

УДК 004.8  
08.00.13 - Математические и инструментальные  
методы экономики (экономические науки)

UDC 004.8  
08.00.13 - Mathematical and instrumental methods of  
Economics (Economics)

**Прогнозирование рисков невозврата ссуды с  
применением интеллектуального итерационного  
алгоритма учета нетипичных случаев**

**Forecasting the risks of loan non-repayment using  
an intelligent iterative algorithm for accounting for  
atypical cases**

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](mailto:prof.lutsenko@gmail.com) <http://lc.kubagro.ru>  
<https://www.researchgate.net/profile/Eugene-Lutsenko>

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](mailto:prof.lutsenko@gmail.com) <http://lc.kubagro.ru>  
<https://www.researchgate.net/profile/Eugene-Lutsenko>

В данной работе ставится, рассматривается и решается актуальная задача достоверного прогнозирования рисков невозврата ссуды. От успешности решения этой задачи самым непосредственным образом зависят прибыльность, успешность и устойчивость субсидирующих организаций. Предлагается теоретическое и практическое решение этой задачи путем применения интеллектуального итерационного алгоритма учета нетипичных случаев субсидирования, которые обычно и приводят к понижению достоверности прогнозирования. Предлагаемое решение основано на применении автоматизированного системно-когнитивного анализа (АСК-анализ) и его программного инструментария – интеллектуальной системы «Эйдос». Приводится подробный численный пример, основанный на данных портала Kaggl

In this work we set, examine and solve the actual problem of reliable forecasting of the risks of non-repayment of a loan. The profitability, success, and sustainability of subsidizing organizations directly depend on the success of this task. We have also proposed a theoretical and practical solution to this problem is by applying an intelligent iterative algorithm for accounting for atypical cases of subsidies, which usually lead to a decrease in the reliability of forecasting. The proposed solution is based on the use of automated system-cognitive analysis (ASC-analysis) and its software tools which is "Eidos", the intelligent system. We also provide a detailed numerical example based on data from the Kaggl portal

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

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

## СОДЕРЖАНИЕ

<b>ВВЕДЕНИЕ</b> .....	<b>2</b>
<b>1. ПРОБЛЕМАТИКА РАБОТЫ</b> .....	<b>3</b>
1.1. ОПИСАНИЕ ПРЕДМЕТНОЙ ОБЛАСТИ, ОБЪЕКТ, ПРЕДМЕТ, ПРОБЛЕМА, ЦЕЛЬ И ЗАДАЧИ РАБОТЫ.....	3
1.2. ОБОСНОВАНИЕ ТРЕБОВАНИЙ К МЕТОДУ РЕШЕНИЯ ПРОБЛЕМЫ .....	4
1.3. ВЫБОР МЕТОДА РЕШЕНИЯ ПРОБЛЕМЫ, СООТВЕТСТВУЮЩЕГО ОБОСНОВАННЫМ ТРЕБОВАНИЯМ .....	5
1.4. Выводы.....	7
<b>2. НАУЧНОЕ (ТЕОРЕТИЧЕСКОЕ) РЕШЕНИЕ ПРОБЛЕМЫ</b> .....	<b>8</b>
2.1. ИДЕЯ И КОНЦЕПЦИЯ НАУЧНОГО РЕШЕНИЯ ПРОБЛЕМЫ.....	8
2.2. ОПИСАНИЕ ОБЩЕЙ ТЕОРИИ МЕТОДА РЕШЕНИЯ ПРОБЛЕМЫ: СУТЬ МЕТОДА АСК-АНАЛИЗА.....	8
2.3. ОПИСАНИЕ ЛИЧНОГО ВКЛАДА АВТОРА В ТЕОРЕТИЧЕСКОЕ РЕШЕНИЕ ПРОБЛЕМЫ (НАУЧНАЯ НОВИЗНА) .....	13
2.4. Выводы.....	13
<b>3. ИНЖЕНЕРНОЕ (ПРАКТИЧЕСКОЕ) РЕШЕНИЕ ПРОБЛЕМЫ: ТЕХНОЛОГИЯ (ПРОГРАММНЫЙ ИНСТРУМЕНТАРИЙ) И МЕТОДИКА РЕШЕНИЯ ПРОБЛЕМЫ</b> .....	<b>13</b>
3.1. ТЕХНИКО-ЭКОНОМИЧЕСКОЕ ОБОСНОВАНИЕ ЦЕЛЕСООБРАЗНОСТИ РАБОТЫ (ТЭО) .....	13
3.2. ТЕХНИЧЕСКОЕ ЗАДАНИЕ (ТЗ): ФУНКЦИОНАЛЬНОЕ ОПИСАНИЕ ТОГО, ЧТО НЕОБХОДИМО СДЕЛАТЬ (ТЗ) .....	14

<b>3.3. ТЕХНИЧЕСКИЙ ПРОЕКТ (ТП) и РАБОЧИЙ ПРОЕКТ (РП): СТРУКТУРЫ И ОТНОШЕНИЯ ДАННЫХ, АЛГОРИТМЫ ИХ ОБРАБОТКИ, РЕАЛИЗАЦИЯ</b> .....	14
3.3.1. <i>Суть математической модели АСК-анализа и частные критерии</i> .....	14
3.3.2. <i>Интегральные критерии и принятие управляющих решений</i> .....	18
3.3.3. <i>Интеллектуальный итерационный алгоритм учета нетипичных случаев</i> .....	20
<b>3.4. Выводы</b> .....	23
<b>4. ПРИМЕНЕНИЕ РАЗРАБОТАННОЙ ТЕХНОЛОГИИ НА ПРАКТИКЕ, ЕЕ ВНЕДРЕНИЕ И ОЦЕНКА ЭФФЕКТИВНОСТИ РЕШЕНИЯ ПРОБЛЕМЫ</b> .....	<b>23</b>
<b>4.1. МЕТОДИКА, ПЛАН И ОБЕСПЕЧЕНИЕ ВНЕДРЕНИЯ</b> .....	23
<b>4.2. МЕТОДИКА ОЦЕНКИ ЭФФЕКТИВНОСТИ ВНЕДРЕНИЯ</b> .....	24
<b>4.3. ОПИСАНИЕ ВНЕДРЕНИЯ И ЕГО РЕЗУЛЬТАТОВ (ЧИСЛЕННЫЙ ПРИМЕР)</b> .....	25
4.3.1. <i>Задача 1: когнитивная структуризация предметной области</i> .....	25
4.3.2. <i>Задача 2: подготовка исходных данных и формализация предметной области</i> .....	27
4.3.3. <i>Задача 3: применение итерационного алгоритма разделения обобщенных образов классов на типичную и нетипичную части</i> .....	33
4.3.3.1. <i>Исходная модель</i> .....	33
4.3.3.1.1. <i>Синтез и верификация статистических и системно-когнитивных моделей</i> .....	33
4.3.3.1.2. <i>Оценка достоверности моделей</i> .....	36
4.3.3.1.3. <i>Задание текущей модели и распознавание в ней</i> .....	38
4.3.3.2. <i>Первая итерация. Синтез и верификация статистических и системно-когнитивных моделей. Оценка достоверности моделей. Задание текущей модели и распознавание в ней</i> .....	39
4.3.3.3. <i>Вторая итерация. Синтез и верификация статистических и системно-когнитивных моделей. Оценка достоверности моделей. Задание текущей модели и решение в ней всех поставленных задач</i> .....	47
4.3.4. <i>Задача 4: решение различных задач в наиболее достоверной модели</i> .....	48
4.3.4.1. <i>Подзадача 4.1. Прогнозирование (диагностика, классификация, распознавание, идентификация)</i> .....	48
4.3.4.2. <i>Подзадача 4.2. Поддержка принятия решений в простейшем варианте (SWOT-анализ)</i> .....	50
4.3.4.3. <i>Подзадача 4.2. Развитый алгоритм принятия решений</i> .....	54
4.3.4.4. <i>Подзадача 4.3. Исследование моделируемой предметной области путем исследования ее модели</i> .....	56
4.3.4.4.1. <i>Когнитивные диаграммы классов</i> .....	56
4.3.4.4.2. <i>Агломеративная когнитивная кластеризация классов</i> .....	57
4.3.4.4.3. <i>Когнитивные диаграммы значений факторов</i> .....	58
4.3.4.4.4. <i>Агломеративная когнитивная кластеризация значений факторов</i> .....	60
4.3.4.4.5. <i>Нелокальные нейроны и нелокальные нейронные сети</i> .....	61
4.3.4.4.6. <i>3d-интегральные когнитивные карты</i> .....	63
4.3.4.4.7. <i>Когнитивные функции</i> .....	64
4.3.4.4.8. <i>Сила и направление влияния значений факторов на принадлежность к классам</i> .....	67
4.3.4.4.9. <i>Степень детерминированности классов значениями обуславливающих их факторов</i> ..	70
4.3.5. <i>Выводы</i> .....	72
<b>4.4. Выводы</b> .....	72
<b>ЗАКЛЮЧЕНИЕ</b> .....	<b>73</b>
<b>ЛИТЕРАТУРА</b> .....	<b>74</b>
<b>ПРИЛОЖЕНИЕ 1. ПРОГРАММНАЯ РЕАЛИЗАЦИЯ В СИСТЕМЕ «ЭЙДОС» ИНТЕЛЛЕКТУАЛЬНОГО ИТЕРАЦИОННОГО АЛГОРИТМА УЧЕТА НЕТИПИЧНЫХ СЛУЧАЕВ</b> .....	<b>78</b>

## Введение

В данной работе ставится, подробно рассматривается и решается актуальная задача достоверного прогнозирования рисков невозврата ссуды. От успешности решения этой задачи самым непосредственным образом зависят прибыльность, успешность и устойчивость субсидирующих организаций. Предлагается теоретическое и практическое решение этой задачи путем применения интеллектуального итерационного алгоритма учета нетипичных случаев субсидирования, которые обычно и приводят к понижению достоверности прогнозирования. Предлагаемое

решение основано на применении автоматизированного системно-когнитивного анализа (АСК-анализ) и его программного инструментария – интеллектуальной системы «Эйдос». Приводится подробный численный пример, основанный на данных портала Каггл. Структура изложения в данной работе на сколько это возможно приближена к рекомендациям, приведенным в [1], а также [2, 3].

## **1. Проблематика работы**

### **1.1. Описание предметной области, объект, предмет, проблема, цель и задачи работы**

Задача прогнозирования рисков невозврата финансовых средств (ссуд, кредитов, инвестиций)<sup>1</sup> физическими и юридическими лицами является весьма актуальной для всех, систематически занимающихся этим видом деятельности. Если достоверно прогнозировать риск невозврата ссуды и предоставлять ее только тогда, когда риск ее невозврата мал и не предоставлять ее, когда он велик, то прибыль от субсидирования будет высока<sup>2</sup>. Понятно, что чем ниже достоверность прогнозирования риска невозврата ссуды, тем ниже прибыльность от субсидирования. Более того, при высоком риске невозврата субсидирование может привести к убыточности от этого вида деятельности или даже к банкротству («неоправданное кредитование»)<sup>3</sup>.

Поэтому прибыль от субсидирования самым непосредственным образом зависит от достоверности прогнозирования риска невозврата ссуды.

Из этого следует объект и предмет исследования, а также проблема, решаемая в работе, ее цель и задачи.

Объект исследования – прогнозирование риска невозврата ссуды.

Предмет исследования – прогнозирование рисков невозврата ссуды с применением интеллектуального итерационного алгоритма учета нетипичных случаев.

Проблема – возникающая при прогнозировании риска невозврата ссуды – ложно-отрицательные решения, возникающие, когда модель не относит ситуацию к классу (обобщенной категории), к которой она в действительности относится. Этот вид ошибок возникает в нетипичных ситуациях, резко снижает достоверность прогнозирования и приводит к ошибочным рекомендациям о выдаче или невыдаче ссуды, противоположным адекватным. *Фактически* стандартные алгоритмы прогнозирования основаны на обобщении типовых ситуаций и, поэтому, нетипичные ситуации ими обрабатываются ошибочно. *Желательно* было

---

<sup>1</sup> Далее в работе будем говорить о ссудах

<sup>2</sup> Это очень напоминает вечный двигатель, основанный на «демоне Максвелла», непосредственно преобразующий информацию в энергию.

<sup>3</sup> Случаи целенаправленного банкротства в данной работе не рассматриваются.

бы, чтобы алгоритм прогнозирования корректно обрабатывал не только типовые, но и нетипичные случаи.

Таким образом, налицо противоречие между фактическим и желаемым и это противоречие *неприемлемо*.

Целью данной работы является решение поставленной проблемы путем разработки алгоритма, обеспечивающего формирование обобщенных образов классов не только на основе типичных относящихся к ним наблюдений, но и на основе исключений, т.е. нетипичных случаев.

Путем декомпозиции цели получим последовательность **задач**, решение которых является этапами достижения цели:

Задача 1: когнитивная структуризация предметной области.

Задача 2: подготовка исходных данных и формализация предметной области (разработка классификационных и описательных шкал и градаций и кодирование исходных с их помощью, т.е. получение обучающей выборки).

