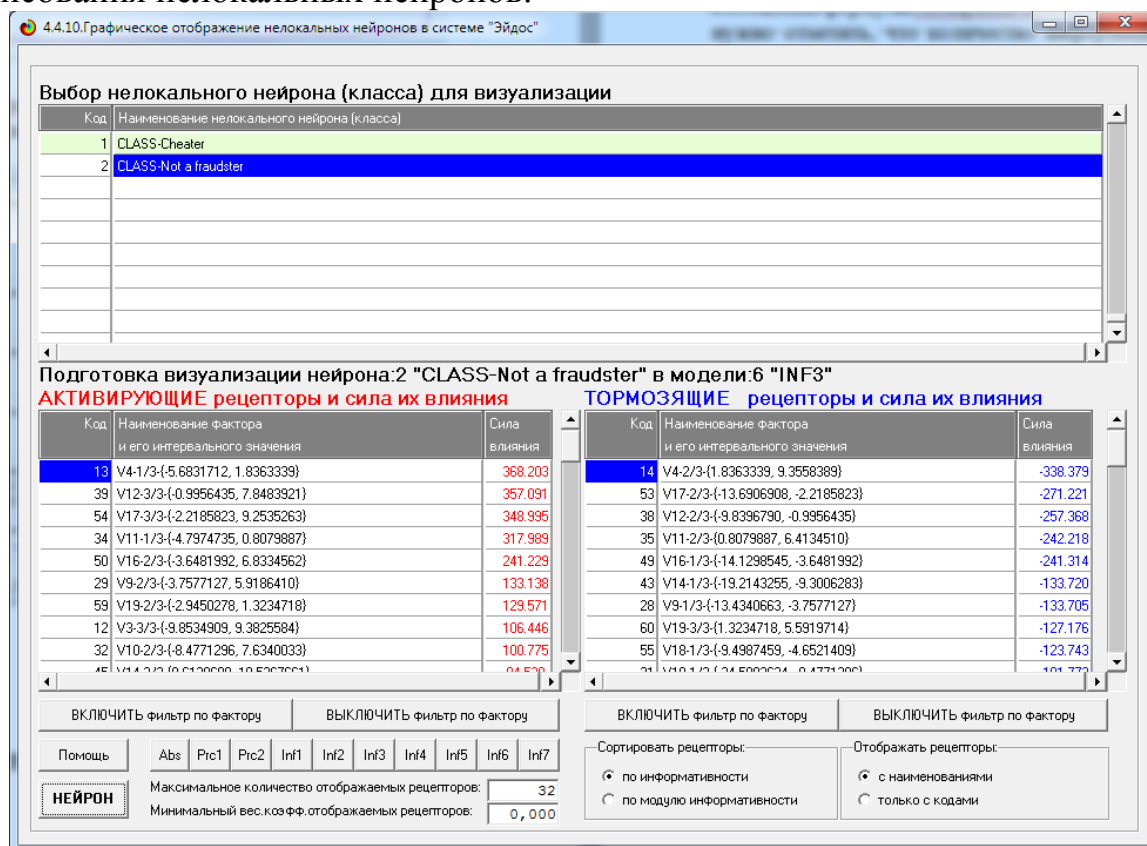


Рисунок 48

Сами изображения нелокальных нейронов для всех классов приведены на рисунках 49.

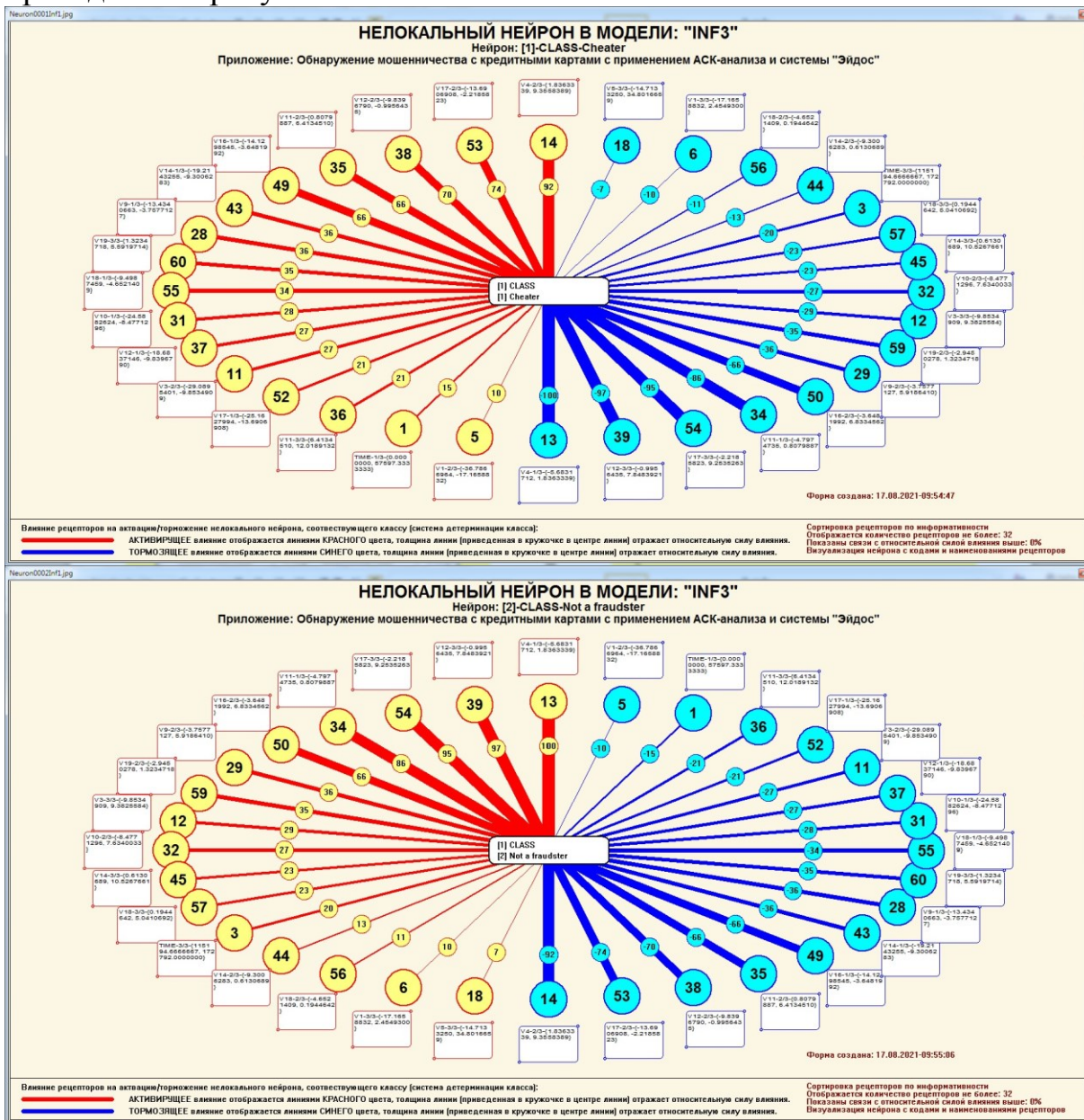


Рисунок 49

По смыслу нелокальные нейроны отображают систему детерминации классов значениям характеристик транзакций, т.е. имеют тот же смысл, что и SWOT-диаграммы классов [10].

В текущей версии системы «Эйдос» нелокальные нейроны отображаются в режиме 4.4.10.

5.5. Нелокальная нейронная сеть

Как уже упоминалось в предыдущем разделе нелокальные нейроны и нелокальная нейронная сеть были предложены Луценко Е.В. в 2003 году [10]. В системе «Эйдос» они были реализованы всегда, с самых первых

версий, т.е. с 1981 года. В системе «Эйдос» один слой нелокальной нейронной сети отображается в режиме 4.4.11. Параметры рисования слоя приведены на рисунке 50, а сама нейронная сеть на рисунке 51.

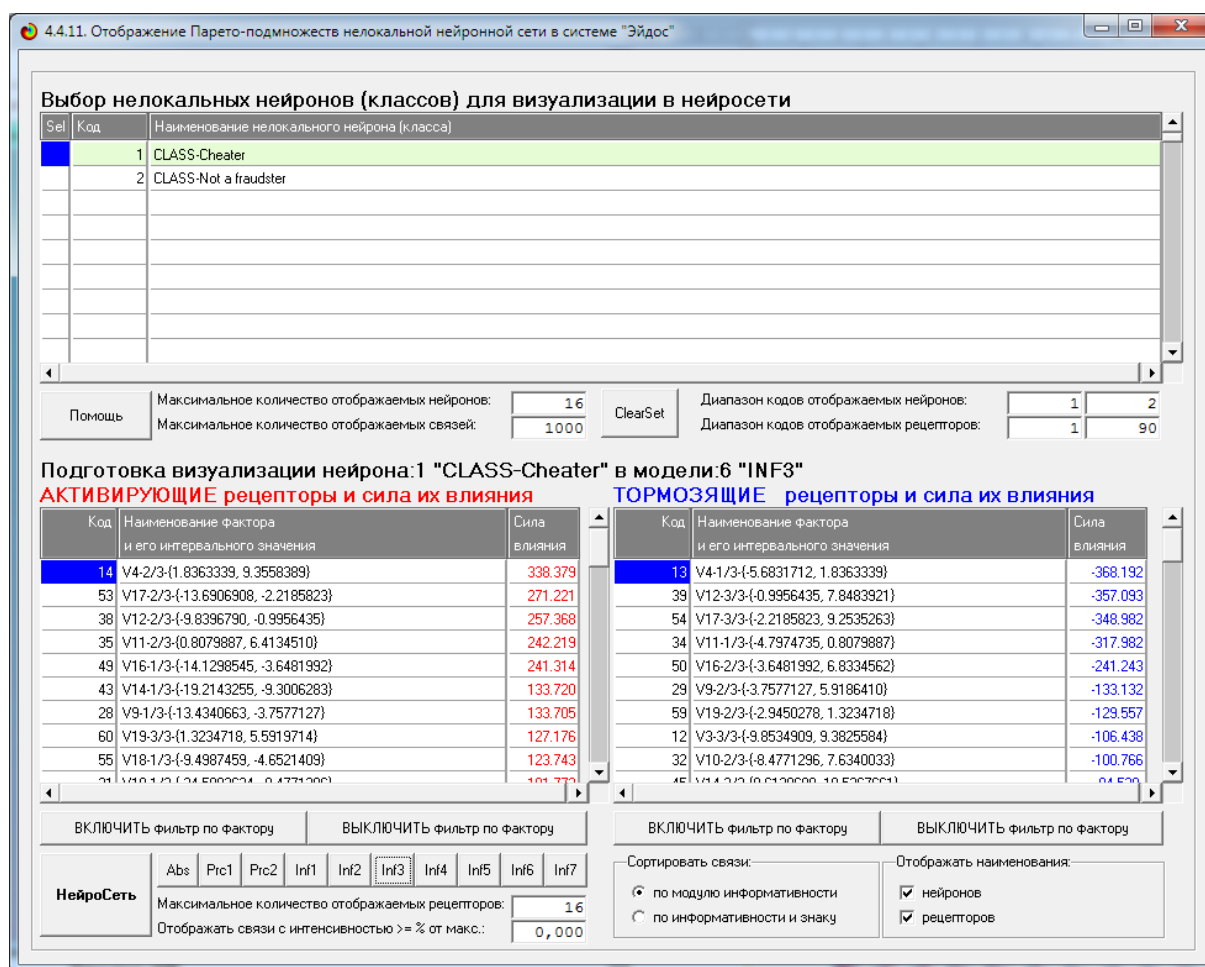


Рисунок 50

На рисунках 51 отображен один слой нейронной сети полностью (вверху), а также читабельный фрагмент сети (внизу), содержащий около 70% наиболее значимых значений характеристик транзакций.

Каждое значение характеристик транзакций влияет на все сейсмические события, соответствующие классам. Это влияние имеет знак (способствует или препятствует) и силу.

Сила влияния каждого значения характеристик транзакций на параметры сейсмической активности (землетрясений), соответствующие классам отображается в виде линии связи, цвет которой означает: способствует (**красный**) или препятствует (**синий**) данное значение характеристик транзакций наступлению события, соответствующего классу. Толщина линии связи соответствует модулю силы влияния.

На рисунках 51 все линии связи рассортированы в порядке убывания модуля силы влияния слева направо.