Задача 3: синтез и верификация статистических и системно-когнитивных моделей и выбор наиболее достоверной модели. Улучшение наиболее достоверной модели путем применения итерационного алгоритма разделения обобщенных образов классов на типичную и нетипичную части.

Задача 4: решение различных задач в наиболее достоверной модели:

– подзадача 4.1. Прогнозирование (диагностика, классификация, распознавание, идентификация);

– подзадача 4.2. Поддержка принятия решений;

– подзадача 4.3. Исследование моделируемой предметной области путем исследования ее модели: когнитивные диаграммы классов и значений факторов, агломеративная когнитивная кластеризация классов и значений факторов, нелокальные нейроны и нейронные сети, 3d-интегральные когнитивные карты, когнитивные функции), исследование силы и направления влияния факторов и степени детерминированности классов, обуславливающими их значениями факторов.

## **1.2. Обоснование требований к методу решения проблемы**

Обычно при решении реальных, а не чисто учебных проблем, а в данной работе рассматривается именно такой случай, метод должен обеспечивать 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). Очень важно также, чтобы данный метод был реализован в реально доступной пользователям программной системе.

### 1.3. Выбор метода решения проблемы, соответствующего обоснованным требованиям

В работах [4-9], посвященных вопросам прогнозирования рисков невозврата заемных средств (скоринг) само понятие «нетипичные случаи» вообще не встречается. В работе [10] оно встречается, но не приводится ни способа выявления таких нетипичных случаев с применением технологий искусственного интеллекта, ни *способа их обработки и учета в интеллектуальных моделях* (о которых в работе [10] вообще не упоминается).

В работах автора [11-18] приводятся исследования и разработки учета нетипичных случаев в интеллектуальных скоринговых моделях с применением Автоматизированного системно-когнитивного анализа (АСК-анализ) и его программного инструментария интеллектуальной системы «Эйдос» [19, 20, 31].

Эти разработки соответствуют всем обоснованным в разделе 1.2 требованиям, за исключением одного, но критически важного требования: описанная в них DOS-версия системы «Эйдос» морально устарела и давно отсутствует в открытом доступе.

Поэтому с 2012 года автором разрабатывается и непрерывно совершенствуется органичная для MS Windows новая версия этой системы: «Эйдос-X++». Эта система:

- 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](http://lc.kubagro.ru/_AidosALL.txt)): открытая лицензия: CC 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 293, соответственно) ([http://lc.kubagro.ru/aidos/Presentation\\_Aidos-online.pdf](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 решение 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](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.

По перечисленным выше причинам метод «АСК-анализ» и его инструментарий – интеллектуальная система «Эйдос» выбраны в качестве **основы** для решения поставленной проблемы. В рамках решения поставленной в данной работе проблемы в системе «Эйдос» должен быть разработан новый режим, реализующий предложенные в данной работе идею, концепцию и интеллектуальный итерационный алгоритм учета в модели прогнозирования нетипичных случаев возврата и невозврата ссуды. Описание данных идеи, концепции, алгоритма и режима в открытой печати осуществляется впервые. В этом состоит актуальность данной работы.

#### **1.4. Выводы**

По результатам 1-й главы сформулированы объект, предмет, проблема, цель и задачи работы. В качестве метода решения поставленной проблемы выбран АСК-анализ, который полностью соответствует обоснованным требованиям.

Вместе с тем этот метод должен быть доработан в следующих аспектах:

1. Должна быть предложена идея и разработана концепция учета в модели прогнозирования нетипичных случаев возврата и невозврата ссуды.

2. Данная концепция должна быть конкретизирована и детализирована до уровня математической модели и соответствующей методики численных расчетов (алгоритма и структуры данных) и реализована в виде режима в интеллектуальной системе «Эйдос».

В этой доработке наиболее подходящего метода решения проблемы и его апробации на основе реальных данных и состоит содержание последующих разделов данной работы.

## **2. Научное (теоретическое) решение проблемы**

### **2.1. Идея и концепция научного решения проблемы**

Идея научного решения проблемы состоит в следующем. Обобщенные образы классов создаются путем обобщения всех относящихся к ним наблюдений обучающей выборки. При верификации моделей выясняется, что некоторые наблюдения модель правильно относит к тем классам, к которым они действительно принадлежат, а некоторые нет. Естественно считать, что верно идентифицируются те наблюдения, которые являются типичными для данной предметной области. Нетипичные же наблюдения, являющиеся исключениями из правила (т.е. выявленных и отраженных в модели закономерностей), модель не относит к тем классам, к которым они в действительности относятся, т.е. дает ложно-отрицательные решения. Предлагается создать собственные классы с теми же наименованиями для нетипичных наблюдений (исключений).

Концепция научного решения проблемы. Осуществить формализацию предметной области: подготовка исходных данных и формализация предметной области (разработка классификационных и описательных шкал и градаций и кодирование исходных с их помощью, т.е. получение обучающей выборки). Выполнить синтез моделей: Abs, Prc1, Prc2, Inf1, Inf2, Inf3, Inf4, Inf5, Inf6, Inf7. Провести верификацию моделей путем распознавания в них наблюдений из обучающей выборки. Выбрать модель, наиболее достоверную по F-критерию Ван Ризбергена. Если количество ложно-отрицательных решений в результатах распознавания в наиболее достоверной модели не уменьшилось, то переход на выход. Иначе для наблюдений обучающей выборки, идентификация которых привела к ложно-отрицательным решениям (т.е. они не были отнесены к классам, к которым в действительности относятся) создать для них в исходных данных свои классы с теми же наименованиями, но с добавлением к названию символов, означающих, что этот класс создан для исключений. Перейти на начало. Выход.

### **2.2. Описание общей теории метода решения проблемы: суть метода АСК-анализа**

Суть метода АСК-анализа состоит в последовательном повышении степени формализации модели и преобразовании данных в информацию, а ее в знания и решении на основе этих знаний задач идентификации (расознавания, классификации и прогнозирования), поддержки принятия решений и исследования моделируемой предметной области [19, 20] (рисунки 1 и 2):

О соотношении содержания понятий: «Данные», «Информация» и «Знания»

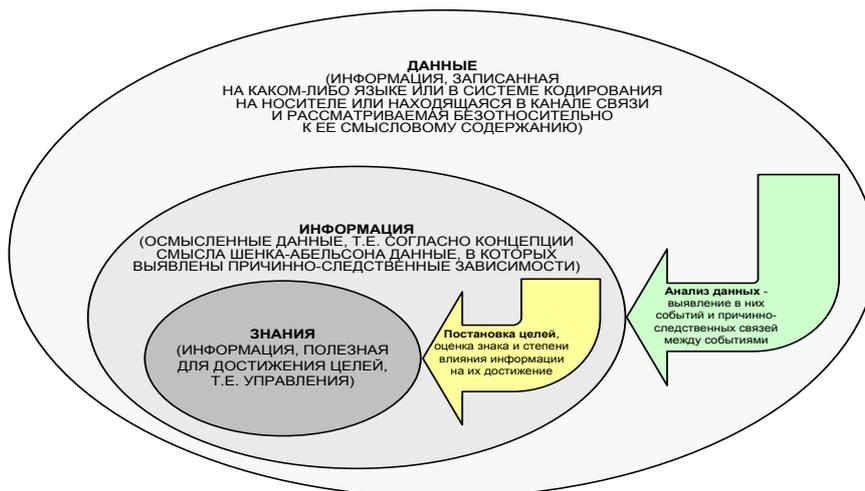


Рисунок 1. О соотношении содержания понятий:  
«данные», «информация» и «знания» в АСК-анализе

Понятие данных, по-видимому, является одним из предельно общих понятий, выработанных человечеством, может быть даже это наиболее общее понятие. По мнению автора, понятие данных является даже более общим, чем понятия материи и сознания, бытия и небытия, пространства и времени. О чем бы мы не говорили, о чем бы не думали, что бы не представляли, все это мы делаем лишь оперируя определенными данными. Таким образом, человек находится как бы в мыльном пузыре или коконе из данных, за пределы которого он в принципе не может выйти. И цветные движущиеся узоры на поверхности этого мыльного пузыря человек и принимает за реальный мир. А чтобы определить, что такое данные именно это и нужно сделать, т.е. необходимо выйти за пределы данных к чему-то еще более общему. Но как это сделать не совсем понятно, даже в принципе. Это создает большие сложности для традиционного определения понятия данных путем подведения его под более общее понятие и выделение специфического признака или признаков<sup>4</sup>.

Поэтому предлагается два варианта.

1-й вариант: указать только специфический признак данных, а более общее понятие не указывать. Этот вариант приводит к такому определению: **данные** – это все, связанное с теми или иными различиями в наличии или степени выраженности тех или иных свойств.

2-й вариант: определить понятие данных через частное понятие, входящее в него, и сделать это не смотря на то, что это не принято<sup>5</sup>.

<sup>4</sup> Например, как определяют млекопитающее: это животное (более общее понятие), выкармливающее своих детенышей молоком (специфический признак, выделяющий из более общего понятия определяемое подмножество).

<sup>5</sup> Например, животные (более общее понятие), это и млекопитающие, и не млекопитающие, т.е. как выкармливающее своих детенышей молоком (специфический признак, выделяющий из более общего понятия известное частное понятие), так и не молоком и вообще не выкармливающие.

**Данные** – это информация, записанная на каком-либо носителе или находящаяся в каналах связи и представленная на каком-то языке или в системе кодирования и рассматриваемая как с семантическим аспектом, так и безотносительно к ее смысловому содержанию.

Исходные данные об объекте управления обычно представлены в форме баз данных, чаще всего временных рядов, т.е. данных, привязанных ко времени. В соответствии с методологией и технологией автоматизированного системно-когнитивного анализа (АСК-анализ), для прогнозирования, управления и принятия решений использовать непосредственно исходные данные не представляется возможным. Точнее сделать это можно, но результат прогнозирования и управления при таком подходе оказывается мало чем отличающимся от случайного. Для реального же решения задач прогнозирования и управления необходимо предварительно преобразовать данные в информацию, а ее в знания о том, какие воздействия на объект моделирования к каким его изменениям обычно, как показывает опыт, приводят.

**Информация есть осмысленные данные.**

Смысл данных, в соответствии с концепцией смысла Шенка-Абельсона, состоит в том, что известны причинно-следственные зависимости между событиями, которые описываются этими данными. Таким образом, данные преобразуются в информацию в результате операции, которая называется «Анализ данных», состоящей из двух этапов:

1. Выявление событий в данных (разработка классификационных и описательных шкал и градаций и преобразование с их использованием исходных данных в обучающую выборку, т.е. в базу событий – эвентологическую базу).

2. Выявление причинно-следственных зависимостей между событиями.

В случае систем прогнозирования и управления этими событиями в данных являются совпадения определенных значений входных факторов и выходных параметров объекта управления, т.е. по сути, случаи перехода объекта управления в определенные будущие состояния под действием определенных сочетаний значений управляющих факторов. Качественные значения входных факторов и выходных параметров естественно формализовать в форме лингвистических переменных. Если же входные факторы и выходные параметры являются числовыми, то их значения измеряются с некоторой погрешностью и фактически представляют собой интервальные числовые значения, которые также могут быть представлены или формализованы в форме лингвистических переменных (типа: «малые», «средние», «большие» значения экономических показателей) в форме числовых диапазонов.

Какие же *математические меры* могут быть использованы для количественного измерения силы и направления причинно-следственных зависимостей?

Наиболее очевидным ответом на этот вопрос, который обычно первым всем приходит на ум, является: «Корреляция». Однако, в статистике это хорошо известно, что это совершенно не так. Для преобразования исходных данных в информацию необходимо не только выявить события в этих данных, но и найти причинно-следственные связи между этими событиями. В АСК-анализе предлагается 7 количественных мер причинно-следственных связей, основными из которых являются заслуженно знаменитая мера Пирсона хи-квадрат и математически связанная с ней семантическая мера целесообразности информации по Александра Харкевича.

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

Значит для преобразования информации в знания необходимо:

1. Поставить цель (классифицировать будущие состояния моделируемого объекта на целевые и на нежелательные).
2. Оценить полезность информации для достижения этой цели (знак и силу влияния).

Второй пункт выполнен, по сути, при преобразовании данных в информацию. Поэтому остается выполнить только первый пункт, т.е. классифицировать будущие состояния объекта управления как желательные (целевые) и нежелательные.