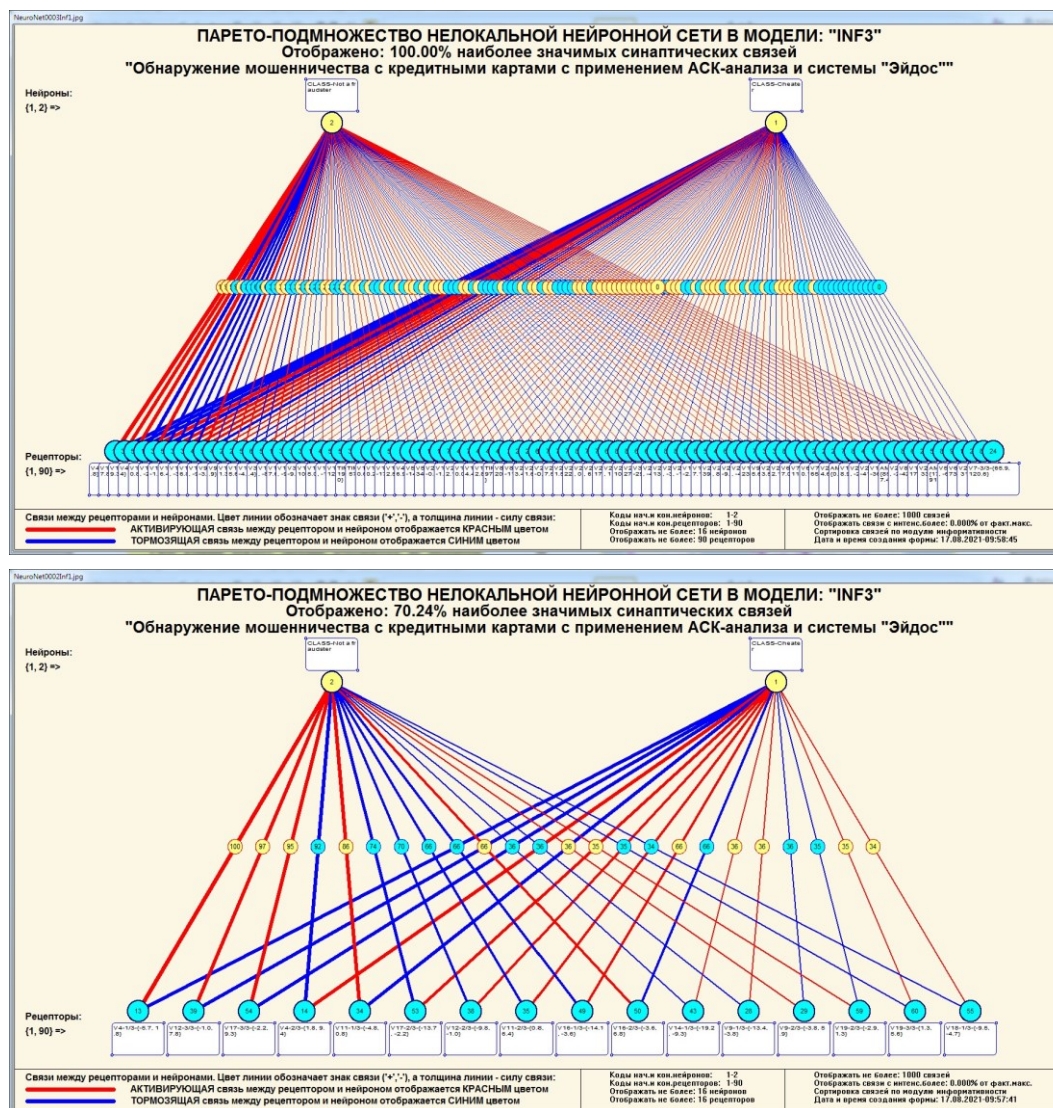
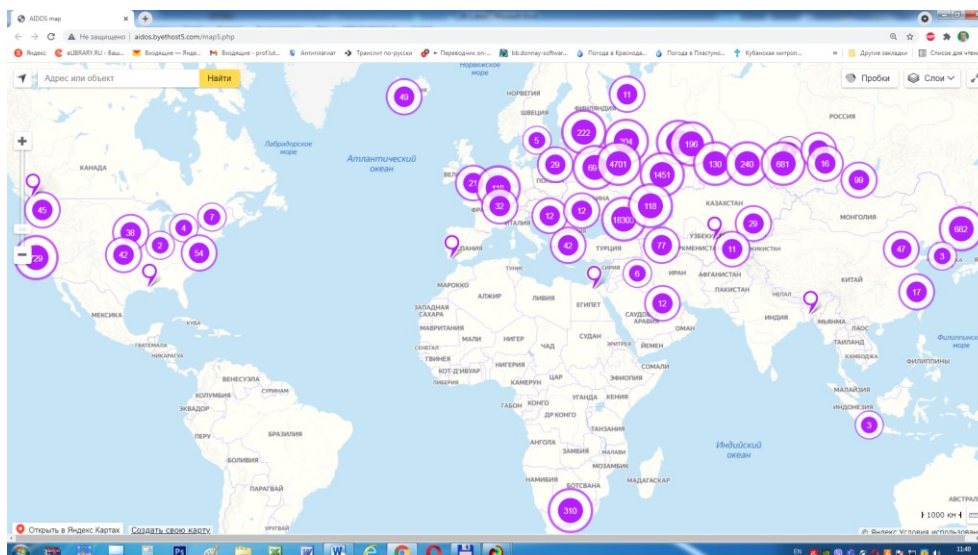


Рисунок 51

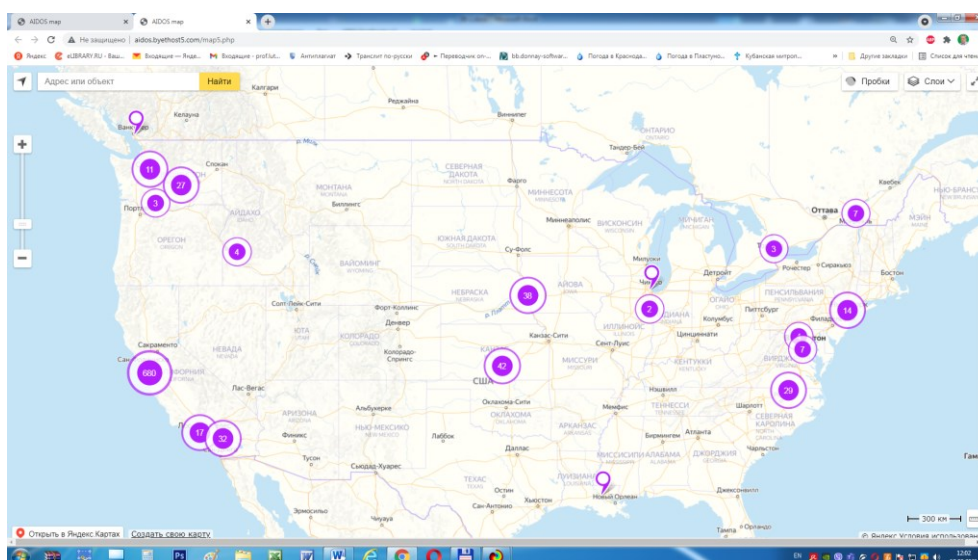
5.6. 3D-интегральные когнитивные карты

В то время как современные ученые пишут о том, что в принципе могут быть созданы автоматизированные системы, отображающие 3d-интегральные когнитивные карты, в системе «Эйдос» эта возможность была реализована всегда, *еще с начала 90-х годов*. Причем все модели в системе «Эйдос» рассчитываются непосредственно на основе эмпирических данных, и выходные формы отображают различные аспекты этих моделей, а не являются визуализацией экспертных оценок, как традиционно.

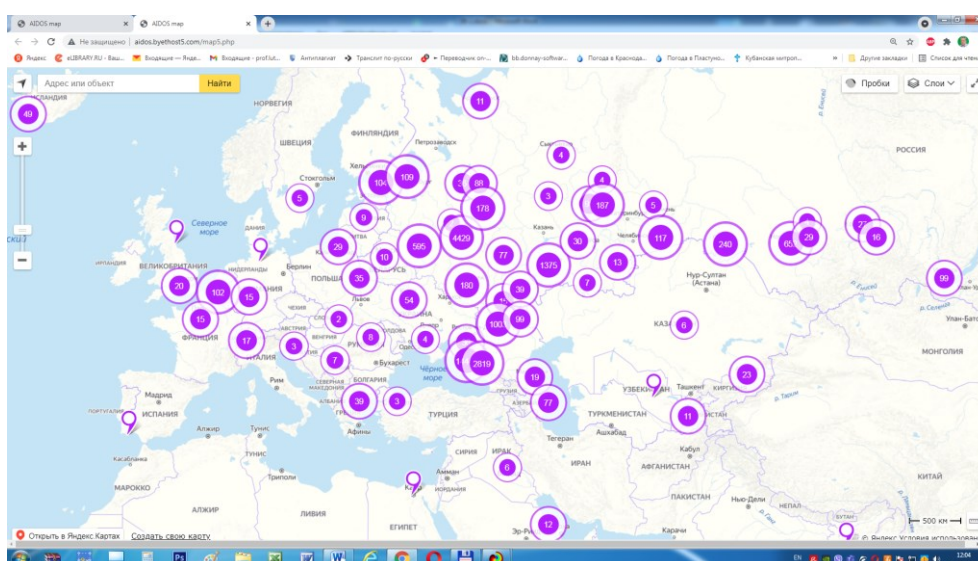
Автор написал об этом: в 647 публикациях, среди которых на момент написания данной работы 31 монография, 31 патент на системы искусственного интеллекта, в том числе первые в России; в нескольких сотнях статей, из которых 332 в изданиях, входящих в перечень ВАК РФ. Система «Эйдос» довольно широко используется практически во всем мире (рисунок 52):



Весь мир



США и Канада



Европа и часть России

Рисунок 52

Однако на данный момент эта информация остается практически неизвестной научной сообществу, если судить ссылкам из публикаций ученых и разработчиков.

3d-интегральной когнитивной карты, по сути, являются изображением на одном рисунке 2d-круговых диаграмм классов и значений факторов, соединенных слоем нейронной сети.

На рисунках ниже приведена экранная форма задания параметров визуализации фрагмента 3d-интегральной когнитивной карты (рисунок 53) и сам фрагмент⁸ 3d-интегральной когнитивной карты (рисунок 54).

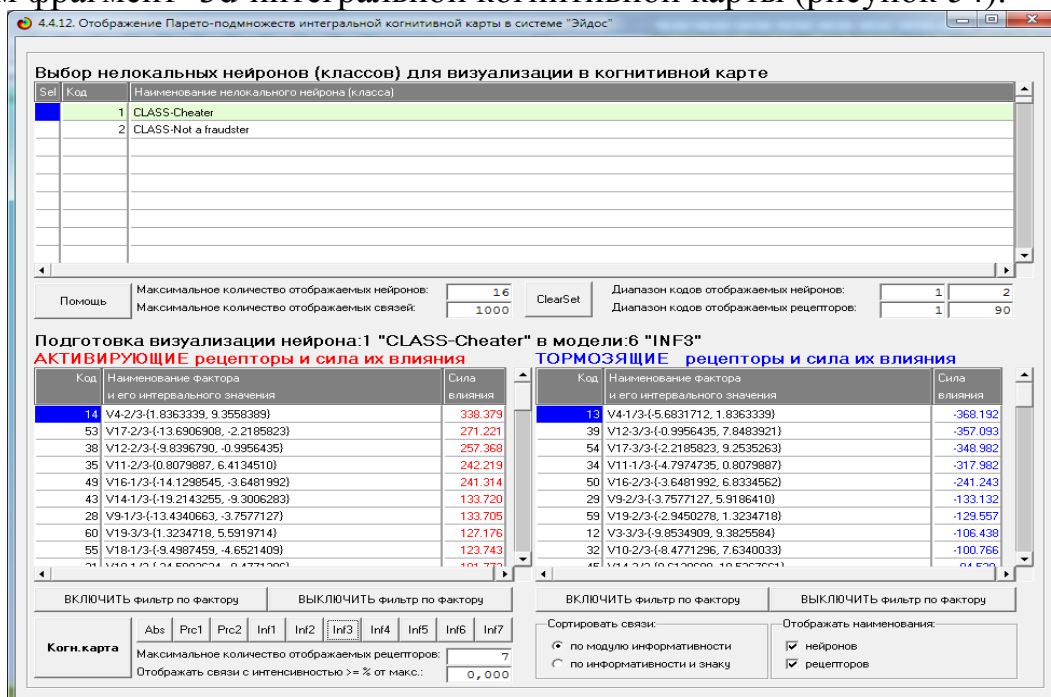


Рисунок 53

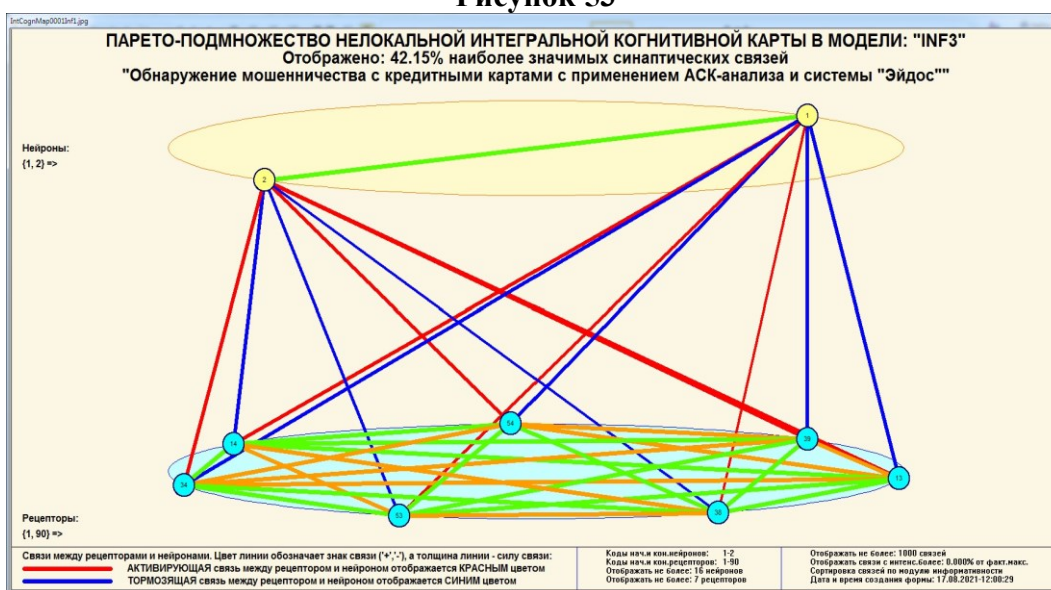


Рисунок 54

⁸ полная карта, как правило, является нечитабельной из-за очень большого объема содержащейся в ней информации

5.7. 2D-интегральные когнитивные карты содержательного сравнения классов

На рисунке 55 приведена 2d-когнитивная диаграмма сравнения классов по системе их детерминации. Из этой диаграммы и матрицы сходства классов (таблица 16), визуализацией которой она является, мы видим, насколько сходны или насколько отличаются друг от друга классы по значениям обуславливающих их факторов.

Однако мы не видим из этой диаграммы, чем именно сходны и чем именно отличаются эти классы по значениям обуславливающих их факторов. Это мы можем увидеть из когнитивной диаграммы содержательного сравнения классов (рисунок 55), которая отображается в режиме 4.2.3 системы «Эйдос».

На рисунке 55 приведена экранная форма задания параметров отображения этих когнитивных диаграмм, на рисунке 56 help этого режима, а на рисунке 57 – пример такой диаграммы.

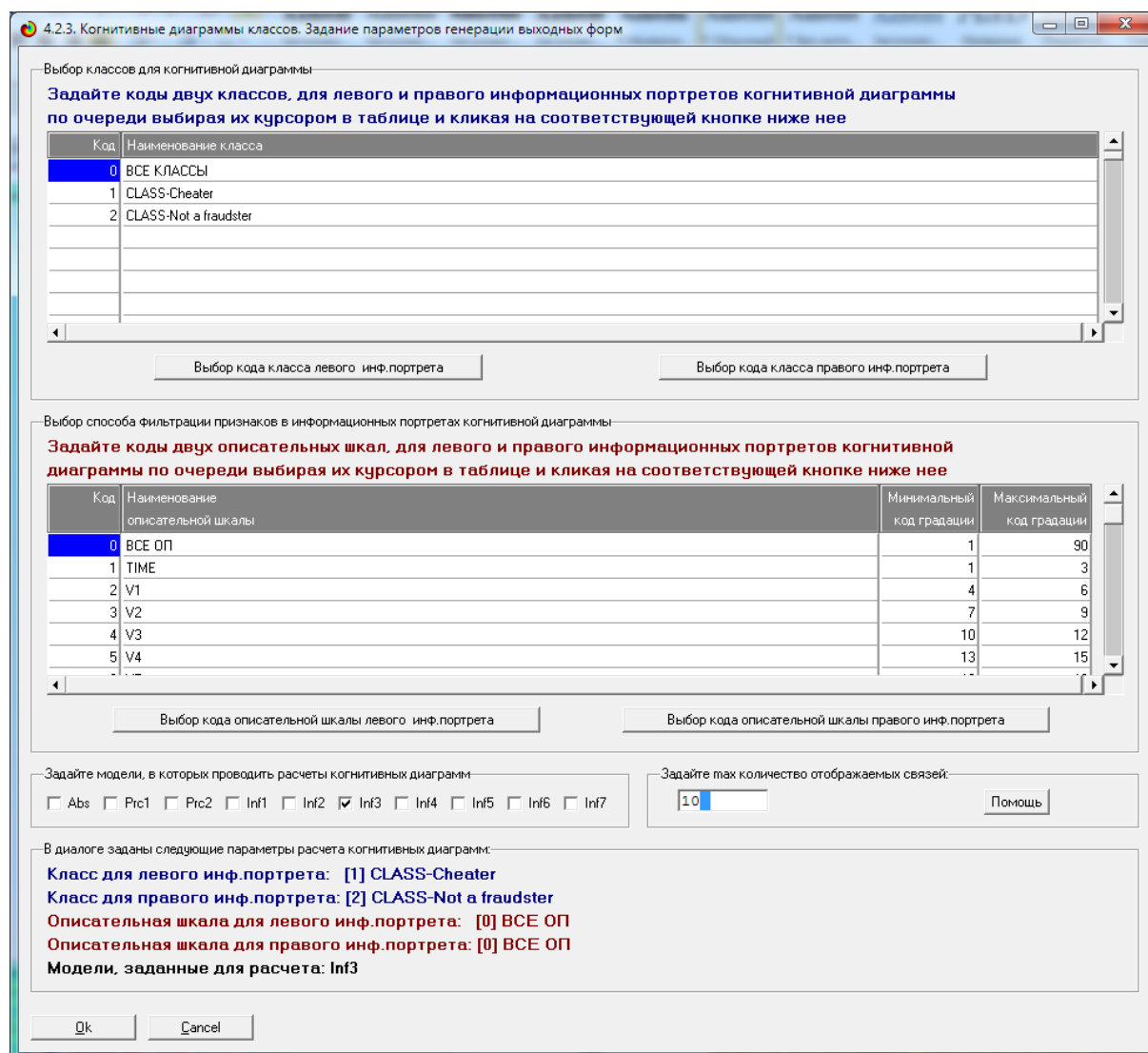


Рисунок 55

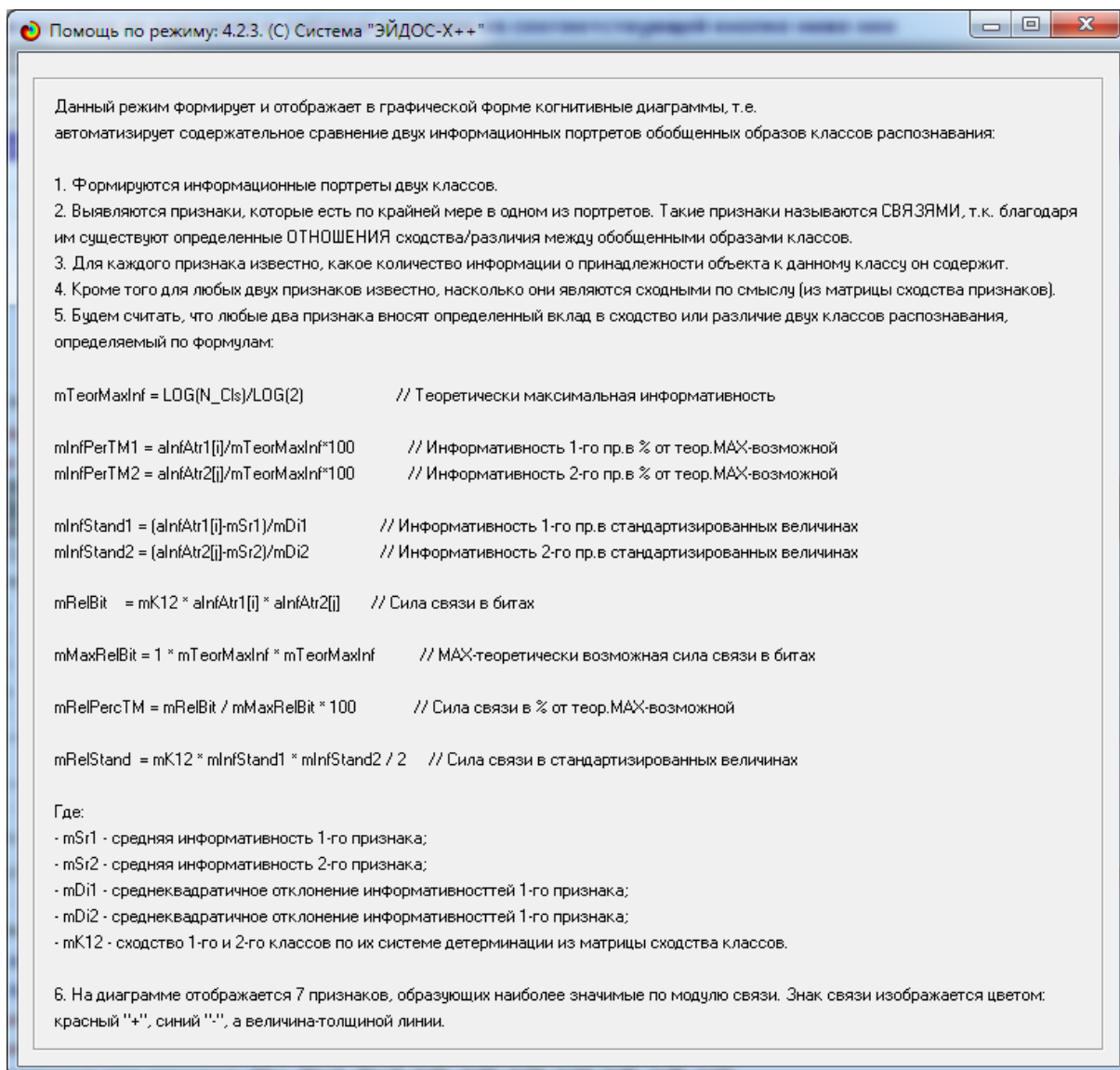


Рисунок 56



Рисунок 57

Когнитивные диаграммы, приведенные на рисунке 57, являются примерами опосредованных нечетких логических заключений. Впервые об автоматизированной реализации рассуждений подобного типа в интеллектуальной системе «Эйдос» написано в 2002 году в работе [10] на странице 521⁹. Позже об этом писалось в работе [10]¹⁰ и ряде других работ автора, поэтому здесь подробнее рассматривать этот вопрос нецелесообразно.

Например, нам известно, что один человек имеет голубые глаза, а другой черные волосы. Спрашивается, эти признаки вносят вклад в сходство или в различие этих двух людей? В АСК-анализе и системе «Эйдос» этот вопрос решается так. В модели на основе кластерно-конструктивного анализа классов и значений факторов (признаков) известно, насколько те или иные признаки сходны или отличаются по их влиянию на объект моделирования. Поэтому понятно, что человек с голубыми глазами вероятнее всего блондин, а брюнет, скорее всего, имеет темные глаза. Так что понятно, что эти признаки вносят вклад в различие этих двух людей.

5.8. 2D-интегральные когнитивные карты содержательного сравнения значений факторов

На рисунке 43 приведена 2d-когнитивная диаграмма сравнения значений факторов по их влиянию на объект моделирования, т.е. на его переходы в состояния, соответствующие классам.

Иначе говоря, из этой диаграммы и матрицы сходства значений факторов (таблица 15), визуализацией которой она является, вполне понятно, насколько сходны или отличаются любые два значения факторов по их смыслу.

Напомним, что смысл, согласно концепции смысла Шенка-Абельсона, используемой в АСК-анализе, состоит в знании причин и последствий [10].

Однако из этой диаграммы не видно, чем именно конкретно сходны или отличаются значения факторов.

Это видно из когнитивных диаграмм, которые можно получить в режиме 4.3.3 системы «Эйдос». Пример такой когнитивной диаграммы приведен на рисунке 59.

На рисунке 58 приведены две экранные формы: форма настроек параметров вывода когнитивных диаграмм и help данного режима.

⁹ https://www.elibrary.ru/download/elibrary_18632909_64818704.pdf, Таблица 7. 17, стр. 521

¹⁰ <http://ej.kubagro.ru/2013/07/pdf/15.pdf>, стр.44.