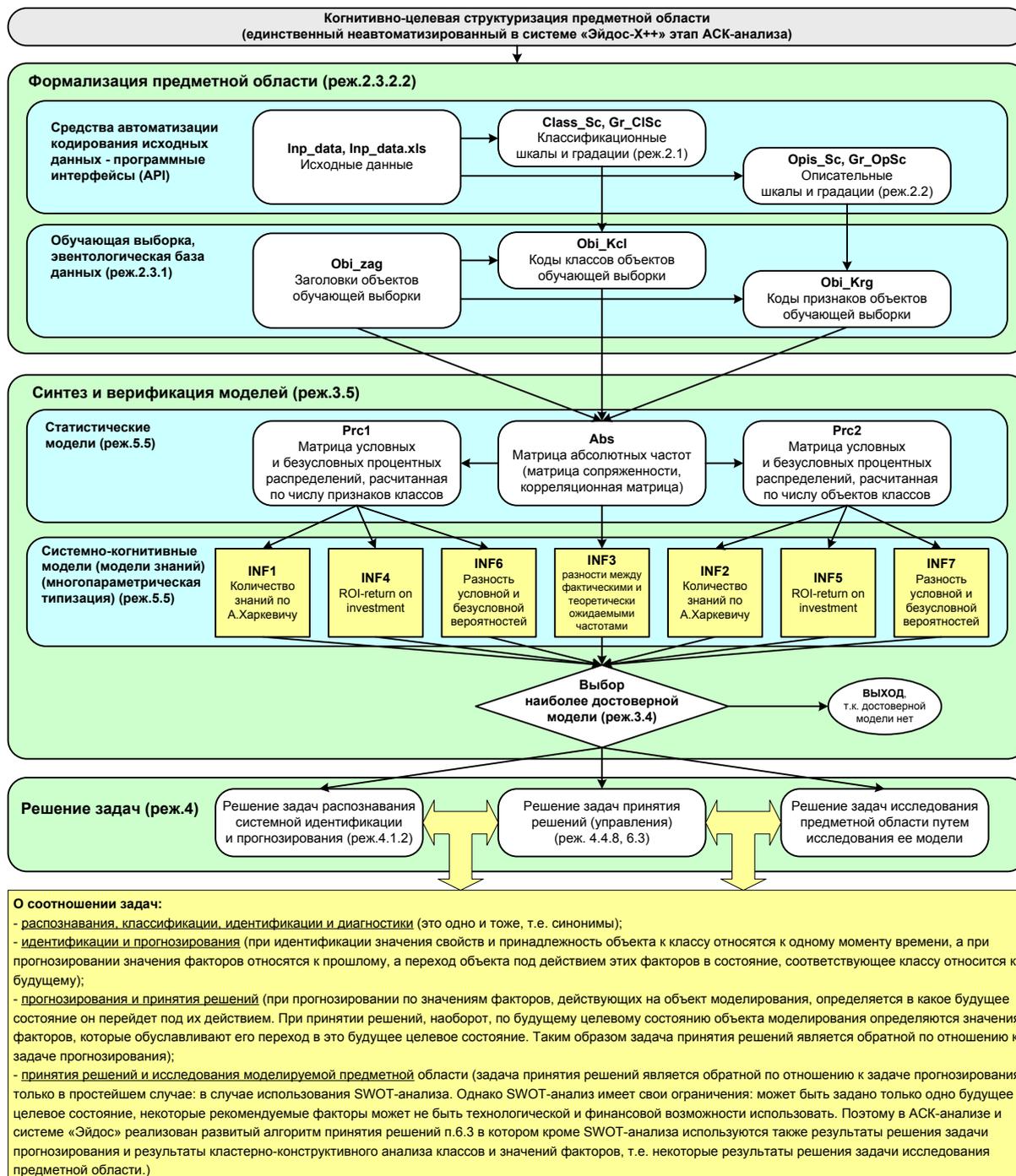
Знания могут быть представлены в различных формах, характеризующихся различной степенью формализации:

- вообще неформализованные знания, т.е. знания в своей собственной форме, ноу-хау (мышление без вербализации есть медитация);
- знания, формализованные в естественном вербальном языке;
- знания, формализованные в виде различных методик, схем, алгоритмов, планов, таблиц и отношений между ними (даталогические и инфологические модели баз данных);
- знания в форме технологий, организационных, производственных, социально-экономических и политических структур;
- знания, формализованные в виде математических моделей и методов представления знаний в автоматизированных интеллектуальных системах (логическая, фреймовая, сетевая, продукционная, нейросетевая, нечеткая и другие);
- знания, формализованные в форме различных математических, физических, химических и других научных моделей.

Таким образом, для решения сформулированной проблемы необходимо осознанно и целенаправленно **последовательно повышать степень формализации** исходных данных до уровня, который позволяет ввести исходные данные в интеллектуальную систему, а затем:

- преобразовать исходные данные в информацию;
- преобразовать информацию в знания;
- использовать знания для решения задач прогнозирования, управления, принятия решений и исследования предметной области (рисунок 2).

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



**Рисунок 2. Последовательность преобразования данных  
в информацию, а ее в знания и решения задач  
в АСК-анализе и системе «Эйдос»**

### **2.3. Описание личного вклада автора в теоретическое решение проблемы (научная новизна)**

- *предлагается* идея и концепция учета нетипичных случаев;
- которая *в отличие от традиционного подхода* (Луценко, 1995, 1996);
  - *путем* разделения классов на типичную и нетипичную части (нетипичными считаются наблюдения, порождающие ложно-отрицательные решения);
  - *обеспечивает 61,5%:* повышение достоверности прогнозирования рисков невозврата ссуды по критерию L1 проф.Е.В.Луценко (мера L1 является нечетким мультиклассовым обобщением известной F-меры Ван Ризбергена [24]).

В этом и состоит научная новизна исследования.

### **2.4. Выводы**

В данном разделе автором *предложен* интеллектуальный итерационный алгоритм учета нетипичных случаев, который *в отличие от традиционного подхода* (Луценко, 1995, 1996) *путем* разделения классов на типичную и нетипичную части (нетипичными считаются наблюдения, порождающие ложно-отрицательные решения) *обеспечивает 61,5%:* повышение достоверности прогнозирования рисков невозврата ссуды.

Этим самым поставленная проблема решена теоретически и задача главы выполнена.

Однако для применения предложенного теоретического решения поставленной проблемы на практике необходимо разработать и создать инфраструктуру практического применения этого решения. Решению этой задачи и посвящен следующий раздел.

## **3. Инженерное (практическое) решение проблемы: технология (программный инструментарий) и методика решения проблемы**

### **3.1. Техничко-экономическое обоснование целесообразности работы (ТЭО)**

Ответим на вопрос: «*Зачем это нужно?*». Ответ на этот вопрос, по сути, является основным содержанием «Техничко-экономического обоснования» (ТЭО).

Задача прогнозирования рисков невозврата финансовых средств (ссуд, кредитов, инвестиций) физическими и юридическими лицами является весьма актуальной для всех, систематически занимающихся этим видом деятельности. Если достоверно прогнозировать риск невозврата ссуды и предоставлять ее только тогда, когда риск ее невозврата мал и не предоставлять ее, когда он велик, то прибыль от субсидирования будет высока. Понятно, что чем ниже достоверность прогнозирования риска невозврата ссуды, тем ниже прибыльность от субсидирования. Более того,

при высоком риске невозврата субсидирование может привести к убыточности от этого вида деятельности или даже к банкротству («неоправданное кредитование»). Поэтому прибыль от субсидирования самым непосредственным образом зависит от достоверности прогнозирования риска невозврата ссуды.

Из этого следует целесообразность проведения исследования.

### **3.2. Техническое задание (ТЗ): функциональное описание того, что необходимо сделать (ТЗ)**

Ответим на вопрос: «*Что нужно сделать*», чтобы решить поставленную проблему, т.е. какие *функции* необходимо реализовать в программной системе, чтобы она обеспечивала решение поставленной проблемы. Ответ на этот вопрос является основным содержанием «Технического задания» (ТЗ).

В системе «Эйдос» необходимо реализовать режим, обеспечивающий реализацию сформулированных выше идеи и концепции решения проблемы.

### **3.3. Технический проект (ТП) и рабочий проект (РП): структуры и отношения данных, алгоритмы их обработки, реализация**

Ответим на вопрос: «*Как это сделать?*», т.е. рассмотрим математические модели и методику численных расчетов (алгоритмы и структуры данных).

#### **3.3.1. Суть математической модели АСК-анализа и частные критерии**

**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 [21, 22].

Суть математической модели АСК-анализа состоит в следующем.

Непосредственно на основе эмпирических данных рассчитывается матрица абсолютных частот (таблица 1).

На ее основе рассчитываются матрицы условных и безусловных процентных распределений (таблица 2).

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 1 – Матрица абсолютных частот

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

Таблица 2 – Матрица условных и безусловных процентных распределений

		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				$P_{\Sigma j}$			

Затем на основе таблицы 2 с использованием частных критериев, приведенных таблице 3 рассчитываются матрицы системно-когнитивных моделей (таблица 4).

Таблица 3 – Различные аналитические формы частных критериев знаний

Наименование модели знаний и частный критерий	Выражение для частного критерия	
	через относительные частоты	через абсолютные частоты
<b>ABS</b> , матрица абсолютных частот	---	$N_{ij}$
<b>PRC1</b> , матрица условных и безусловных процентных распределений, в качестве $N_{\Sigma j}$ используется суммарное количество признаков по классу	---	$P_{ij} = \frac{N_{ij}}{N_{\Sigma j}}$
<b>PRC2</b> , матрица условных и безусловных процентных распределений, в качестве $N_{\Sigma j}$ используется суммарное количество объектов обучающей выборки по классу	---	$P_{ij} = \frac{N_{ij}}{N_{\Sigma j}}$
<b>INF1</b> , 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	$I_{ij} = \Psi \times \text{Log}_2 \frac{P_{ij}}{P_i}$	$I_{ij} = \Psi \times \text{Log}_2 \frac{N_{ij}N}{N_i N_j}$
<b>INF2</b> , 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.	$I_{ij} = \Psi \times \text{Log}_2 \frac{P_{ij}}{P_i}$	$I_{ij} = \Psi \times \text{Log}_2 \frac{N_{ij}N}{N_i N_j}$
<b>INF3</b> , частный критерий: <b>Хи-квадрат</b> : разности между фактическими и теоретически ожидаемыми абсолютными частотами	---	$I_{ij} = N_{ij} - \frac{N_i N_j}{N}$
<b>INF4</b> , частный критерий: ROI - Return On Investment, 1-й вариант расчета вероятностей: $N_j$ – суммарное количество признаков по $j$ -му классу	$I_{ij} = \frac{P_{ij}}{P_i} - 1 = \frac{P_{ij} - P_i}{P_i}$	$I_{ij} = \frac{N_{ij}N}{N_i N_j} - 1$
<b>INF5</b> , частный критерий: ROI - Return On Investment, 2-й вариант расчета вероятностей: $N_j$ – суммарное количество объектов по $j$ -му классу	$I_{ij} = \frac{P_{ij}}{P_i} - 1 = \frac{P_{ij} - P_i}{P_i}$	$I_{ij} = \frac{N_{ij}N}{N_i N_j} - 1$
<b>INF6</b> , частный критерий: разность условной и безусловной вероятностей, 1-й вариант расчета вероятностей: $N_j$ – суммарное количество признаков по $j$ -му классу	$I_{ij} = P_{ij} - P_i$	$I_{ij} = \frac{N_{ij}}{N_j} - \frac{N_i}{N}$
<b>INF7</b> , частный критерий: разность условной и безусловной вероятностей, 2-й вариант расчета вероятностей: $N_j$ – суммарное количество объектов по $j$ -му классу	$I_{ij} = P_{ij} - P_i$	$I_{ij} = \frac{N_{ij}}{N_j} - \frac{N_i}{N}$

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;

$\Psi$  – 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 .

Таблица 4 – Матрица системно-когнитивной модели

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

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 [19-22].

На основе системно-когнитивных моделей, представленных в таблице 4 (отличаются частыми критериями), 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 [19-22].

Для решения этих задач в АСК-анализе и системе «Эйдос» в настоящее время используется два интегральных критерия.

### 3.3.2. Интегральные критерии и принятие управляющих решений

Задача принятия управляющих решений представляет собой обратную задачу прогнозирования. Если при прогнозировании на основе значений факторов, воздействующих на объект управления, определяется в какое состояние он под их воздействием перейдет, но при принятии решений наоборот, по желательному (целевому) состоянию объекта управления определяется система значений факторов, обуславливающих переход объекта в это целевое состояние.

Не все модели обеспечивают решение обратной задачи прогнозирования. Для этого они должны обеспечивать многопараметрическую типизацию, т.е. создавать обобщенные образы в будущих состояний объекта управления. Как влияет на поведение объекта управления одно значение фактора отражено в системно-когнитивных моделях. Как влияние система факторов определяется с помощью интегральных критериев. В настоящее время в системе «Эйдос» используется два аддитивных интегральных критерия:

- сумма знаний;
- резонанс знаний.

1-й интегральный критерий «Сумма знаний» представляет собой суммарное количество знаний, содержащееся в системе значений 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:

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

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:  $M$  – 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).

2-й интегральный критерий «Семантический резонанс знаний» 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;

$\sigma_j$  – среднее квадратичное отклонение частных критериев знаний вектора класса;

$\sigma_l$  – среднее квадратичное 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).

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.

Система «Эйдос» обеспечивает построение интеллектуальных информационно-измерительных систем в различных предметных областях. В системе «Эйдос» реализовано большое количество программных интерфейсов, обеспечивающий автоматизированный ввод в систему данных различных типов: текстовых, табличных и графических.