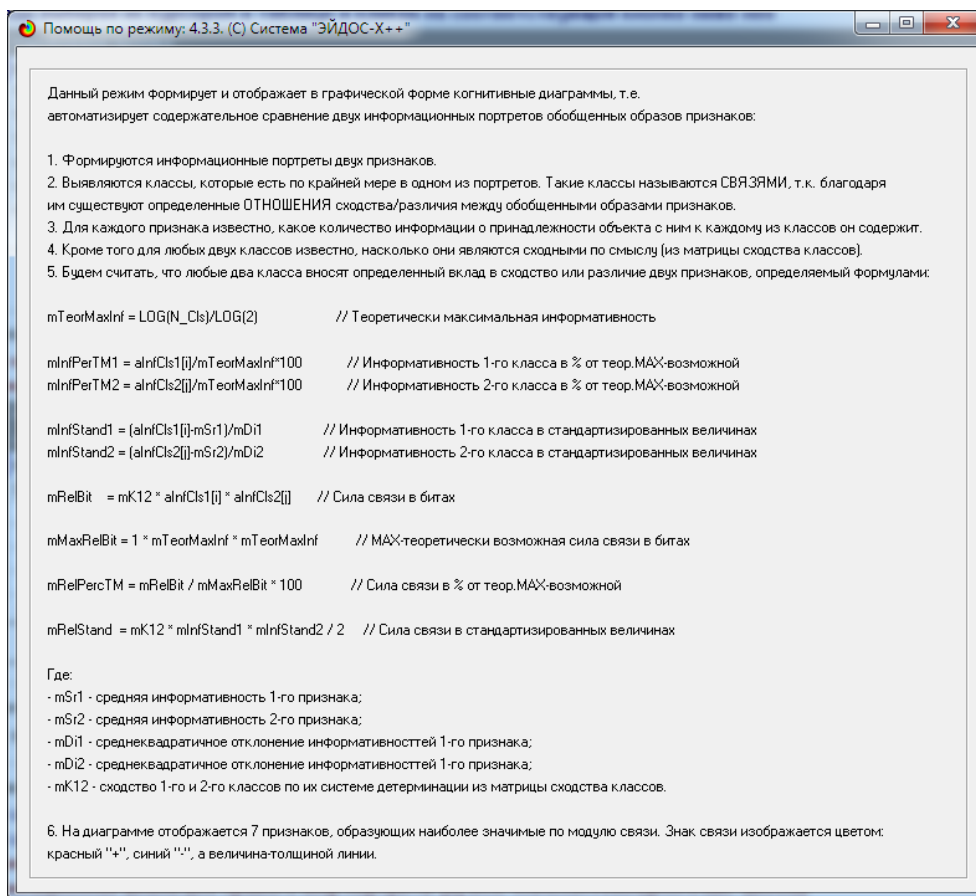
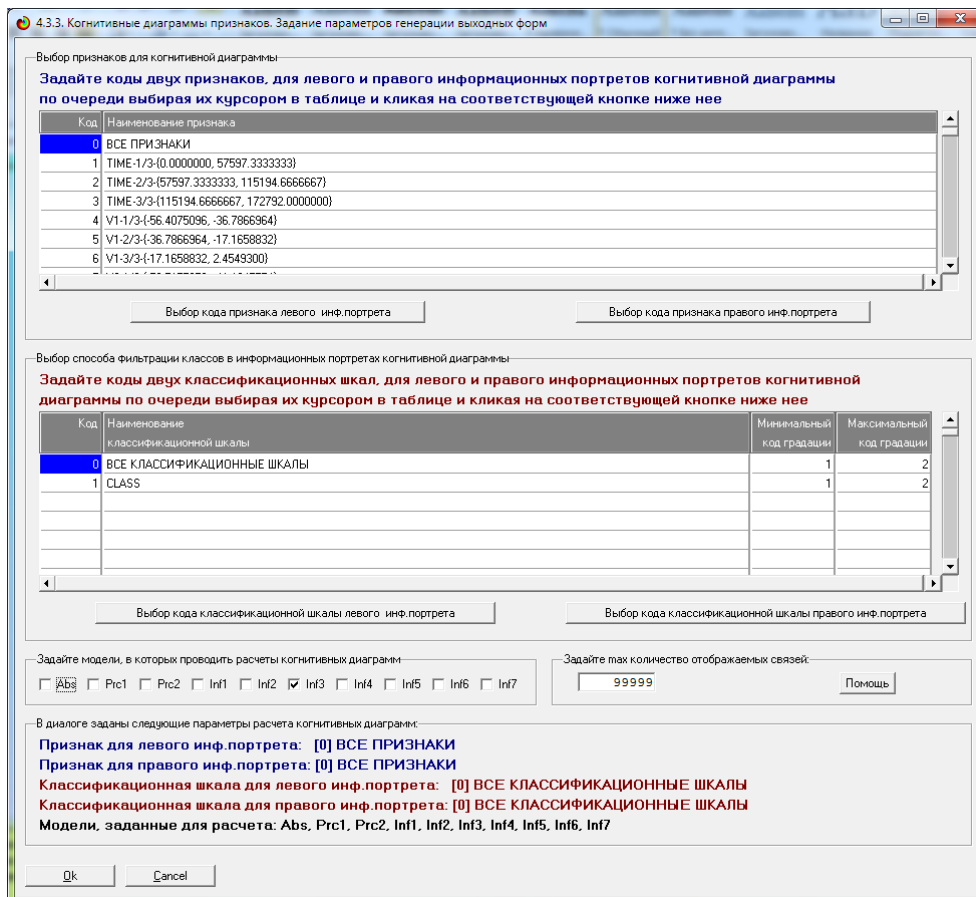


Рисунок 58

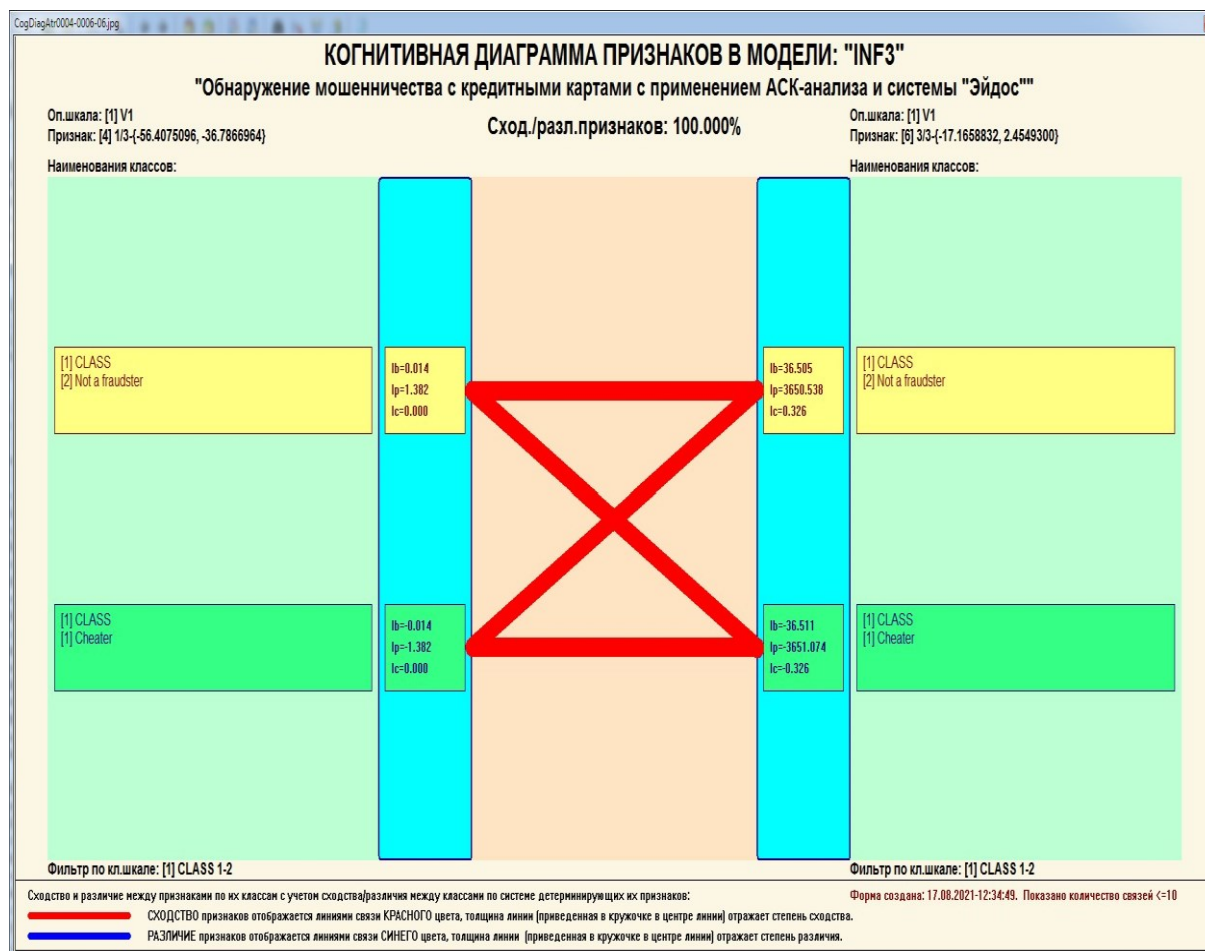


Рисунок 59

5.9. Когнитивные функции

Когнитивные функции являются обобщением классического математического понятия функции на основе системной теории информации и предложены Е.В.Луценко в 2005 году [10].

Когнитивные функции отображают, какое количество информации содержится в градациях описательной шкалы о переходе объекта моделирования в состояния, соответствующие градациям классификационной шкалы. При этом в статистических и системно-когнитивных моделях в каждой градации описательной шкалы содержится информация обо всех градациях классификационной шкалы, т.е. *каждому значению аргумента соответствуют все значения функции, но соответствуют в разной степени, причем как положительной, так и отрицательной, которая отображается цветом.*

В системе «Эйдос» когнитивные функции отображаются в режиме 4.5 (рисунки 60 и 61):