Путем многопараметрической типизации в системе создается системно-когнитивная модель, с применением которой, если модель окажется достаточно достоверной, могут решаться задачи системной идентификации, прогнозирования, классификации, поддержки принятия решений и исследования моделируемого объекта путем исследования его системно-когнитивной модели.

### **3.3.3. Интеллектуальный итерационный алгоритм учета нетипичных случаев**

Предлагается следующий алгоритм решения проблемы (рисунок 3):

Шаг 1. Осуществить формализацию предметной области: подготовка исходных данных и формализация предметной области (разработка классификационных и описательных шкал и градаций и кодирование исходных с их помощью, т.е. получение обучающей выборки).

Шаг 2. Выполнить синтез моделей: Abs, Prc1, Prc2, Inf1, Inf2, Inf3, Inf4, Inf5, Inf6, Inf7.

Шаг 3. Провести верификацию моделей путем распознавания в них наблюдений из обучающей выборки.

Шаг 4. Выбрать в качестве текущей модель, наиболее достоверную по F-критерию Ван Ризбергена и провести в ней распознавание.

Шаг 5. Если количество ложно-отрицательных решений в результатах распознавания в наиболее достоверной модели не уменьшилось, то переход на шаг 7 (выход).

Шаг 6. Для наблюдений обучающей выборки, идентификация которых привела к ложно-отрицательным решениям (т.е. они не были

отнесены к классам, к которым в действительности относятся) создать для них в исходных данных свои классы с теми же наименованиями, но с добавлением к названию символов, означающих, что этот класс создан для исключений. Перейти на шаг 1.

Шаг 7. Выход.

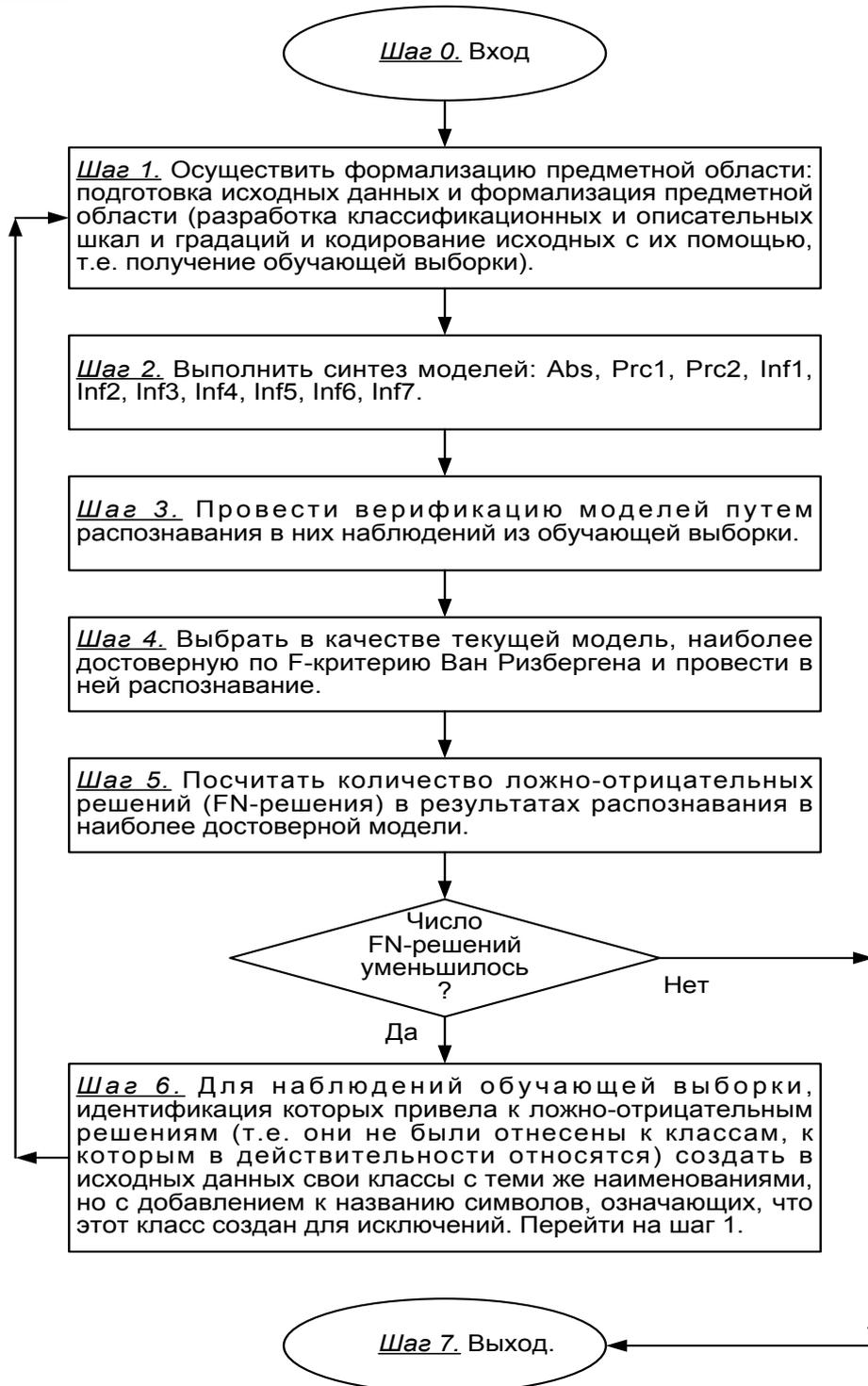


Рисунок 3. Интеллектуальный итерационный алгоритм учета нетипичных случаев

Рабочий проект (РП) – это сама программная реализация данного алгоритма и соответствующих структур данных.

В разработке технического и рабочего проекта АСК-анализа и интеллектуальной системы «Эйдос» были следующие этапы:

1-й этап, «подготовительный»: 1979-1992 годы. Математическая модель системы "Эйдос" разработана в 1979 и впервые прошла экспериментальную апробацию в 1981 году (первый расчет на компьютере на основе модели). С 1981 по 1992 система "Эйдос" неоднократно реализовалась на платформе Wang (на компьютерах Wang-2200C). В 1987 году впервые получен акт внедрения на одну из ранних версий системы «Эйдос», реализованную в среде персональной технологической системы «Вега-М» разработки автора (см.2-й акт по ссылке: <http://lc.kubagro.ru/aidos/aidos02/PR-4.htm>).

2-й этап, «эра IBM PC и MS DOS»: 1992-2012 годы. Для IBM-совместимых персональных компьютеров система "Эйдос" впервые реализована на языках CLIPPER-87 и CLIPPER-5.01 (5.02) в 1992 году, а в 1994 году уже были получены свидетельства РосПатента, первые в Краснодарском крае и, возможно, в России на системы искусственного интеллекта (слева приведена титульная видеограмма финальной DOS-версии системы «Эйдос-12.5», июнь 2012 года). С тех пор и до настоящего времени система непрерывно совершенствуется на IBM PC. Исходные тексты DOS-версии системы «Эйдос» вместе докторской диссертацией по специальности 08.00.13 «Математические инструментальные методы экономики» (экономические, технические и физ.-мат. науки) находятся по ссылке: [http://lc.kubagro.ru/FTP/den\\_Dissert.exe](http://lc.kubagro.ru/FTP/den_Dissert.exe). Незащищенная докторская диссертация по специальности 05.13.01 «Системный анализ, обработка информации и управление» (технические науки) находятся по ссылке: <https://search.rsl.ru/ru/record/01002284002>.

3-й этап, «эра MS Windows XP, 8, 7: 2012-2020 годы. С июня 2012 по 14.12.2020 система «Эйдос» развивалась на языке Аляска-1.9 + Экспресс++ + библиотека для работы с Internet xh2net. Система «Эйдос-X1.9» хорошо работала на всех версиях MS Windows кроме Windows-10, которая требовала специальной настройки. Наиболее трудоемкие в вычислительном отношении операции синтеза моделей и распознавания реализует с помощью графического процессора (GPU), что на некоторых задачах обеспечивает ускорение решение 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).

4-й этап, «эра MS Windows-10: с 2020 года по настоящее время. С 13.12.2020 года по настоящее время система «Эйдос» развивается на языке Аляска-2.0 + Экспресс++. Библиотека xh2net в ней больше не используется, т.к. все возможности работы с Internet входят в базовые

возможности языка программирования. Всегда актуальные исходные тексты текущей версии системы «Эйдос» находятся по адресу: [http://lc.kubagro.ru/\\_AidosALL.txt](http://lc.kubagro.ru/_AidosALL.txt).

Обычно считается, что программирование является не научной, а инженерной работой (приложение 1). Автор всю сознательную жизнь профессионально занимается программированием (с 1976 года) и совершенно не согласен с этой распространённой точкой зрения. По мнению автора, программирование, являясь инженерным трудом, безусловно, включает серьёзную научную компоненту, а также является видом искусства.

### **3.4. Выводы**

По результатам данного раздела можно сделать обоснованный вывод о том, что инфраструктура применения на практике ранее разработанного научного решения проблемы успешно спроектирована и создана, чем научное решение доведено до статуса инновации.

Следующим шагом является *внедрение, применение на практике* новой версии системы «Эйдос», в состав которой входит режим, реализующий предложенные в данной работе идею, концепцию и интеллектуальный алгоритм учета нетипичных случаев при прогнозировании рисков невозврата ссуды.

В данной работе в качестве применения на практике предложенных решений рассмотрим подробный *численный пример*, основанный на реальных данных, находящихся в общем доступе. Для этого необходимо разработать все виды обеспечения внедрения и осуществить само внедрение, что позволит проверить эффективность всех ранее принятых решений и оценить их значение для науки и практики. Все это и является содержанием следующего раздела.

## **4. Применение разработанной технологии на практике, ее внедрение и оценка эффективности решения проблемы**

### **4.1. Методика, план и обеспечение внедрения**

Методика внедрения – это перечень этапов работ или задач, которые необходимо выполнить для осуществления внедрения.

В разделе данной работы путем декомпозиции ее цели получена последовательность **задач**, решение которых является этапами достижения этой цели:

Задача 1: когнитивная структуризация предметной области.

Задача 2: подготовка исходных данных и формализация предметной области (разработка классификационных и описательных шкал и градаций и кодирование исходных с их помощью, т.е. получение обучающей выборки).

Задача 3: синтез и верификация статистических и системно-когнитивных моделей и выбор наиболее достоверной модели. Улучшение

наиболее достоверной модели путем применения итерационного алгоритма разделения обобщенных образов классов на типичную и нетипичную части.

Задача 4: решение различных задач в наиболее достоверной модели:

- подзадача 4.1. Прогнозирование (диагностика, классификация, распознавание, идентификация);
- подзадача 4.2. Поддержка принятия решений;
- подзадача 4.3. Исследование моделируемой предметной области путем исследования ее модели: когнитивные диаграммы классов и значений факторов, агломеративная когнитивная кластеризация классов и значений факторов, нелокальные нейроны и нейронные сети, 3d-интегральные когнитивные карты, когнитивные функции), исследование силы и направления влияния факторов и степени детерминированности классов, обуславливающими их значениями факторов.

План внедрения – это утвержденный руководителем внедрения перечень мероприятий с датами выполнения и ответственными по каждому пункту. В принципе план внедрения это методика внедрения, дополненная различной информацией, необходимой для ее реализации.

В данном случае автор в одном лице совмещал роли и руководителя, и ответственного исполнителя внедрения. Как исполнитель автор решал все поставленные задачи в сроки, которые он заранее жестко не определял, но так, чтобы успеть подготовить материал к очередному номеру журнала.

Виды обеспечения внедрения:

- финансовое (финансирование работ по внедрению);
- организационное (описание схем и регламентов информационного, финансового и материального взаимодействия участников внедрения, логистика);
- юридическое (документы, придающие внедрению и формируемым в его результате отношениям и документам необходимый юридический статус);
- кадровое (специалисты, участвующие во внедрении, их количество и компетенции);
- техническое (компьютерная и орг.техника, компьютерные сети, доступ в Internet, а транспортное обеспечение, здания и сооружения, служебные помещения).

Все эти виды обеспечения внедрения необходимы для его успеха.

Данная работа выполнялась по инициативе и за счет автора на его собственном компьютере.

#### **4.2. Методика оценки эффективности внедрения**

Эффективность внедрения может рассматриваться в натуральном или/и стоимостном выражении.

В натуральном выражении эффективность это в основном количество и качество продукции.

В стоимостном выражении эффективность это в основном прибыль и рентабельность.

В случае данной работы эффективность предложенных решений в *натуральном* выражении может измеряться повышением достоверности моделей прогнозирования, в частности уменьшением количества ложно-отрицательных решений (ошибок не отнесения ситуаций к классам, к которым они на самом деле относятся). В данной работе мы оценим этот вид эффективности в численном примере, основанном на реальных данных.

В стоимостном выражении эффективность предлагаемых решений можно определить путем сравнения прибыли от субсидирования за определённый период до их применения и после их применения. Ожидается увеличение этой прибыли за счет снижения доли неадекватных решений.

### **4.3. Описание внедрения и его результатов (численный пример)**

В данном разделе рассмотрим в интеллектуальной системе «Эйдос» подробный численный пример, основанный на реальных данных. Система «Эйдос» будет использована, т.к. в настоящее время именно она представляет собой программный инструмент автоматизированного системно-когнитивного анализа (АСК-анализ). На этом численном примере мы наглядно проиллюстрируем принятые решения в той форме, в которой они в настоящее время реализованы в системе «Эйдос», и исследуем их эффективность.

В ходе рассмотрения численного примера решим задачи, поставленные в разделах 1.1 и 4.1 данной работы.

#### **4.3.1. Задача 1: когнитивная структуризация предметной области**

На этапе когнитивно-целевой структуризации предметной области мы неформализуемым путем решаем на качественном уровне, что будем рассматривать в качестве факторов, действующих на моделируемый объект (причин), а что в качестве результатов действия этих факторов (последствий). По сути это постановка решаемой проблемы.

Описательные шкалы служат для формального описания факторов, а классификационные – результатов их действия на объект моделирования. Шкалы могут быть числовые и текстовые [21].

При этом необходимо отметить, что статистические и системно-когнитивные модели (СК-модели) отражают лишь сам *факт* наличия зависимостей между значениями факторов и результатами их действия. Но они не отражают причин и механизмов такого влияния. Это значит, что система «Эйдос» выявляет *эмпирические закономерности* в моделируемой предметной области и отображает их в различных формах: табличной, графической и аналитической. Это соответствует

эмпирическому этапу развития. Этим самым она вплотную подводит исследователя к теоретическому уровню познания [23]<sup>6</sup>

Это значит:

– *во-первых*, что содержательная интерпретация СК-моделей – это компетенция специалистов-экспертов хорошо разбирающихся в данной предметной области. Иногда встречается ситуация, когда и то, что на первый взгляд является причинами, и то, что, казалось бы, является их последствиями, на самом деле является последствиями неких глубинных причин, которых мы не видим и никоим образом непосредственно не отражаем в модели<sup>7</sup>;

– *во-вторых*, даже если содержательной интерпретации обнаруженных эмпирических закономерностей не разработано, то в принципе это совершенно не исключает возможности эффективно пользоваться их знанием на *практике* для достижения заданных результатов и поставленных целей, т.е. для управления.

Данная работа основана на исходных данных, размещенных на портале Kaggle: <https://www.kaggle.com/kmlDas/loan-default-prediction>.

По описанию задачи на портале Kaggle можно сделать вывод о том, что ее смысл состоит в том, чтобы по характеристикам получателя ссуды спрогнозировать риск ее невозврата. На основе этого прогноза можно принять обоснованное решение о выдаче или невыдаче ссуды данному конкретному ссудополучателю.

Исходные данные содержат следующие параметры:

- серийный номер или уникальный идентификатор ссудополучателя;
- занятость ссудополучателя (работает ли он или является безработным);
- банковский баланс ссудополучателя;
- годовая зарплата ссудополучателя;
- результат возврата/невозврата ссуды.

В данной работе в качестве классификационной шкалы выберем результат возврата/невозврата ссуды (таблица 1), а в качестве факторов, влияющих на этот результаты – банковский баланс ссудополучателя и его годовую зарплату (таблицы 5 и 6):

---

<sup>6</sup> См., также: [http://lc.kubagro.ru/aidos/Works\\_on\\_identification\\_presentation\\_and\\_use\\_of\\_knowledge.htm](http://lc.kubagro.ru/aidos/Works_on_identification_presentation_and_use_of_knowledge.htm)

<sup>7</sup> Например, если исследовать зависимость поведения людей от положения стрелок часов, то получится довольно тесная взаимосвязь. Но это не означает, что существуют некие физические силы, типа сил гравитации, с помощью которых стрелки часов влияют на поведение людей. На самом деле на поведение людей влияет положение Солнца над горизонтом, а не положение стрелок часов, а *часы просто адекватно отражают это положение Солнца*. Аналогичная ситуация с геномом, который влияет и на почерк, и на успеваемость, поэтому почерк и успеваемость выглядят взаимосвязанными или влияющими друг на друга, хотя на самом деле они связаны не друг с другом, а с геномом.

Таблица 5 – Классификационная

шкала	
Код	Наименование
1	DEFAULTED?

Таблица 6 – Описательные шкалы

Код	Наименование
1	EMPLOYED
2	BANK BALANCE
3	ANNUAL SALARY

#### 4.3.2. Задача 2: подготовка исходных данных и формализация предметной области

Исходные данные для данной работы (рисунок 2) получены непосредственно с портала Kaggle по прямой ссылке: <https://www.kaggle.com/kmldas/loan-default-prediction/download>.

Эти данные представлены в виде CSV-файла. После скачивания этого файла для ввода в систему «Эйдос» с ним было выполнено несколько простых преобразований:

1. CSV-файл был переименован с «Default\_Fin.csv» на «Inp\_data.csv» и размещен в папке: ..\Aidos-X\AID\_DATA\Inp\_data\ системы «Эйдос» для исходных данных табличного типа.

2. CSV-файл был преобразован в XLS-файл для удобства дальнейшей корректировки и ввода в систему «Эйдос».

Само CSV-XLS преобразование (конвертирование) может быть осуществлено онлайн с помощью одного из онлайн-конвертеров. Рекомендуется использовать следующие CSV-XLS-онлайн конвертеры, которые очень хорошо работают со стандартными CSV-файлами:

<https://convertio.co/ru/csv-xls/>,

<https://onlineconvertfree.com/ru/convert-format/csv-to-xls/>;

<https://document.online-convert.com/ru/convert/csv-to-excel>.

В простейшем случае CSV-файл это текст, состоящий из строк, в каждой из которых содержится *одинаковое* количество элементов, разделенных каким-либо разделителем, чаще всего запятой. Таким образом, строки CSV-файла можно поставить в соответствие строкам таблицы, а элементы строк – колонкам таблицы.

Но следует иметь в виду, что сам CSV-стандарт (форматированный текст) еще не совсем устоялся. Но в CSV-файлах в качестве разделителя могут быть использованы и другие символы, например, точка с запятой или табуляция. Иногда, когда необходимо, чтобы внутри элементов использовалась запятая, эти элементы выделяют кавычками. Поэтому иногда (достаточно редко) встречаются CSV-файлы с необычными форматами, которые не всякий конвертер сможет корректно преобразовать. В этом случае рекомендуется попробовать подобрать другой конвертер, которых очень много в открытом доступе. Потратив на это некоторое время, обычно удается получить желаемый результат.

3. После преобразования CSV-файла в XLS-файл в нем средствами MS-Excel были произведены следующие корректировки:

– в колонках Employed и Defaulted логические переменные 0/1 замены на Yes, No соответственно;

– колонка Defaulted, советующая классификационной шкале, выделена желтым фоном;

– в числовых колонках сделано два знака после запятой (по числу максимальному числу разрядов после запятой, которое встречается в исходных данных в данной колонке);

– при разработке реальных научных интеллектуальных приложений убедительно рекомендуется в числовых колонах в обязательном порядке указывать единицы измерения, в нашем случае это доллары США;

– XLS-файл (стандарт MS Excel-2003) записан в стандарте более новых версий MS Excel как XLSX. Это сделано потому, что в новом стандарте файл имеет размер примерно в два раза меньше, чем в старом.

В результате всех этих операций получилась таблица исходных данных (таблица 7):

Таблица 7 – Исходные данные для ввода в систему «Эйдос» (фрагмент)

Index	Employed	Bank Balance, \$	Annual Salary, \$	Defaulted?
1	Yes	8754,36	532339,56	No
2	No	9806,16	145273,56	No
3	Yes	12882,60	381205,68	No
4	Yes	6351,00	428453,88	No
5	Yes	9427,92	461562,00	No
6	No	11035,08	89898,72	No
7	Yes	9906,12	298862,76	No
8	No	9704,04	211205,40	No
9	Yes	13932,72	449622,36	No
10	Yes	0,00	351303,24	No
11	No	0,00	262452,84	No
12	No	14646,96	159222,72	No
13	Yes	2844,60	339020,40	No
14	Yes	7280,88	539934,72	No
15	Yes	13355,64	285722,04	No
16	Yes	3434,76	540508,92	No
17	Yes	0,00	603183,72	No
18	No	6330,48	211638,48	No
19	Yes	5831,28	738793,32	No
20	Yes	13140,84	317575,56	No
21	Yes	2747,40	606002,16	No
22	Yes	11451,12	389490,12	No
23	Yes	12671,52	615814,56	No
24	Yes	7703,76	365593,20	No
25	Yes	9278,52	412239,72	No
26	Yes	10260,12	302535,96	No
27	Yes	7716,00	497682,12	No
28	Yes	17458,32	386269,08	No
29	Yes	7388,40	472516,68	No
30	No	13434,84	198672,84	No
31	Yes	5937,84	652617,36	No
32	No	5386,56	189593,64	No
33	No	7018,80	269159,28	No
34	Yes	10963,08	562886,76	No
35	No	17087,28	271613,88	No
36	No	17996,64	158287,80	No
37	Yes	8304,36	573679,32	No

В таблице 3 приведен лишь небольшой фрагмент исходных данных, т.к. в этой таблице 10000 строк. Полностью ее можно скачать из Эйдос-облака по прямой ссылке:

[http://aidos.byethost5.com/Source\\_data\\_applications/Applications-000293/Inp\\_data.xlsx](http://aidos.byethost5.com/Source_data_applications/Applications-000293/Inp_data.xlsx).

Затем с параметрами, показанными на рисунке 4, запустим режим 2.3.2.2 системы «Эйдос», представляющий собой автоматизированный программный интерфейс (API) с внешними числовыми и текстовыми данными табличного типа:

2.3.2.2. Универсальный программный интерфейс импорта данных в систему «ЭЙДОС-X»

Автоматическая формализация предметной области: генерация классификационных и описательных шкал и градаций, а также обучающей и распознаваемой выборки на основе базы исходных данных: "Inp\_data"

Задайте тип файла исходных данных: "Inp\_data":

XLS - MS Excel 2003      Стандарт XLS-файла

XLSV - MS Excel 2007(2010)      Стандарт XLSV-файла

DBF - DBASE IV (DBF/NTX)      Стандарт DBF-файла

CSV - CSV => DBF конвертер      Стандарт CSV-файла

Задайте диапазон столбцов классификационного шкала:

Начальный столбец классификационного шкала:

Конечный столбец классификационного шкала:

Задайте диапазон столбцов описательного шкала:

Начальный столбец описательного шкала:

Конечный столбец описательного шкала:

Задайте режим:

Формализация предметной области (на основе "Inp\_data")

Генерация распознаваемой выборки (на основе "Inp\_test")

Задайте способ выбора размера интервалов:

Разные интервалы с разным числом наблюдений

Разные интервалы с равным числом наблюдений

Задайте параметры формирования сценариев или способа интерпретации текстовых полей "Inp\_data":

Не применять сценарный метод АКК-анализа

Применить сценарный метод АКК-анализа

**Параметры интерпретации значений текстовых полей "Inp\_data":**

Интерпретация TXT-полей классов: Значения полей текстовых классификационных шкалов исходных данных "Inp\_data" рассматриваются как целые

Интерпретация TXT-полей признаков: Значения полей текстовых описательных шкалов исходных данных "Inp\_data" рассматриваются как целые

Какие наименования ГРАДАЦИЙ числовых шкалов использовать:

Только интервальные числовые значения (например: "1/3-(59873.000000, 178545.666667)")

Только наименования интервальных числовых значений (например: "Минимальное")

И интервальные числовые значения, и их наименования (например: "Минимальное: 1/3-(59873.000000, 178545.666667)")

OK      Cancel

---

2.3.2.2. Задание размерности модели системы «ЭЙДОС-X»

ЗАДАНИЕ В ДИАЛОГЕ РАЗМЕРНОСТИ МОДЕЛИ

Сценарное количество градаций классификационных и описательных шкал (2 x 8)

Тип шкалы	Количество классификационных шкалов	Количество градаций классификационных шкалов	Среднее количество градаций на классификационный шкал	Количество описательных шкалов	Количество градаций описательных шкалов	Среднее количество градаций на описательный шкал
Числовые	0	0	0,00	2	6	3,00
Текстовые	1	2	2,00	1	2	2,00
ВСЕГО:	1	2	2,00	3	8	2,67

Задайте число интервалов (градаций) в шкале:

В описательных шкалах:

Пересчитать шкалы и градации      Параметры числовых шкалов и градаций      Выйти на создание модели

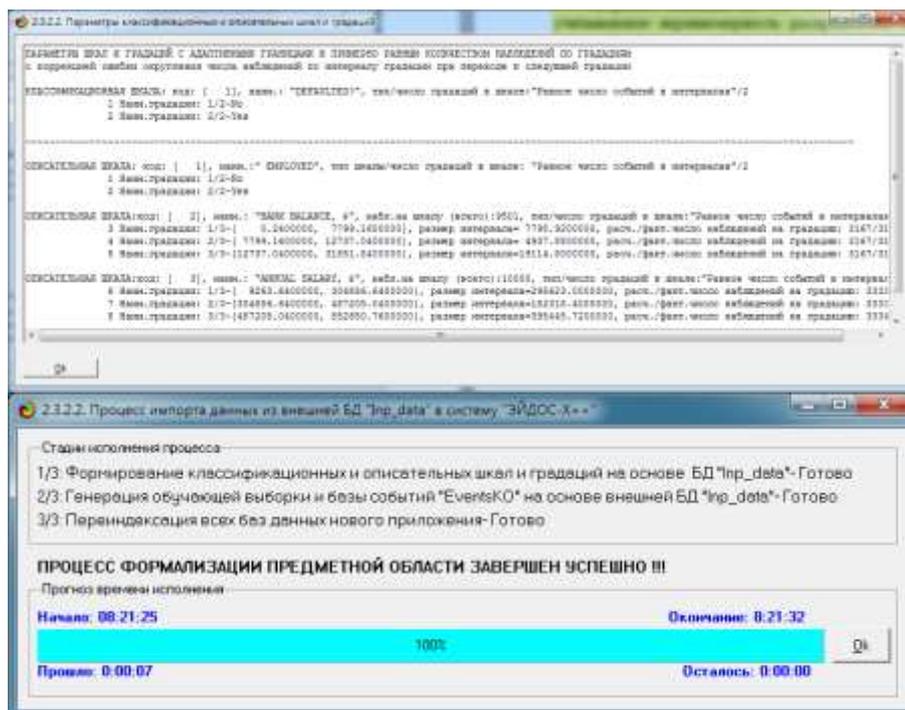


Рисунок 4. Экранная форма программного интерфейса (API) 2.3.2.2 системы «Эйдос» с внешними данными табличного типа

Как видно из рисунка 4 весь процесс ввода исходных данных в систему «Эйдос» занял 7 секунд.