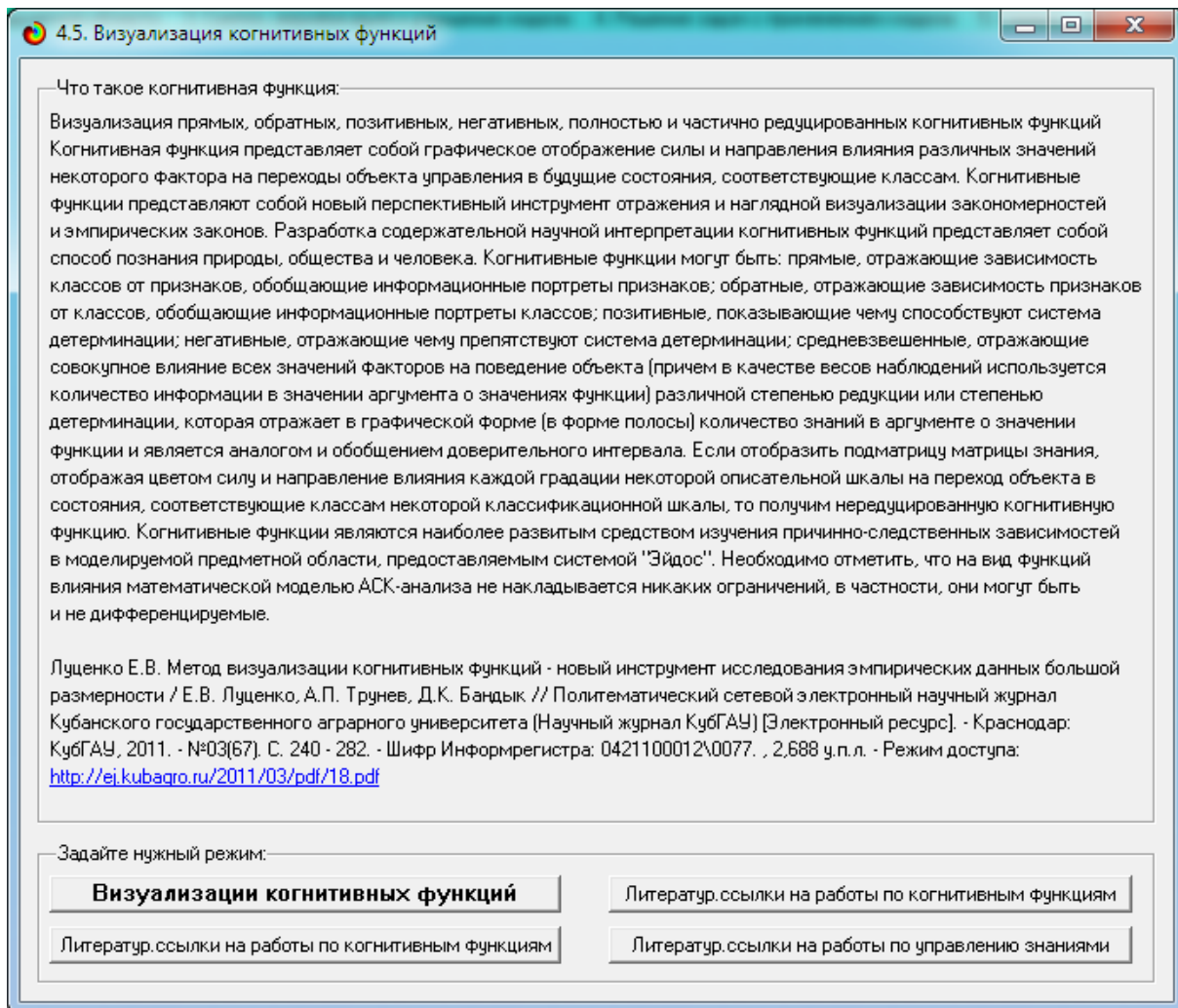
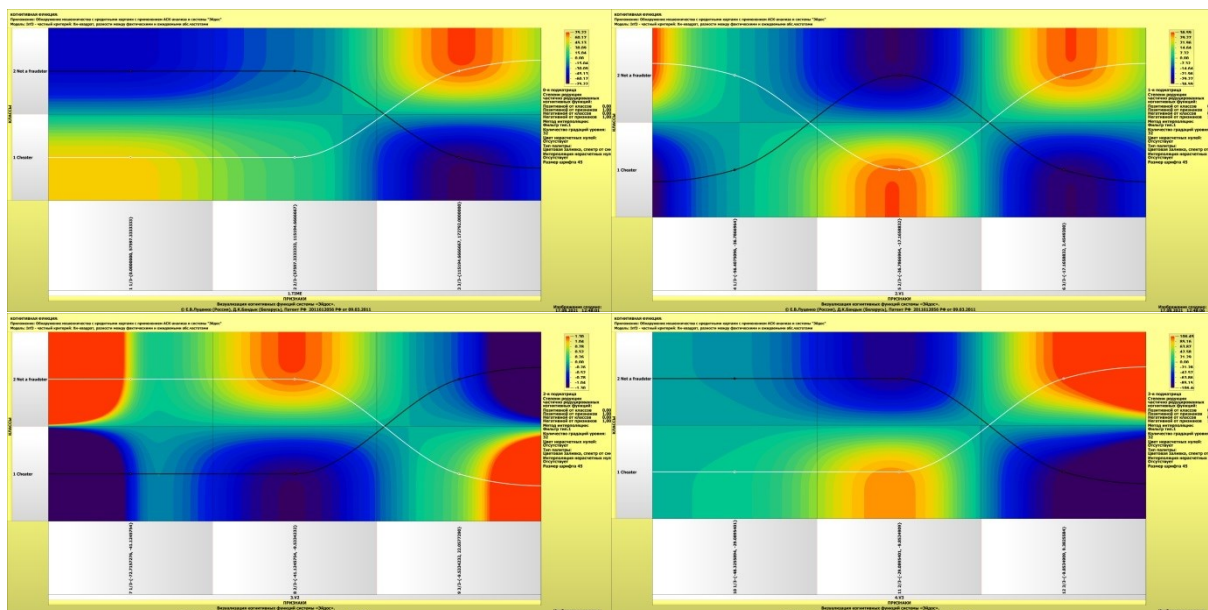
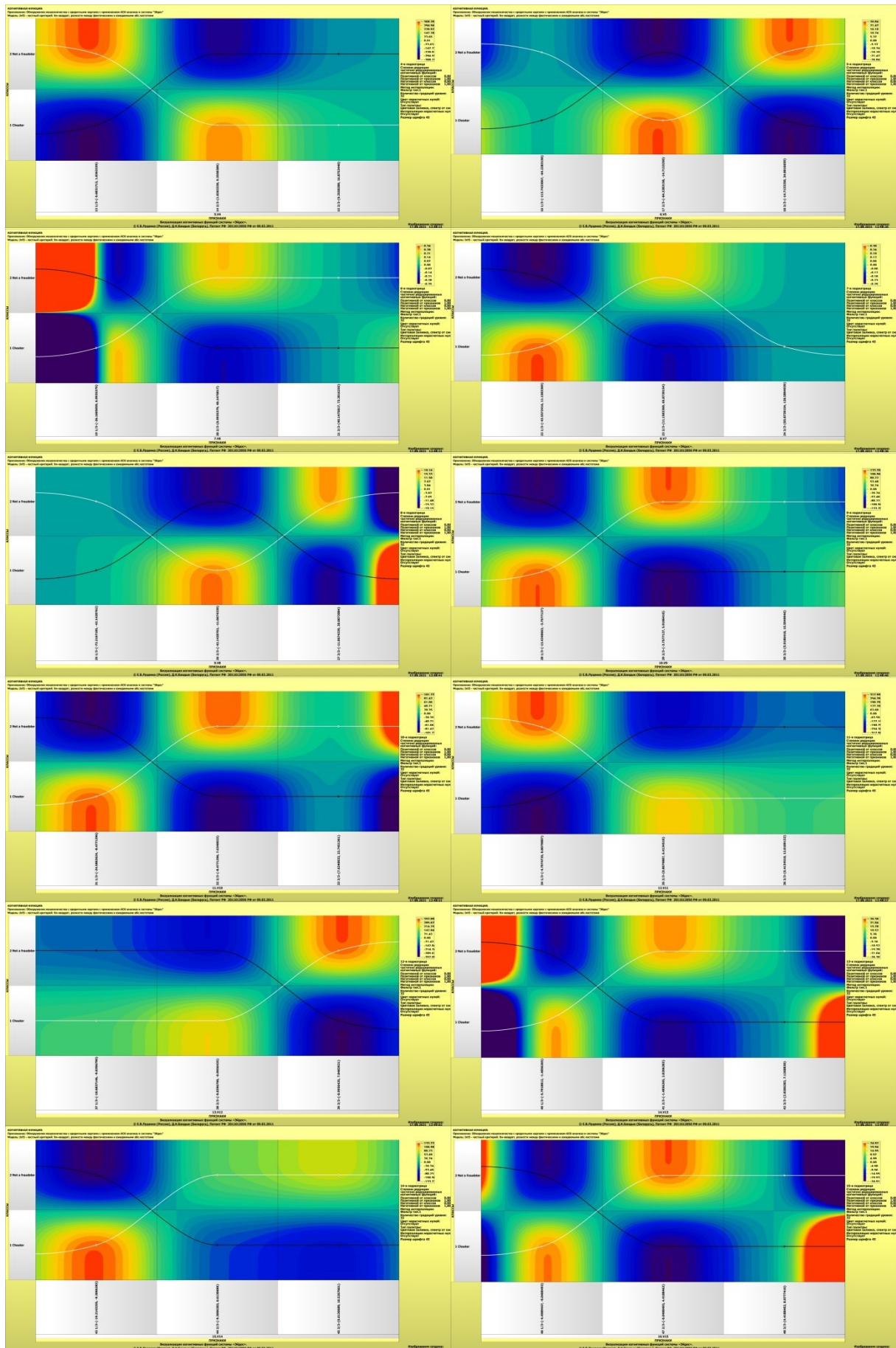
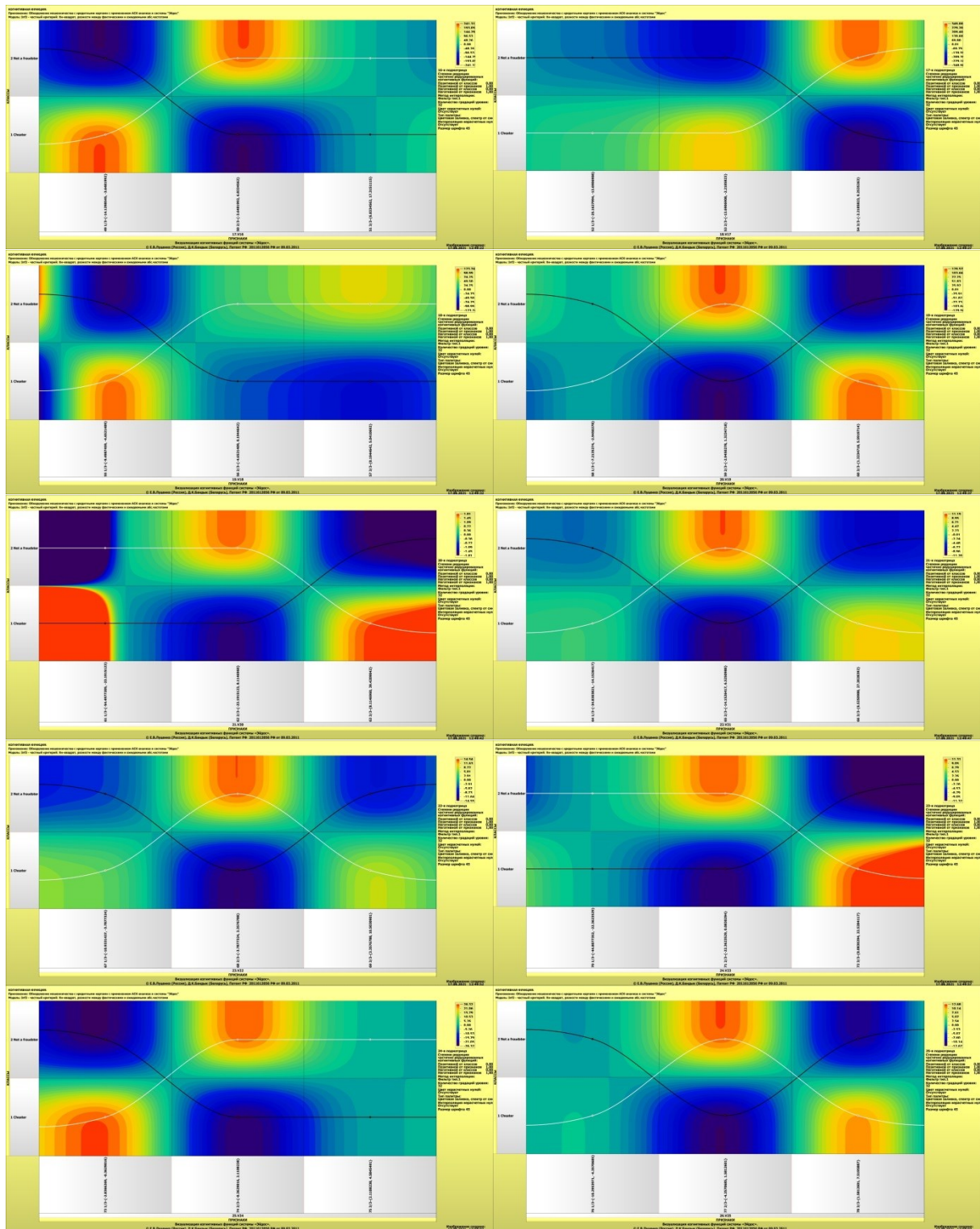