Обратим внимание на то, что заданы адаптивные интервалы, учитывающие неравномерность распределения данных по диапазону значений, что важно при относительно небольшом числе наблюдений. Если бы интервалы были заданы равными по величине, то в различные интервалы попало бы сильно отличающееся число наблюдений, а в некоторых интервалах их бы могло не оказаться вовсе.

Классификационная шкала текстового типа, поэтому для нее число интервалов не задается.

В описательных шкалах задано 3 адаптивных числовых интервала. Как видно из рисунка 4 на каждое интервальное числовое значение приходится около 3167 наблюдений.

На рисунке 5 приведен исчерпывающий Help данного режима, в котором объясняется принцип организации таблицы исходных данных для данного режима. Здесь же обратим внимание на то, что в таблице 3 значения параметров могут быть представлены как числовыми, так и текстовыми значениями.

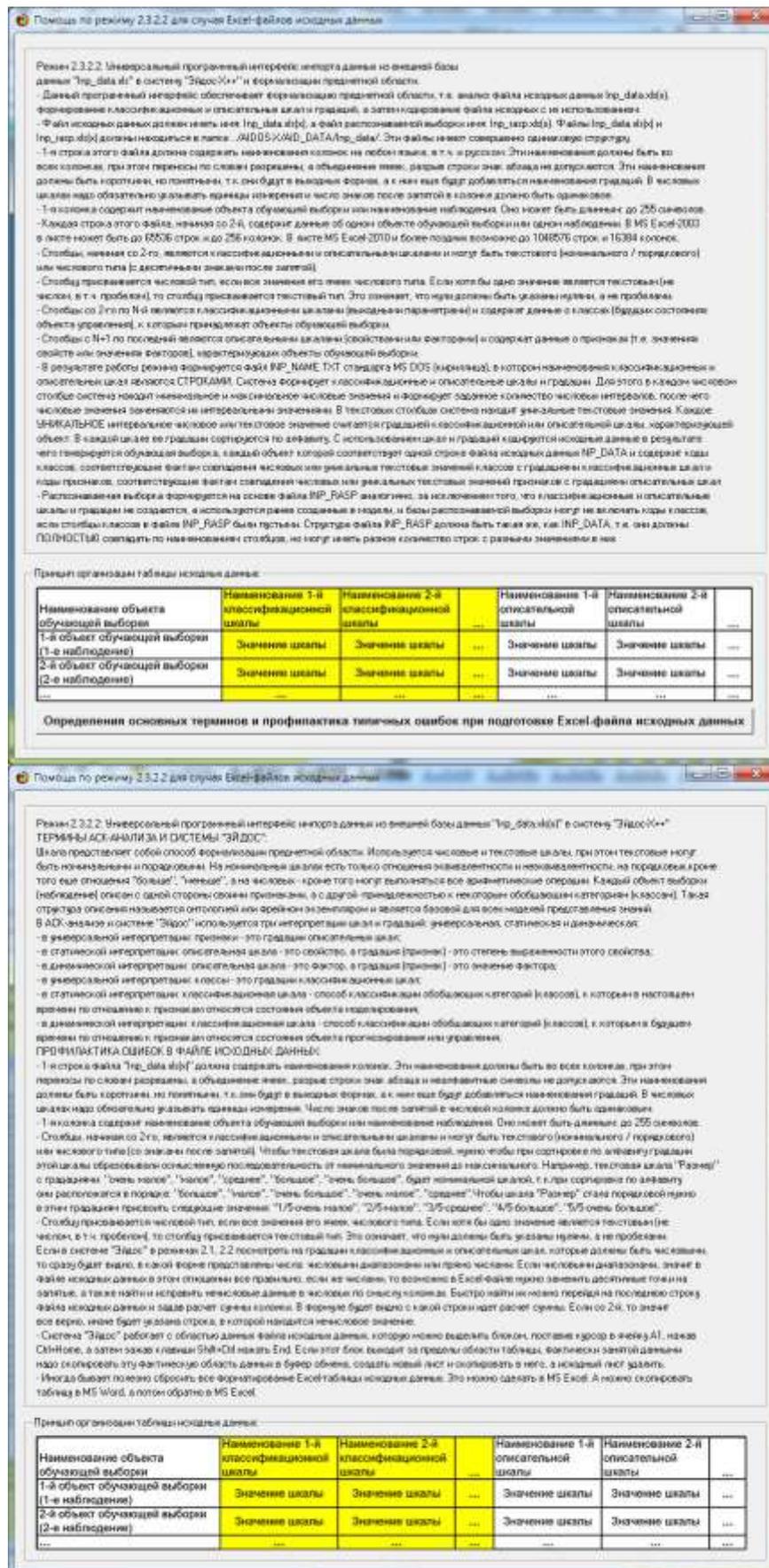


Рисунок 5. Экранные формы HELP программного интерфейса (API) 2.3.2.2

В результате работы API-2.3.2.2 сформирована одна классификационная шкала с суммарным количеством градаций (классов) две (таблица 8) и 3 описательных шкалы с суммарным числом градаций 45 (таблица 9).

С использованием классификационных и описательных шкал и градаций исходные данные (таблица 7) были закодированы и в результате получена обучающая выборка (таблица 10):

Таблица 8 – Классификационные шкалы и градации (возврат/невозврат ссуды)

Код	Наименование
1	DEFAULTED?-1/2-No
2	DEFAULTED?-2/2-Yes

Таблица 9 – Описательные шкалы и градации (характеристики ссудополучателей)

Код	Наименование
1	EMPLOYED-1/2-No
2	EMPLOYED-2/2-Yes
3	BANK BALANCE, \$-1/3-{0.2400000, 7799.1600000}
4	BANK BALANCE, \$-2/3-{7799.1600000, 12737.0400000}
5	BANK BALANCE, \$-3/3-{12737.0400000, 31851.8400000}
6	ANNUAL SALARY, \$-1/3-{9263.6400000, 304886.6400000}
7	ANNUAL SALARY, \$-2/3-{304886.6400000, 487205.0400000}
8	ANNUAL SALARY, \$-3/3-{487205.0400000, 882650.7600000}

Таблица 10 – Обучающая выборка (фрагмент)

ID ссудополучателя	EMPLOYED	BANK BALANCE, \$-	ANNUAL SALARY, \$-	DEFAULTED?-
1	2	4	8	1
2	1	4	6	1
3	2	5	7	1
4	2	3	7	1
5	2	4	7	1
6	1	4	6	1
7	2	4	6	1
8	1	4	6	1
9	2	5	7	1
10	2		7	1
11	1		6	1
12	1	5	6	1
13	2	3	7	1
14	2	3	8	1
15	2	5	6	1
16	2	3	8	1
17	2		8	1
18	1	3	6	1
19	2	3	8	1
20	2	5	7	1
21	2	3	8	1
22	2	4	7	1
23	2	4	8	1
24	2	3	7	1
25	2	4	7	1
26	2	4	6	1
27	2	3	8	1
28	2	5	7	1
29	2	3	7	1
30	1	5	6	1
31	2	3	8	1
32	1	3	6	1
33	1	3	6	1
34	2	4	8	1
35	1	5	6	1
36	1	5	6	1
37	2	4	8	1

Обучающая выборка (таблица 5), по сути, представляет собой нормализованные исходные данные, т.е. таблицу исходных данных (таблица 3), закодированную с помощью классификационных и описательных шкал и градаций (таблицы 4 и 5).

Таким образом, в результате формализации предметной области созданы все необходимые и достаточные условия для выполнения следующего этапа АСК-анализа: т.е. для синтеза и верификации моделей.

### **4.3.3. Задача 3: применение итерационного алгоритма разделения обобщенных образов классов на типичную и нетипичную части**

В соответствии с предложенным интеллектуальным итерационным алгоритмом учета нетипичных случаев при прогнозировании невозврата ссуды сначала создадим базовую (исходную) модель, а затем будем повышать ее достоверность путем создания в исходных данных классов для нетипичных случаев и синтеза и верификации моделей в последовательных итерациях. Этот процесс останавливается тогда, когда модели перестают улучшаться в его результате.

#### **4.3.3.1. Исходная модель**

##### **4.3.3.1.1. Синтез и верификация статистических и системно-когнитивных моделей**

Синтез и верификация статистических и системно-когнитивных моделей (СК-моделей) осуществляется в режиме 3.5 системы «Эйдос» (рисунок 6).

Обратим внимание на то, что на рисунке 6 в правом нижнем углу окна задана опция: «Расчеты проводить на графическом процессоре (GPU)».

Из рисунка 6 видно, что весь процесс синтеза и верификации моделей занял 5 минут 34 секунды. Отметим, что при синтезе и верификации моделей использовался графический процессор (GPU) видеокарты. На центральном процессоре (CPU) выполнение этих операций занимает значительно большее время (на некоторых задачах это происходит в десятки, сотни и даже тысячи раз дольше). Таким образом, неграфические вычисления на графических процессорах видеокарты делает возможной обработку больших объемов исходных данных за разумное время. В процесс синтеза и верификации моделей осуществляется также расчет 10 выходных форм, на что уходит более 99% времени исполнения.

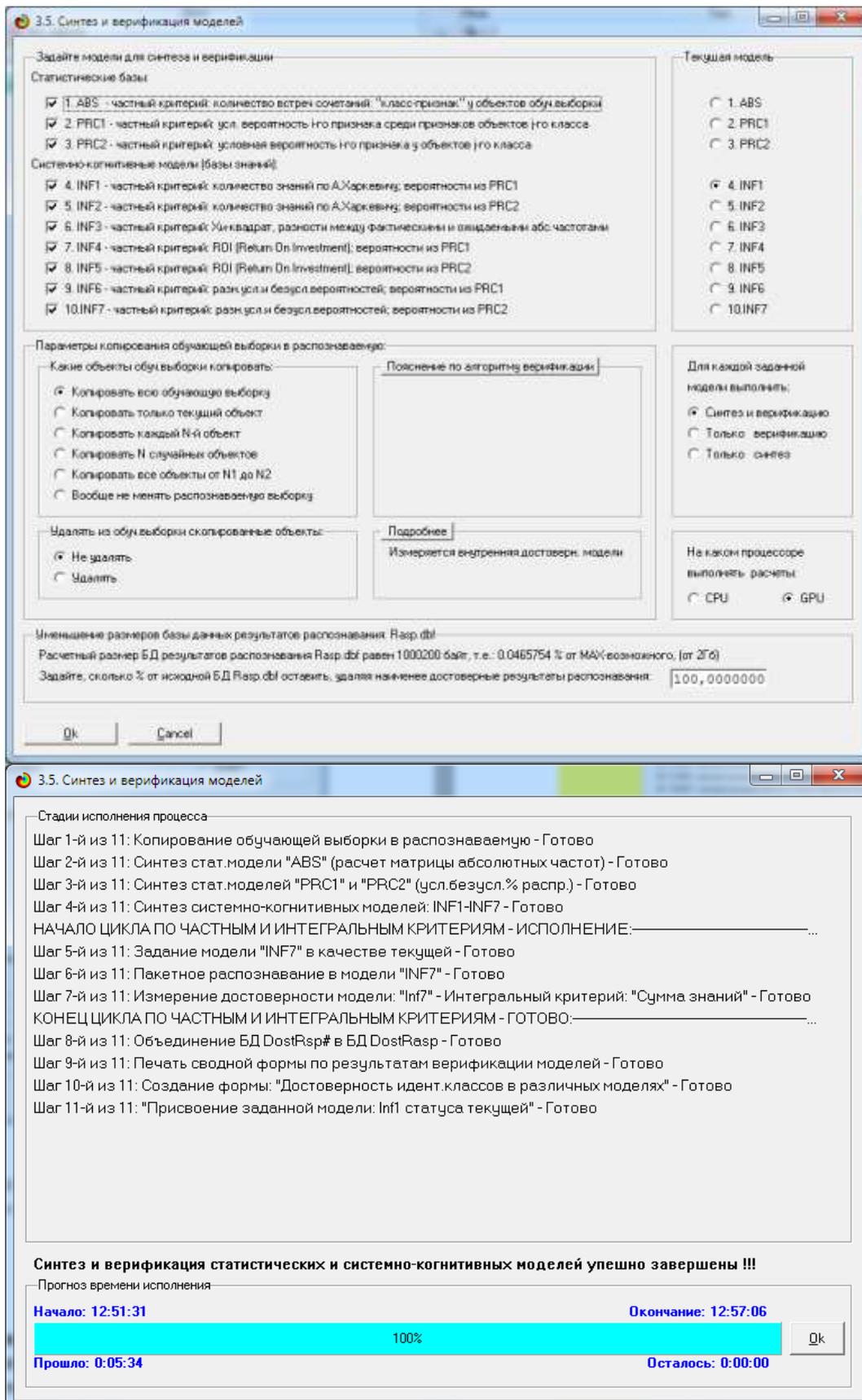


Рисунок 6. Экранные формы режима синтеза и верификации моделей системы «Эйдос» (режим 3.5)

Некоторые созданные статистические и системно-когнитивные модели (СК-модели) приведены на рисунках 7 - 10:

5.5. Модели: "1. ABS - частный критерий: количество встреч сочетаний: "Класс-признак" у объектов обуч.выборки"

Код признака	Наименование описательной шкалы и градации	1. DEFAULTED? 1/2 NO	2. DEFAULTED? 2/2 YES	Сумма	Среднее	Средн. квадрат. откл.
1	EMPLOYED-1/2-No	2817	127	2944	1472.00	1902.12
2	EMPLOYED-2/2-Yes	6850	206	7056	3528.00	4690.02
3	BANK BALANCE, \$-1/3-(0.2400000, 7799.1600000)	3167		3167	1583.50	2239.41
4	BANK BALANCE, \$-2/3-(7799.1600000, 12737.0400000)	3158	9	3167	1583.50	2224.68
5	BANK BALANCE, \$-3/3-(12737.0400000, 31851.8400000)	2843	324	3167	1583.50	1781.20
6	ANNUAL SALARY, \$-1/3-(9263.6400000, 304886.6400000)	3199	134	3333	1666.50	2167.28
7	ANNUAL SALARY, \$-2/3-(304886.6400000, 487205.0400000)	3235	98	3333	1666.50	2218.19
8	ANNUAL SALARY, \$-3/3-(487205.0400000, 882650.7600000)	5233	101	3334	1667.00	2214.66
	Сумма числа признаков	28502	999	29501		
	Среднее	3563	129		1843.61	
	Среднеквадратичное отклонение	1339	105			917.45
	Сумма числа объектов обуч. выборки	9667	333	10000		

Рисунок 7. Матрица абсолютных частот: статистическая Модель ABS

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

Код признака	Наименование описательной шкалы и градации	1. DEFAULT... 1/2 NO	2. DEFAULT... 2/2 YES	Безусл. вероятн.	Среднее	Средн. квадрат. откл.
1	EMPLOYED-1/2-No	29.140	38.138	29.440	33.639	6.362
2	EMPLOYED-2/2-Yes	70.860	61.862	70.560	66.361	6.362
3	BANK BALANCE, \$-1/3-(0.2400000, 7799.1600000)	32.761		31.670	16.380	29.165
4	BANK BALANCE, \$-2/3-(7799.1600000, 12737.0400000)	32.668	2.703	31.670	17.685	21.189
5	BANK BALANCE, \$-3/3-(12737.0400000, 31851.8400000)	29.409	97.297	31.670	63.353	48.004
6	ANNUAL SALARY, \$-1/3-(9263.6400000, 304886.6400000)	33.092	40.240	33.330	36.666	5.055
7	ANNUAL SALARY, \$-2/3-(304886.6400000, 487205.0400000)	33.464	29.429	33.330	31.447	2.853
8	ANNUAL SALARY, \$-3/3-(487205.0400000, 882650.7600000)	33.444	30.330	33.340	31.887	2.201
	Сумма	294.838	300.000	594.838		
	Среднее	36.855	37.500		37.177	
	Среднеквадратичное отклонение	13.851	31.421			23.457

Рисунок 8. Матрица условных и безусловных процентных распределений: статистическая модель PRC2

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

Код признака	Наименование описательной шкалы и градации	1. DEFAULTED? 1/2 NO	2. DEFAULTED? 2/2 YES	Сумма	Среднее	Средн. квадрат. откл.
1	EMPLOYED-1/2-No	-0.001	0.024	0.023	0.011	0.017
2	EMPLOYED-2/2-Yes	0.000	-0.014	-0.014	-0.007	0.011
3	BANK BALANCE, \$-1/3-(0.2400000, 7799.1600000)	0.003		0.003	0.002	0.002
4	BANK BALANCE, \$-2/3-(7799.1600000, 12737.0400000)	0.003	-0.241	-0.238	-0.119	0.172
5	BANK BALANCE, \$-3/3-(12737.0400000, 31851.8400000)	-0.007	0.107	0.100	0.050	0.081
6	ANNUAL SALARY, \$-1/3-(9263.6400000, 304886.6400000)	-0.001	0.017	0.016	0.008	0.012
7	ANNUAL SALARY, \$-2/3-(304886.6400000, 487205.0400000)	0.000	-0.014	-0.013	-0.007	0.010
8	ANNUAL SALARY, \$-3/3-(487205.0400000, 882650.7600000)	0.000	-0.011	-0.010	-0.005	0.006
	Сумма	-0.001	-0.132	-0.133		
	Среднее	0.000	-0.017		-0.008	
	Среднеквадратичное отклонение	0.003	0.099			0.068

Рисунок 9. Матрица информативностей: СК-модель INF1

Код признака	Наименование описательной шкалы и градаций	1. DEFAULTED? 1/2NO	2. DEFAULTED? 2/2YES	Сумма	Среднее	Средн. квар. откл.
1	EMPLOYED-1/2No	-27.307	27.307			38.617
2	EMPLOYED-2/2Yes	32.939	-32.939			46.593
3	BANK BALANCE, \$-1/3-(0.2400000, 7799.1600000)	107.245	-107.245	0.000	0.000	151.667
4	BANK BALANCE, \$-2/3-(7799.1600000, 12737.0400000)	98.245	-98.245	0.000	0.000	138.939
5	BANK BALANCE, \$-3/3-(12737.0400000, 31851.8400000)	-216.755	216.755	0.000	0.000	306.538
6	ANNUAL SALARY, \$-1/3-(9263.6400000, 304886.6400000)	-21.134	21.134	0.000	0.000	29.898
7	ANNUAL SALARY, \$-2/3-(304886.6400000, 487205.0400000)	14.866	-14.866	0.000	0.000	21.024
8	ANNUAL SALARY, \$-3/3-(487205.0400000, 882650.7600000)	11.900	-11.900	0.000	0.000	16.829
	Сумма	0.000	0.000	0.000		
	Среднее	0.000	0.000		0.000	
	Среднеквадратичное отклонение	100.553	100.553			97.143

Рисунок 10. Матрица хи-квадрат: СК-модель INF3

Отметим, что в АСК-анализе и СК-моделях степень выраженности различных свойств объектов наблюдения рассматривается с единственной точки зрения: с точки зрения того, какое **количество информации** содержится в них о том, к каким обобщающим категориям (классам) будут принадлежать или не принадлежать эти объекты. Поэтому не играет никакой роли, в каких единицах измерения измеряются те или иные свойства объектов наблюдения, а также в каких единицах измеряются результаты влияния этих свойств, натуральных, в процентах или стоимостных [21]. Это и есть решение проблемы сопоставимости в АСК-анализе и системе «Эйдос», отличающее их от других интеллектуальных технологий.

#### 4.3.3.1.2. Оценка достоверности моделей

Оценка достоверности моделей в системе «Эйдос» осуществляется путем решения задачи классификации объектов обучающей выборки по обобщенным образам классов и подсчета количества истинных и ложных положительных и отрицательных решений по F-мере Ван Ризбергера, а также по критериям L1- L2-мерам проф. Е.В.Луценко, которые предложены для того, чтобы смягчить или полностью преодолеть некоторые недостатки F-меры [24]. В режиме 3.4 системы «Эйдос» изучается достоверность каждой частной модели в соответствии с этими мерами достоверности (рисунок 11).

Из рисунка 11 мы видим, что в данном интеллектуальном приложении по F-критерию Ван Ризбергера наиболее достоверной является СК-модель INF3 с интегральным критерием «Сумма знаний» ( $F=0,708$  при максимуме 1,000), что является неплохим результатом для моделируемой предметной области.

*Это подтверждает наличие и адекватное отражение в СК-модели INF3 сильной причинно-следственной зависимости между*

характеристиками ссудополучателей и возвратом или невозвратом ими полученной ссуды.

На рисунке 12 приведено частотное распределения числа истинных и ложных положительных и отрицательных решений по результатам прогнозирования возврата/невозврата ссуды в СК-модели INF3 по данным обучающей выборки.

Наименование модели (идентификатор)	Интегральный критерий	Число истинных решений (TP)	Число ложных положительных решений (FP)	Число истинных отрицательных решений (TN)	Число ложных отрицательных решений (FN)	Число верных решений	Сумма квадратов ошибок (MSE)	Среднее значение ошибок (MSE)										
1. A2D1 частный критерий на количество истинно-сходных "х.зн."	Критерий ИФЧ на счет с обр.	2046	2026	0,683	2,727	0,882	4934,690	1400,396	2562,902	1242,544	0,452	0,799	0,722					
2. A2D2 частный критерий на количество истинно-сходных "х.зн."	Критерий ИФЧ, счет на прогноз.	2002	2026	0,683	1,000	0,882	4924,811	1400,396	2562,902	1242,544	0,452	0,799	0,646					
3. PR21 частный критерий: доля верных решений по признаку ссуды.	Критерий доли верных решений с а.	2046	2026	0,683	2,727	0,882	4934,690	1400,396	2562,902	1242,544	0,452	0,799	0,722					
4. PR22 частный критерий: доля верных решений по признаку ссуды.	Критерий доли верных решений с а.	2002	2026	0,683	1,000	0,882	4924,811	1400,396	2562,902	1242,544	0,452	0,799	0,646					
5. PR23 частный критерий: доля верных решений по признаку ссуды.	Критерий доли верных решений с а.	2046	2026	0,683	2,727	0,882	4934,690	1400,396	2562,902	1242,544	0,452	0,799	0,722					
6. PR24 частный критерий: доля верных решений по признаку ссуды.	Критерий доли верных решений с а.	2002	2026	0,683	1,000	0,882	4924,811	1400,396	2562,902	1242,544	0,452	0,799	0,646					
7. PR25 частный критерий: доля верных решений по признаку ссуды.	Критерий доли верных решений с а.	2046	2026	0,683	2,727	0,882	4934,690	1400,396	2562,902	1242,544	0,452	0,799	0,722					
8. PR26 частный критерий: доля верных решений по признаку ссуды.	Критерий доли верных решений с а.	2002	2026	0,683	1,000	0,882	4924,811	1400,396	2562,902	1242,544	0,452	0,799	0,646					
9. PR27 частный критерий: доля верных решений по признаку ссуды.	Критерий доли верных решений с а.	2046	2026	0,683	2,727	0,882	4934,690	1400,396	2562,902	1242,544	0,452	0,799	0,722					
10. PR28 частный критерий: доля верных решений по признаку ссуды.	Критерий доли верных решений с а.	2002	2026	0,683	1,000	0,882	4924,811	1400,396	2562,902	1242,544	0,452	0,799	0,646					
11. PR29 частный критерий: доля верных решений по признаку ссуды.	Критерий доли верных решений с а.	2046	2026	0,683	2,727	0,882	4934,690	1400,396	2562,902	1242,544	0,452	0,799	0,722					
12. PR30 частный критерий: доля верных решений по признаку ссуды.	Критерий доли верных решений с а.	2002	2026	0,683	1,000	0,882	4924,811	1400,396	2562,902	1242,544	0,452	0,799	0,646					
13. PR31 частный критерий: доля верных решений по признаку ссуды.	Критерий доли верных решений с а.	2046	2026	0,683	2,727	0,882	4934,690	1400,396	2562,902	1242,544	0,452	0,799	0,722					
14. PR32 частный критерий: доля верных решений по признаку ссуды.	Критерий доли верных решений с а.	2002	2026	0,683	1,000	0,882	4924,811	1400,396	2562,902	1242,544	0,452	0,799	0,646					
15. PR33 частный критерий: доля верных решений по признаку ссуды.	Критерий доли верных решений с а.	2046	2026	0,683	2,727	0,882	4934,690	1400,396	2562,902	1242,544	0,452	0,799	0,722					
16. PR34 частный критерий: доля верных решений по признаку ссуды.	Критерий доли верных решений с а.	2002	2026	0,683	1,000	0,882	4924,811	1400,396	2562,902	1242,544	0,452	0,799	0,646					
17. PR35 частный критерий: доля верных решений по признаку ссуды.	Критерий доли верных решений с а.	2046	2026	0,683	2,727	0,882	4934,690	1400,396	2562,902	1242,544	0,452	0,799	0,722					
18. PR36 частный критерий: доля верных решений по признаку ссуды.	Критерий доли верных решений с а.	2002	2026	0,683	1,000	0,882	4924,811	1400,396	2562,902	1242,544	0,452	0,799	0,646					
19. PR37 частный критерий: доля верных решений по признаку ссуды.	Критерий доли верных решений с а.	2046	2026	0,683	2,727	0,882	4934,690	1400,396	2562,902	1242,544	0,452	0,799	0,722					
20. PR38 частный критерий: доля верных решений по признаку ссуды.	Критерий доли верных решений с а.	2002	2026	0,683	1,000	0,882	4924,811	1400,396	2562,902	1242,544	0,452	0,799	0,646					