Рисунок 60







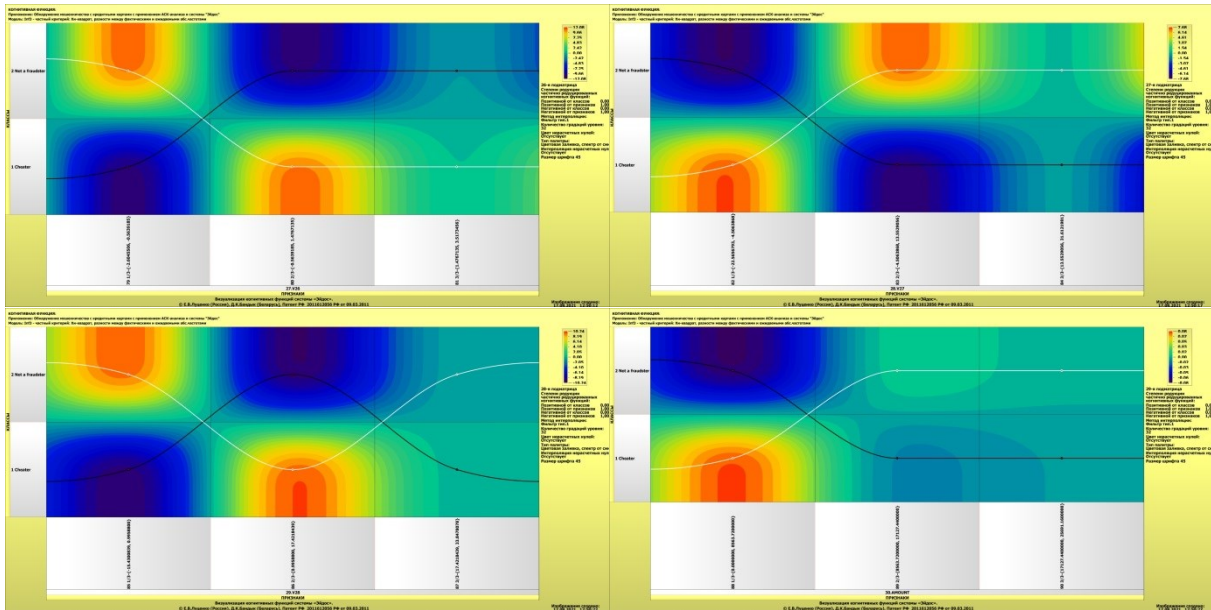
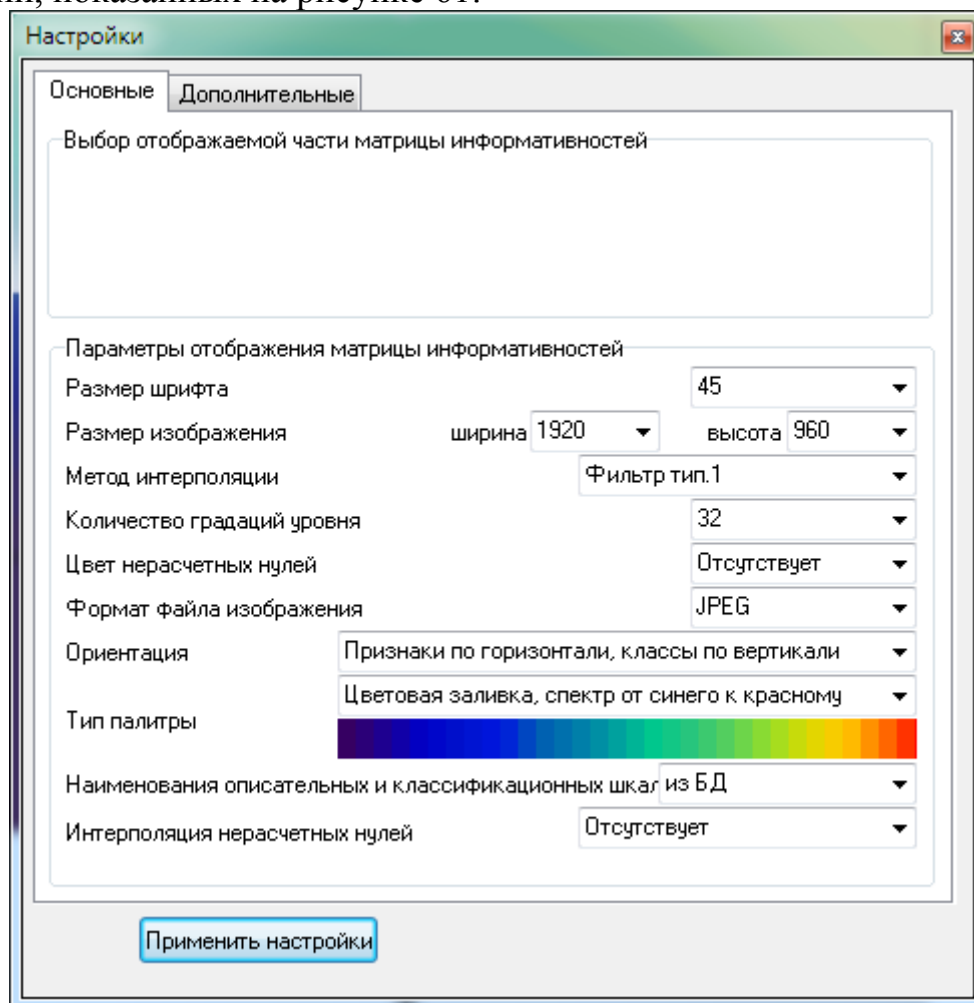


Рисунок 61

На рисунке 62 приведены параметры визуализации когнитивных функций, показанных на рисунке 61:



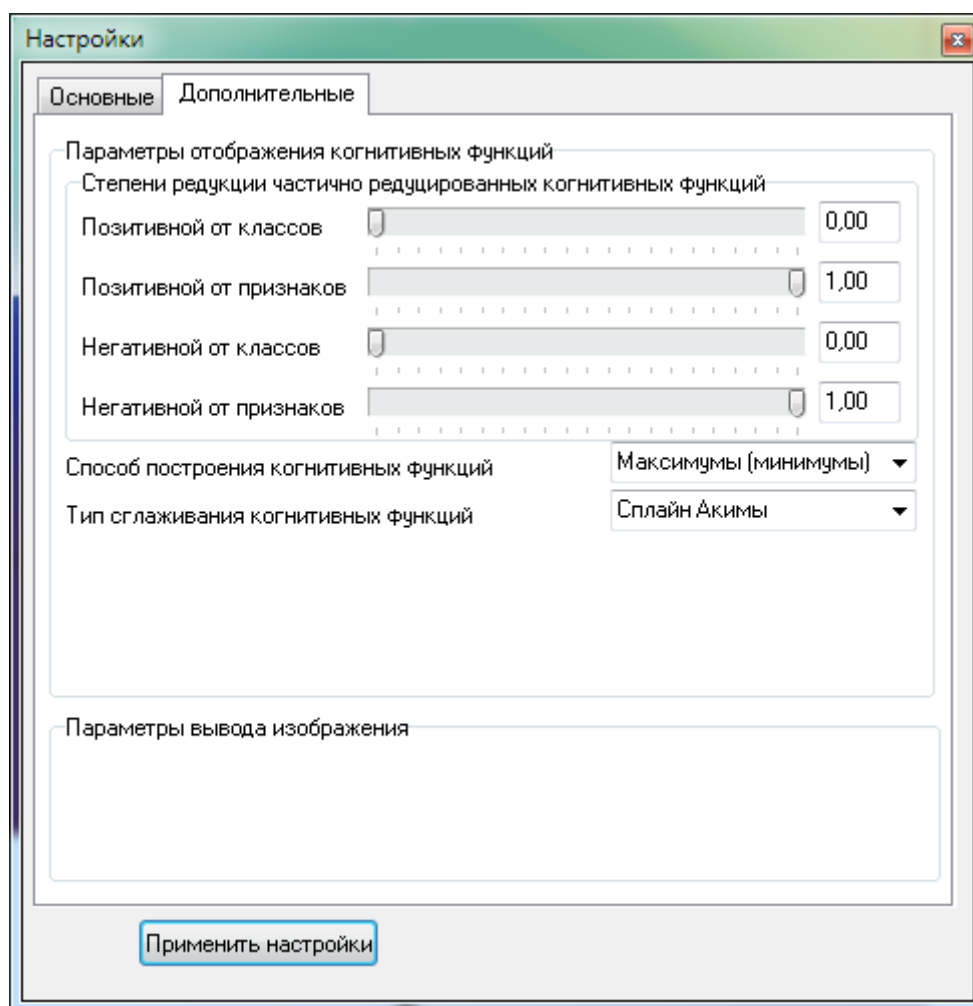


Рисунок 62

Как видно из 1-го рисунка 62, изображения когнитивных функций, приведенные на рисунках 61, получены при достаточно высоком разрешении 1920 на 960 p_{ix}. Поэтому при увеличении масштаба изображения, например до 500%, они вполне читабельны.

5.10. Значимость характеристик транзакций и их значений

В АСК-анализе все факторы рассматриваются с одной единственной точки зрения: сколько информации содержится в их значениях о переходе объекта моделирования и управления, на который они действуют, в определенное будущее состояние, описываемое классом (градация классификационной шкалы), и при этом сила и направление влияния всех значений факторов на объект измеряется в одних общих для всех факторов единицах измерения: единицах количества информации [10].

Значимость (селективная сила) градаций описательных шкал в АСК-анализе – это вариабельность частных критериев в статистических и системно-когнитивных моделях, например в модели $Inf1$, это вариабельность информативностей.

Если рассортировать все градации факторов (признаки) в порядке убывания селективной силы и получить сумму селективной силы системы значений факторов нарастающим итогом, то получим Парето-кривую (рисунки 63 и 64). Это можно сделать в режиме 3.7.5 системы «Эйдос»:

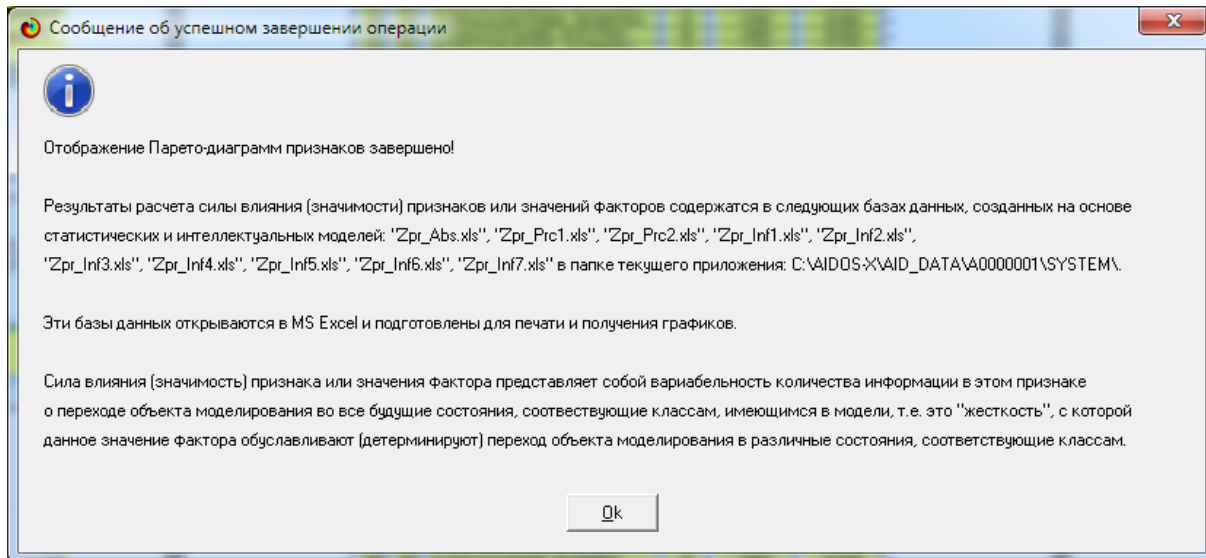


Рисунок 63

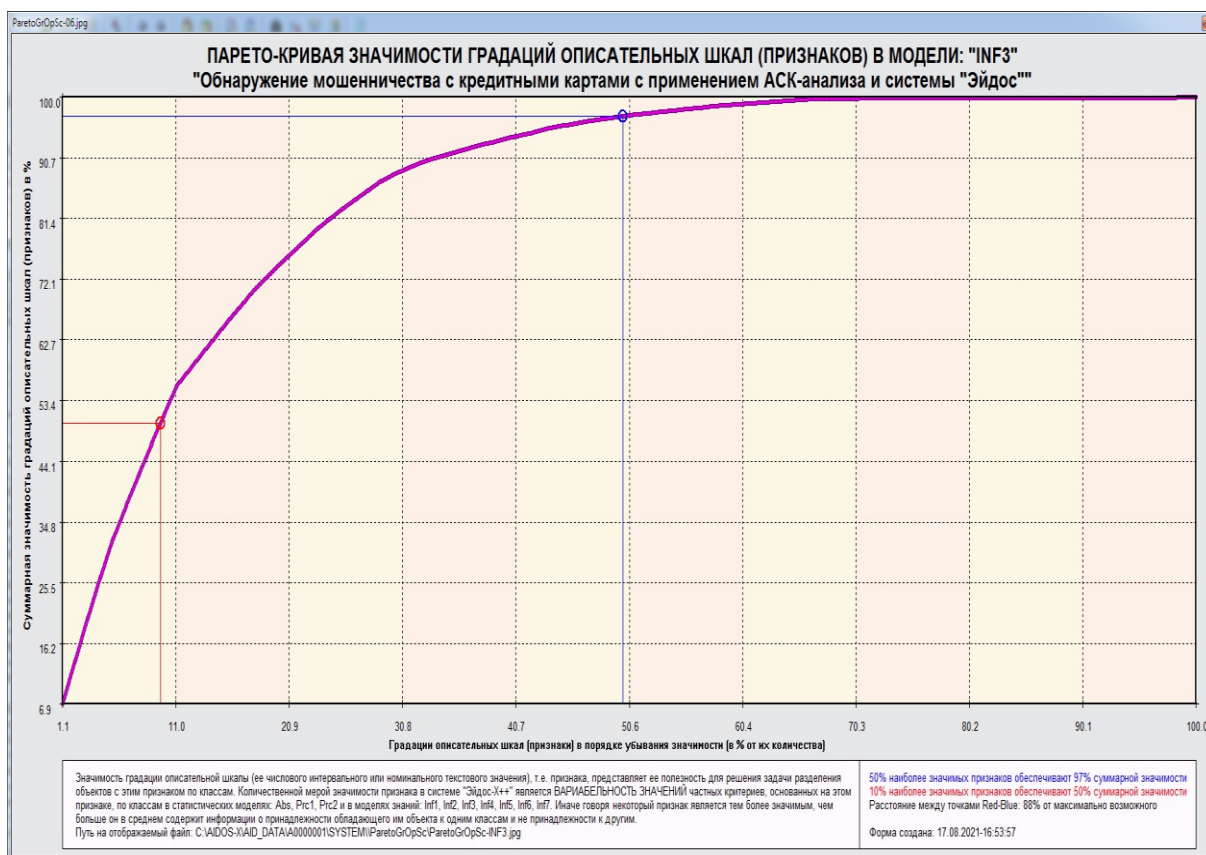


Рисунок 64

В табличной форме эта информация представлена в таблице 16:

Таблица 16 – Значимость значений характеристик транзакций в модели INF3

№	№%	Код	Наименование	Код шкалы	Значимость, %	Значимость, нарастающим итогом, %
1	1,111	13	V4-1/3-{-5.6831712, 1.8363339}	5	6,869	6,869
2	2,222	39	V12-3/3-{-0.9956435, 7.8483921}	13	6,661	13,530
3	3,333	54	V17-3/3-{-2.2185823, 9.2535263}	18	6,510	20,040
4	4,444	14	V4-2/3-{-1.8363339, 9.3558389}	5	6,312	26,353
5	5,556	34	V11-1/3-{-4.7974735, 0.8079887}	12	5,932	32,284
6	6,667	53	V17-2/3-{-13.6906908, -2.2185823}	18	5,060	37,344
7	7,778	38	V12-2/3-{-9.8396790, -0.9956435}	13	4,801	42,145
8	8,889	35	V11-2/3-{-0.8079887, 6.4134510}	12	4,518	46,664
9	10,000	49	V16-1/3-{-14.1298545, -3.6481992}	17	4,502	51,165
10	11,111	50	V16-2/3-{-3.6481992, 6.8334562}	17	4,500	55,665
11	12,222	43	V14-1/3-{-19.2143255, -9.3006283}	15	2,494	58,160
12	13,333	28	V9-1/3-{-13.4340663, -3.7577127}	10	2,494	60,654
13	14,444	29	V9-2/3-{-3.7577127, 5.9186410}	10	2,484	63,138
14	15,556	59	V19-2/3-{-2.9450278, 1.3234718}	20	2,417	65,555
15	16,667	60	V19-3/3-{-1.3234718, 5.5919714}	20	2,372	67,927
16	17,778	55	V18-1/3-{-9.4987459, -4.6521409}	19	2,308	70,235
17	18,889	12	V3-3/3-{-9.8534909, 9.3825584}	4	1,986	72,221
18	20,000	31	V10-1/3-{-24.5882624, -8.4771296}	11	1,899	74,120
19	21,111	32	V10-2/3-{-8.4771296, 7.6340033}	11	1,880	75,999
20	22,222	37	V12-1/3-{-18.6837146, -9.8396790}	13	1,862	77,861
21	23,333	11	V3-2/3-{-29.0895401, -9.8534909}	4	1,838	79,699
22	24,444	45	V14-3/3-{-0.6130689, 10.5267661}	15	1,577	81,276
23	25,556	57	V18-3/3-{-0.1944642, 5.0410692}	19	1,559	82,835
24	26,667	52	V17-1/3-{-25.1627994, -13.6906908}	18	1,452	84,287
25	27,778	36	V11-3/3-{-6.4134510, 12.0189132}	12	1,415	85,701
26	28,889	3	TIME-3/3-{-115194.6666667, 172792.0000000}	1	1,403	87,104
27	30,000	1	TIME-1/3-{-0.0000000, 57597.3333333}	1	1,031	88,135
28	31,111	44	V14-2/3-{-9.3006283, 0.6130689}	15	0,916	89,052
29	32,222	56	V18-2/3-{-4.6521409, 0.1944642}	19	0,748	89,800
30	33,333	5	V1-2/3-{-36.7866964, -17.1658832}	2	0,683	90,483
31	34,444	6	V1-3/3-{-17.1658832, 2.4549300}	2	0,681	91,164
32	35,556	15	V4-3/3-{-9.3558389, 16.8753440}	5	0,557	91,721
33	36,667	17	V5-2/3-{-64.2283158, -14.7133250}	6	0,501	92,222
34	37,778	18	V5-3/3-{-14.7133250, 34.8016659}	6	0,499	92,721
35	38,889	73	V24-1/3-{-2.8366269, -0.3629016}	25	0,491	93,212
36	40,000	40	V13-1/3-{-5.7918812, -1.4856265}	14	0,491	93,703
37	41,111	74	V24-2/3-{-0.3629016, 2.1108238}	25	0,486	94,189
38	42,222	46	V15-1/3-{-4.4989447, -0.0400493}	16	0,465	94,653
39	43,333	47	V15-2/3-{-0.0400493, 4.4188462}	16	0,463	95,117
40	44,444	41	V13-2/3-{-1.4856265, 2.8206283}	14	0,445	95,561
41	45,556	2	TIME-2/3-{-57597.3333333, 115194.6666667}	1	0,373	95,934
42	46,667	27	V8-3/3-{-11.0674339, 20.0072084}	9	0,357	96,292
43	47,778	26	V8-2/3-{-42.1420762, -11.0674339}	9	0,340	96,632
44	48,889	68	V22-2/3-{-3.7877324, 3.3576788}	23	0,271	96,903
45	50,000	77	V25-2/3-{-4.3570685, 1.5812601}	26	0,236	97,139
46	51,111	79	V26-1/3-{-2.6045506, -0.5639185}	27	0,225	97,365
47	52,222	78	V25-3/3-{-1.5812601, 7.5195887}	26	0,219	97,584
48	53,333	80	V26-2/3-{-0.5639185, 1.4767135}	27	0,218	97,802
49	54,444	72	V23-3/3-{-0.0830294, 22.5284117}	24	0,211	98,014
50	55,556	71	V23-2/3-{-22.3623529, 0.0830294}	24	0,209	98,223
51	56,667	65	V21-2/3-{-14.1526417, 6.5250988}	22	0,209	98,432
52	57,778	86	V28-2/3-{-0.9958800, 17.4218439}	29	0,191	98,623
53	58,889	85	V28-1/3-{-15.4300839, 0.9958800}	29	0,190	98,813
54	60,000	69	V22-3/3-{-3.3576788, 10.5030901}	23	0,165	98,978
55	61,111	66	V21-3/3-{-6.5250988, 27.2028392}	22	0,155	99,133
56	62,222	10	V3-1/3-{-48.3255894, -29.0895401}	4	0,149	99,281
57	63,333	82	V27-1/3-{-22.5656793, -4.5063868}	28	0,143	99,425
58	64,444	83	V27-2/3-{-4.5063868, 13.5529056}	28	0,142	99,566
59	65,556	67	V22-1/3-{-10.9331437, -3.7877324}	23	0,107	99,674
60	66,667	64	V21-1/3-{-34.8303821, -14.1526417}	22	0,055	99,729
61	67,778	58	V19-1/3-{-7.2135274, -2.9450278}	20	0,046	99,775
62	68,889	42	V13-3/3-{-2.8206283, 7.1268830}	14	0,045	99,820
63	70,000	63	V20-3/3-{-8.1146960, 39.4209042}	21	0,034	99,854
64	71,111	62	V20-2/3-{-23.1915123, 8.1146960}	21	0,032	99,886
65	72,222	8	V2-2/3-{-41.1245754, -9.5334232}	3	0,024	99,910
66	73,333	76	V25-1/3-{-10.2953971, -4.3570685}	26	0,018	99,928
67	74,444	33	V10-3/3-{-7.6340033, 23.7451361}	11	0,018	99,946
68	75,556	30	V9-3/3-{-5.9186410, 15.5949946}	10	0,009	99,955
69	76,667	81	V26-3/3-{-1.4767135, 3.5173456}	27	0,008	99,964
70	77,778	9	V2-3/3-{-9.5334232, 22.0577290}	3	0,007	99,971
71	78,889	19	V6-1/3-{-26.1605059, 6.9935379}	7	0,006	99,977
72	80,000	22	V7-1/3-{-43.5572416, 11.1583369}	8	0,006	99,983
73	81,111	20	V6-2/3-{-6.9935379, 40.1475817}	7	0,005	99,988
74	82,222	23	V7-2/3-{-11.1583369, 65.8739154}	8	0,004	99,992
75	83,333	75	V24-3/3-{-2.1108238, 4.5845491}	25	0,004	99,996
76	84,444	88	AMOUNT-1/3-{-0.0000000, 8563.7200000}	30	0,002	99,998
77	85,556	48	V15-3/3-{-4.4188462, 8.8777416}	16	0,000	99,998
78	86,667	70	V23-1/3-{-44.8077352, -22.3623529}	24	0,000	99,999
79	87,778	7	V2-1/3-{-72.7157276, -41.1245754}	3	0,000	99,999
80	88,889	4	V1-1/3-{-56.4075096, -36.7866964}	2	0,000	99,999
81	90,000	89	AMOUNT-2/3-{-8563.7200000, 17127.4400000}	30	0,000	99,999

Значимость (селективная сила) описательных шкал в АСК-анализе – среднее значимости их градаций.

Если рассортировать все описательные шкалы в порядке убывания средней селективной силы их градаций в разных моделях, то получим информацию о значимости описательных шкал. В режиме 3.7.4 системы «Эйдос» эта информация заносится в перечисленные на рисунке 65 базы данных, которые открываются в MS Excel и оформлены как выходные формы с наименованиями шкал (таблица 17).

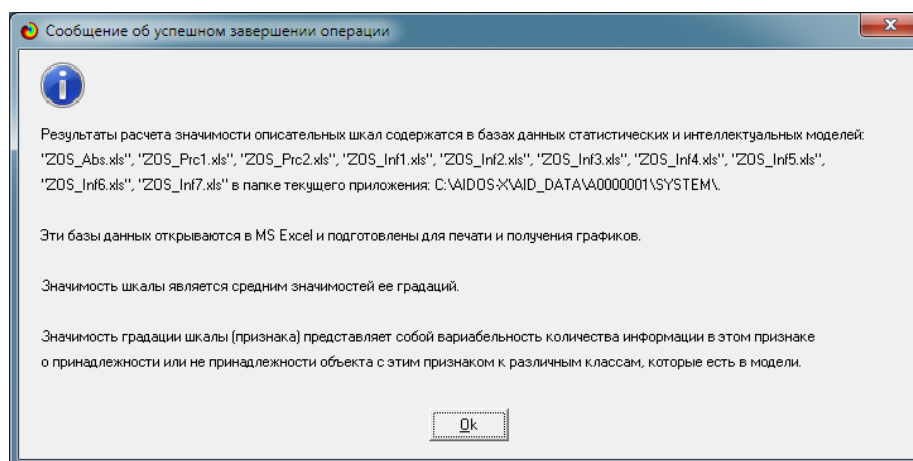


Рисунок 65

Таблица 17 – Значимость характеристик транзакций в модели INF3

№	№%	Код	Наименование	Значимость, %	Значимость, Нарастающим итогом, %
1	3,333	5	V4	13,738	13,738
2	6,667	13	V12	13,324	27,062
3	10,000	18	V17	13,022	40,084
4	13,333	12	V11	11,865	51,949
5	16,667	17	V16	9,002	60,951
6	20,000	15	V14	4,988	65,939
7	23,333	10	V9	4,987	70,926
8	26,667	20	V19	4,835	75,761
9	30,000	19	V18	4,615	80,376
10	33,333	4	V3	3,972	84,349
11	36,667	11	V10	3,796	88,145
12	40,000	1	TIME	2,808	90,952
13	43,333	2	V1	1,364	92,316
14	46,667	6	V5	1,000	93,316
15	50,000	25	V24	0,981	94,297
16	53,333	14	V13	0,980	95,277
17	56,667	16	V15	0,929	96,205
18	60,000	9	V8	0,697	96,903
19	63,333	23	V22	0,544	97,447
20	66,667	26	V25	0,474	97,921
21	70,000	27	V26	0,452	98,373
22	73,333	24	V23	0,421	98,794
23	76,667	22	V21	0,419	99,213
24	80,000	29	V28	0,381	99,593
25	83,333	28	V27	0,285	99,879
26	86,667	21	V20	0,066	99,945
27	90,000	3	V2	0,032	99,977
28	93,333	7	V6	0,012	99,988
29	96,667	8	V7	0,010	99,998
30	100,000	30	AMOUNT	0,002	100,000

На рисунке 66 приведен график значимости характеристик транзакций для оценки риска мошенничества по операциям с кредитными картами нарастающим итогом (в %). График построен по данным таблицы 17.

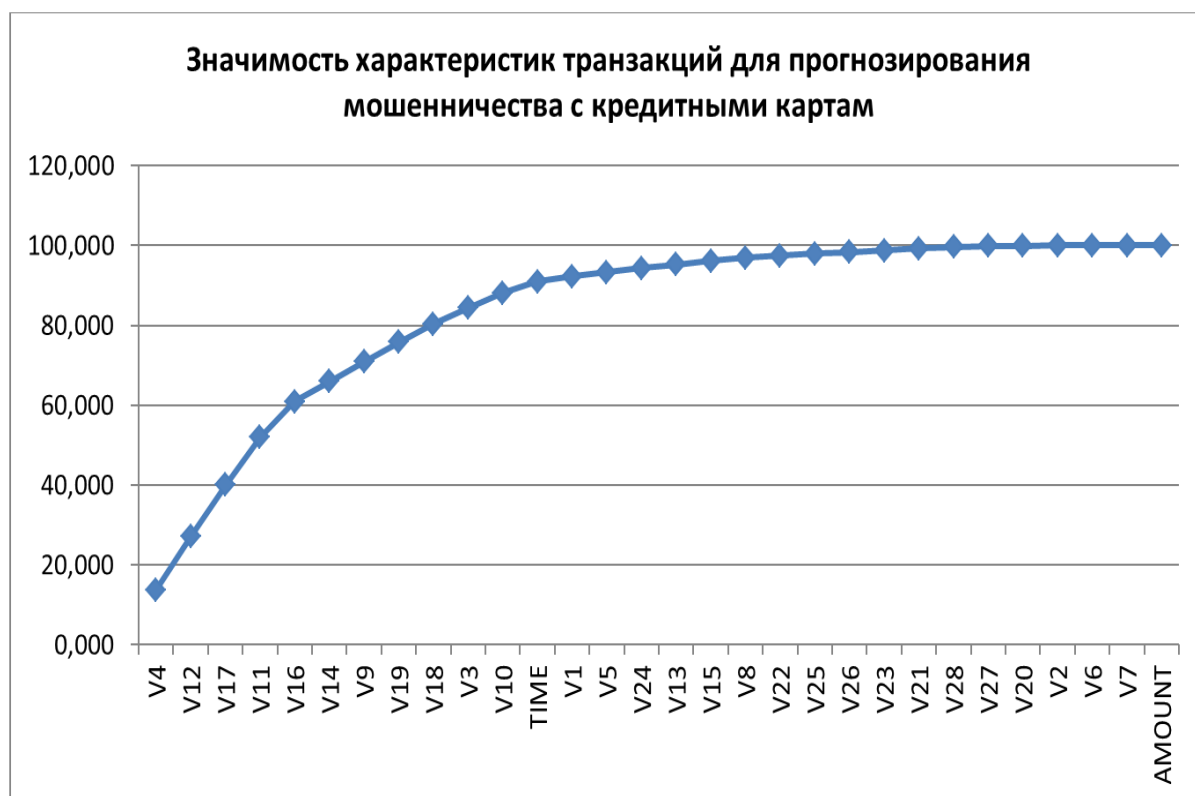


Рисунок 66

Из таблицы 18 и графика 66 видно, что наиболее значимыми являются характеристики: V4, V12, V17, V11. Эти 4 характеристики составляют всего 13% от общего числа характеристик (30), но вместе обеспечивают более 50% суммарной значимости.

Из таблицы 18 и графика 66 видно также, что 50% характеристик, т.е. характеристики: V4, V12, V17, V11, V16, V14, V9, V19, V18, V3, V10, TIME, V1, V5, V24 вместе обеспечивают более 94% суммарной значимости.

5.11. Степень детерминированности классов и классификационных шкал

Степень детерминированности (обусловленности) класса в системе «Эйдос» количественно оценивается **степенью варибельности значений факторов** (градаций описательных шкал) в колонке матрицы модели, соответствующей данному классу.

Чем выше степень детерминированности класса, тем более достоверно он прогнозируется по значениям факторов.

Получить эту информацию можно в режиме 3.7.3 системы «Эйдос» (рисунок 67).

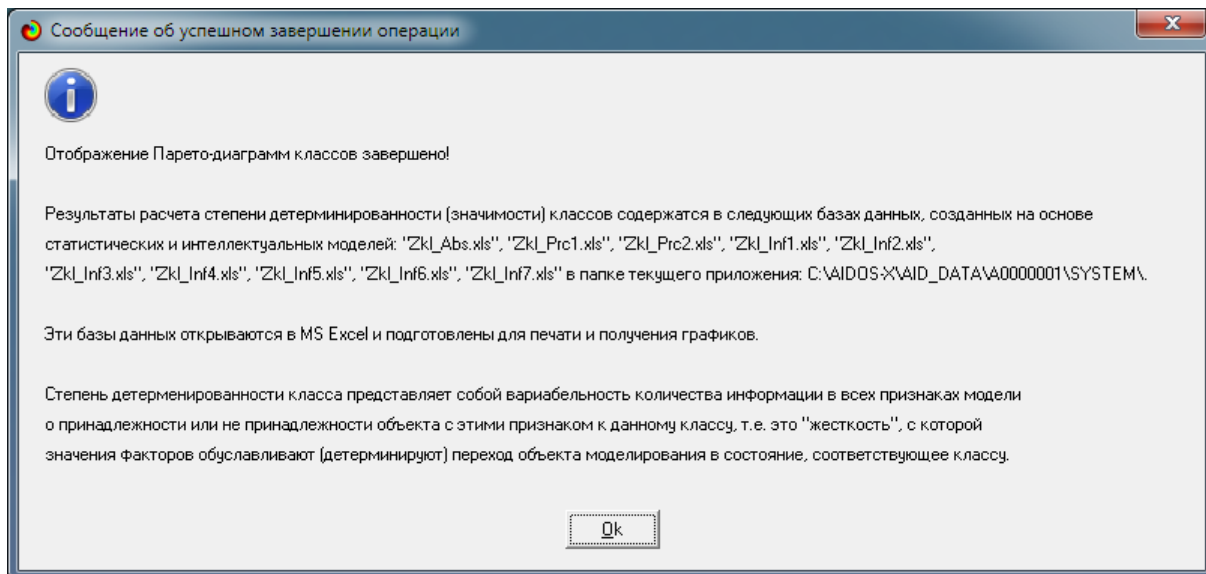


Рисунок 67

Степень детерминированности (обусловленности) всей классификационной шкалы является средним от степени детерминированности ее градаций, т.е. классов (рисунок 68, режим 3.7.2 системы «Эйдос»).

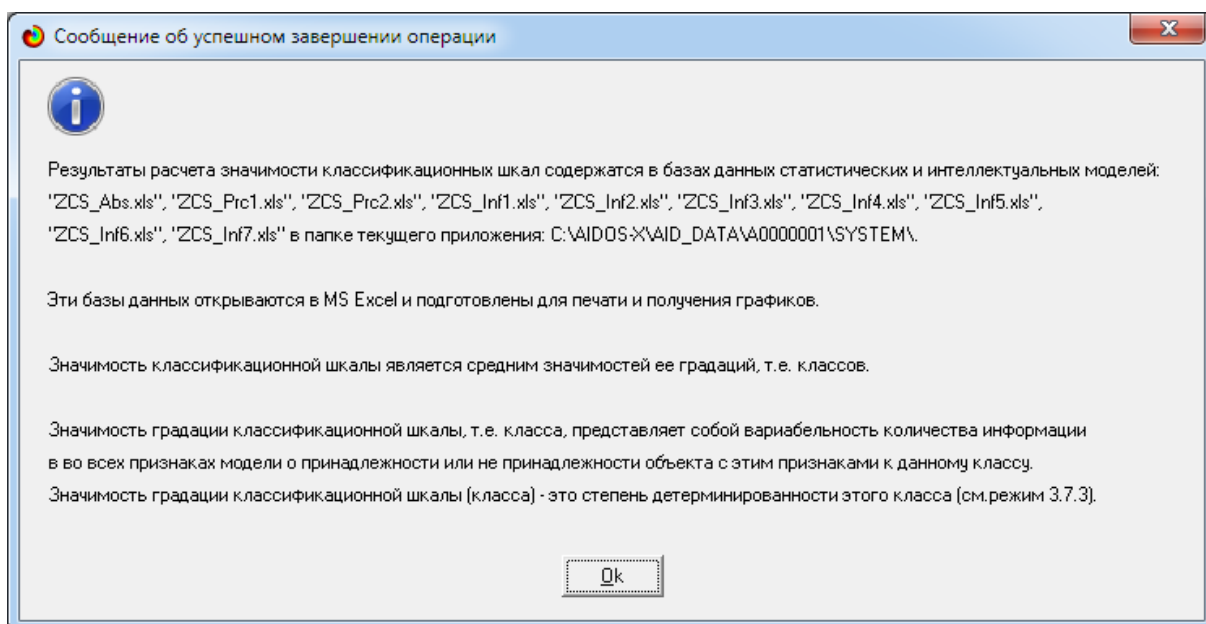


Рисунок 68

3. Выводы

Главные выводы, которые можно обоснованно сделать по результатам данного исследования, состоят в том, что:

1. Характеристики транзакций операций по кредитным картам очень тесно связаны с риском мошенничества.

2. Это связь настолько сильна, что ее знание позволяет практически безошибочно определять случаи мошенничества с применением технологий искусственного интеллекта.

Таким образом, предложенная гипотеза подтверждена.

Конечно, понятно, что в построенных и исследованных моделях отражены не все играющие роль факторы. Эти модели могут просто не отражать всю полноту различных аспектов взаимосвязей характеристик транзакций по операциям с кредитными картами с риском мошенничества. Это может быть потому, что соответствующей информации нет в исходных данных. Поэтому эти модели будут постоянно совершенствоваться с целью повышения достоверности идентификации и прогноза. Возможностей для этого очень много [10].

Со всеми моделями, созданными в данной статье, можно ознакомиться установив облачное Эйдос-приложение №297 в режиме 1.3 системы «Эйдос». Саму систему можно бесплатно скачать с сайта ее автора и разработчика проф.Е.В.Луценко по ссылке: http://lc.kubagro.ru/aidos/_Aidos-X.htm. Единственное различие с этим описанием в облачном Эйдос-приложении: файл исходных данных Inp_data.xlsx содержит 10000 записей, а не 284807. Это сделано для того, чтобы он не удался с ftp-сервера (файлы больше 10 Мб удаляются автоматически), быстрее скачивался, быстрее все считалось. Если будет интерес прогнать задачу на полных данных, то можно скачать их с портала Kaggle По указанным в начале статьи ссылкам.

Литература

1. Andrea Dal Pozzolo, Olivier Caelen, Reid A. Johnson and Gianluca Bontempi. [Calibrating Probability with Undersampling for Unbalanced Classification](#). In Symposium on Computational Intelligence and Data Mining (CIDM), IEEE, 2015
2. Dal Pozzolo, Andrea; Caelen, Olivier; Le Borgne, Yann-Aël; Waterschoot, Serge; Bontempi, Gianluca. [Learned lessons in credit card fraud detection from a practitioner perspective](#), Expert systems with applications,41,10,4915-4928,2014, Pergamon
3. Dal Pozzolo, Andrea; Boracchi, Giacomo; Caelen, Olivier; Alippi, Cesare; Bontempi, Gianluca. [Credit card fraud detection: a realistic modeling and a novel learning strategy](#). IEEE transactions on neural networks and learning systems,29,8,3784-3797,2018,IEEE
4. Dal Pozzolo, Andrea [Adaptive Machine learning for credit card fraud detection](#) ULB MLG PhD thesis (supervised by G. Bontempi)
5. Carcillo, Fabrizio; Dal Pozzolo, Andrea; Le Borgne, Yann-Aël; Caelen, Olivier; Mazzer, Yannis; Bontempi, Gianluca. [Scarff: a scalable framework for streaming credit card fraud detection with Spark](#), Information fusion,41, 182-194,2018,Elsevier
6. Carcillo, Fabrizio; Le Borgne, Yann-Aël; Caelen, Olivier; Bontempi, Gianluca. [Streaming active learning strategies for real-life credit card fraud detection: assessment and visualization](#), International Journal of Data Science and Analytics, 5,4,285-300,2018,Springer International Publishing

7. Bertrand LeBichot, Yann-Aël Le Borgne, Liyun He, Frederic Oblé, Gianluca Bontempi [Deep-Learning Domain Adaptation Techniques for Credit Cards Fraud Detection](#), INNSBDDL 2019: Recent Advances in Big Data and Deep Learning, pp 78-88, 2019
8. Fabrizio Carcillo, Yann-Aël Le Borgne, Olivier Caelen, Frederic Oblé, Gianluca Bontempi [Combining Unsupervised and Supervised Learning in Credit Card Fraud Detection](#) Information Sciences, 2019
9. Yann-Aël Le Borgne, Gianluca Bontempi [Machine Learning for Credit Card Fraud Detection - Practical Handbook](#)
10. Lutsenko E.V. Application of automated system-cognitive analysis of bank databases on credit card transactions for quantitative assessment of fraud risk // August 2021, DOI: [10.13140/RG.2.2.35999.28326](https://doi.org/10.13140/RG.2.2.35999.28326), License [CC BY 4.0](#), <https://www.researchgate.net/publication/353954436>
11. Lutsenko E.V. Forecasting in financial markets using scenario-based ASC-analysis and the Eidos system (using the example of Google shares) // July 2021, DOI: [10.13140/RG.2.2.28157.08168](https://doi.org/10.13140/RG.2.2.28157.08168), License [CC BY-SA 4.0](#), <https://www.researchgate.net/publication/353157032>
12. Lutsenko E.V. Scenario and spectral automated system-cognitive analysis // July 2021, License [CC BY-SA 4.0](#), DOI: [10.13140/RG.2.2.22981.37608](https://doi.org/10.13140/RG.2.2.22981.37608), ISBN 978-5-907474-67-3, <https://www.researchgate.net/publication/353555996>