Рисунок 11. Экранная форма с информацией о достоверности моделей по F-критерию Ван Ризбергера

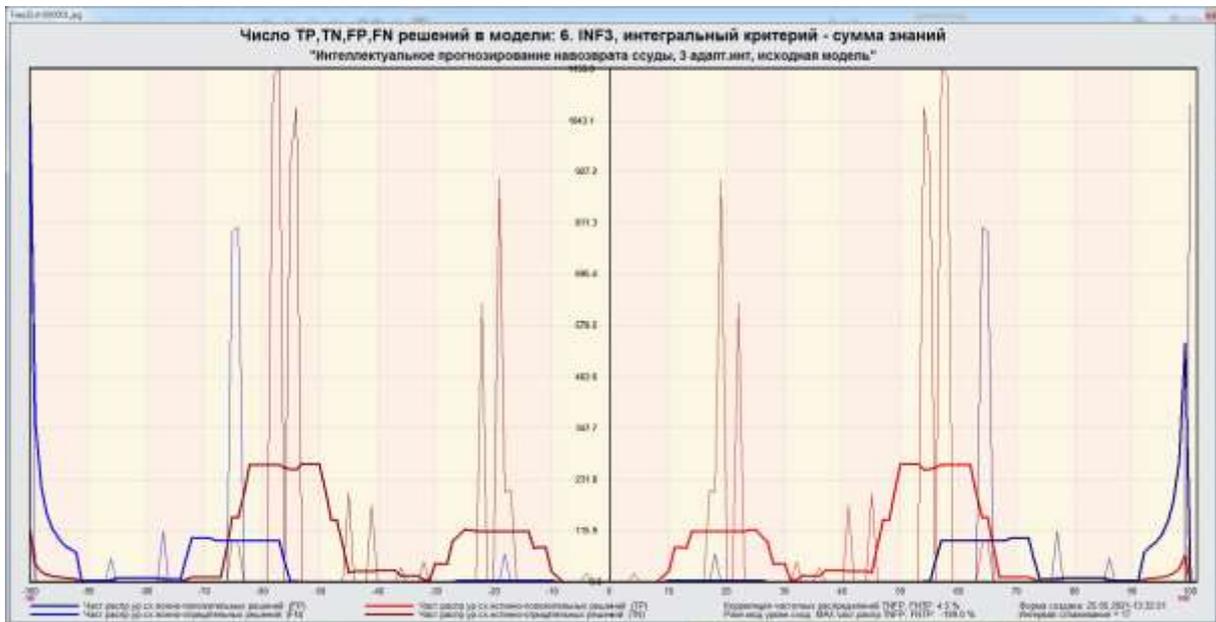


Рисунок 12. Частотные распределения числа истинных и ложных положительных и отрицательных решений и их разности в СК-модели Inf3

Рисунок 12 содержит изображения частотных распределений количества истинных и ложных положительных и отрицательных решений в зависимости от уровня сходства. Из этого рисунка мы видим, что в районе высоких уровней сходства есть аномальный выброс ложноположительных и ложноотрицательных решений, которого не должно

быть в разумной модели. Возможно, этот выброс связан с изменениями, внесенными их автором, разместившим их на портале Kaggle. Он упоминает об этих изменениях и объясняет их необходимостью обезличивания данных.

#### 4.3.3.1.3. Задание текущей модели и распознавание в ней

Важно иметь в виду, что корректируется (путем разделения классов на типичную и нетипичную части) всегда та модель, в которой проводилось распознавание. Если данный режим 3.7.6 запускается сразу после режима синтеза и верификации моделей 3.5 – то это будет СК-модель INF7, как последняя, в которой проводилась верификация в режиме 3.5. Если мы в режиме 3.4 выбрали другую модель для улучшения как наиболее достоверную, то надо ее задать в качестве текущей в режиме 5.6, а затем выполнить **распознавание** в этой модели в режиме 4.1.2.

Поэтому далее выполняем режим 5.6 (рисунки 2 и 13), а затем режим **распознавания** 4.1.2 в текущей модели INF3 (рисунок 14).

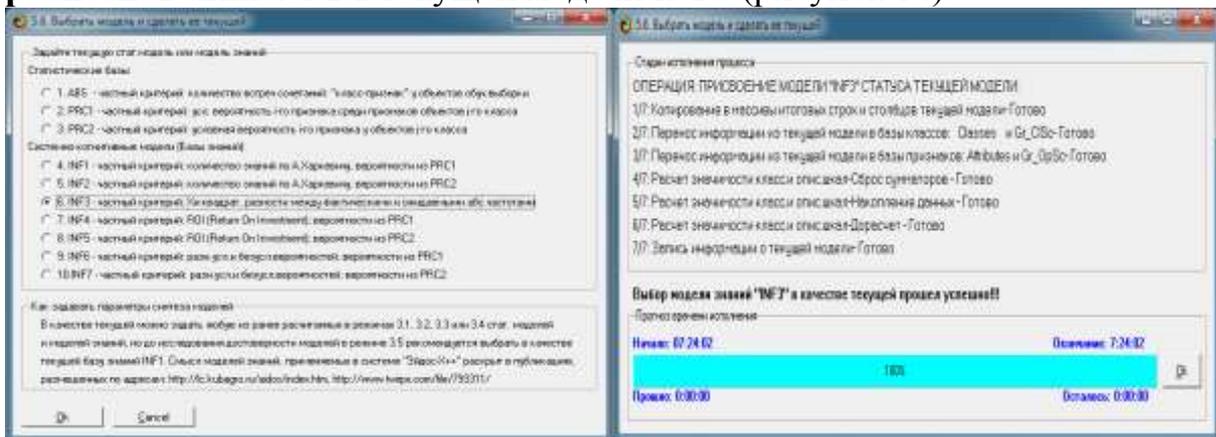


Рисунок 13. Экранные формы придания наиболее достоверной СК-модели Inf3 статуса текущей модели

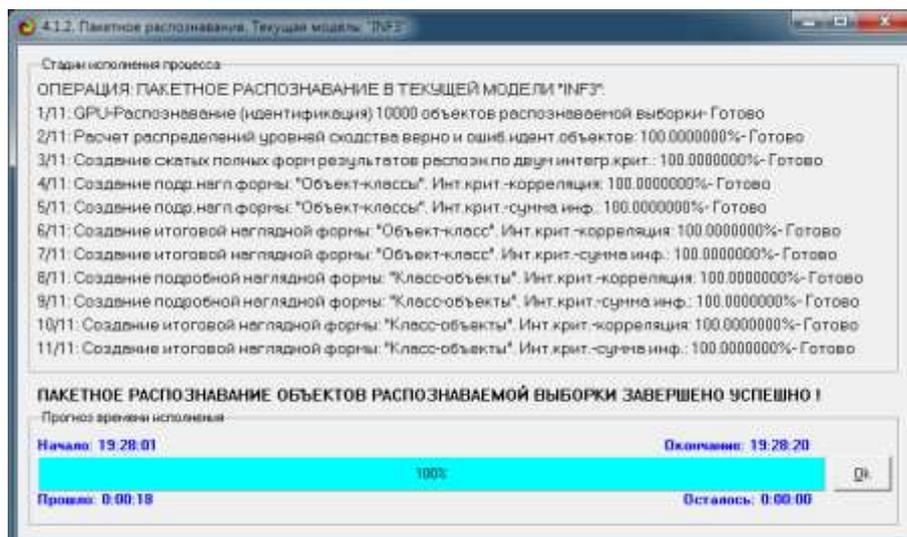


Рисунок 14. Экранная форма пакетного распознавания в текущей СК-модели Inf3

Из экранной формы на рисунке 14 видно, что процесс идентификации 10000 объектов распознаваемой выборки шел 18 секунд. Для самого распознавания использовался графический процессор (GPU), а точнее 1500 шейдерных процессоров видеокарты NVIDIA GeForce GTX 770. Для расчета 10 выходных форм по результатам распознавания использовался центральный процессор (CPU) i7. В основном время было затрачено именно на расчет выходных форм.

#### 4.3.3.2. Первая итерация. Синтез и верификация статистических и системно-когнитивных моделей. Оценка достоверности моделей. Задание текущей модели и распознавание в ней

Для выполнения 1-й итерации запускаем режим 3.7.6 (рисунок 15):

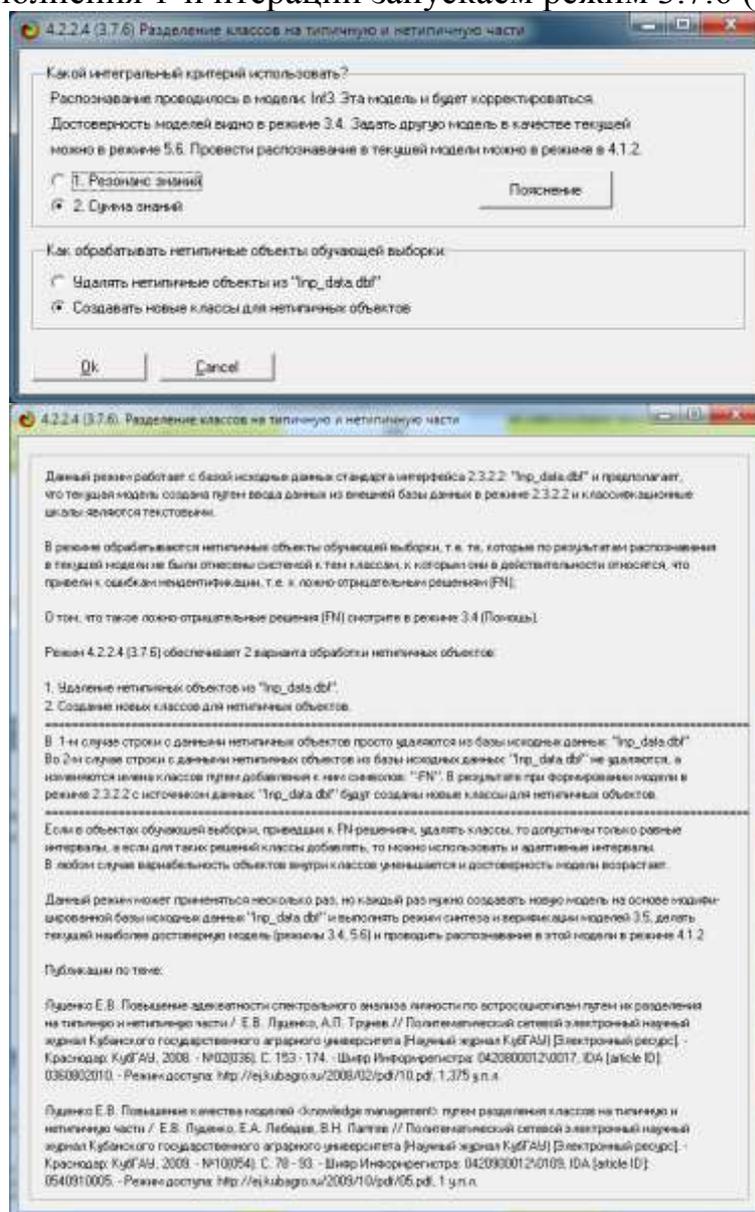


Рисунок 15. Экранные формы режима разделения классов на типичную и нетипичную части

После запуска режима появляется прогресс-бар и после него окно с сообщением и запускается API-2.3.2.2 с параметрами, сформированными данным режимом (рисунок 16). Затем выполняем режимы, следуя инструкциям системы (рисунки 17)<sup>8</sup>.

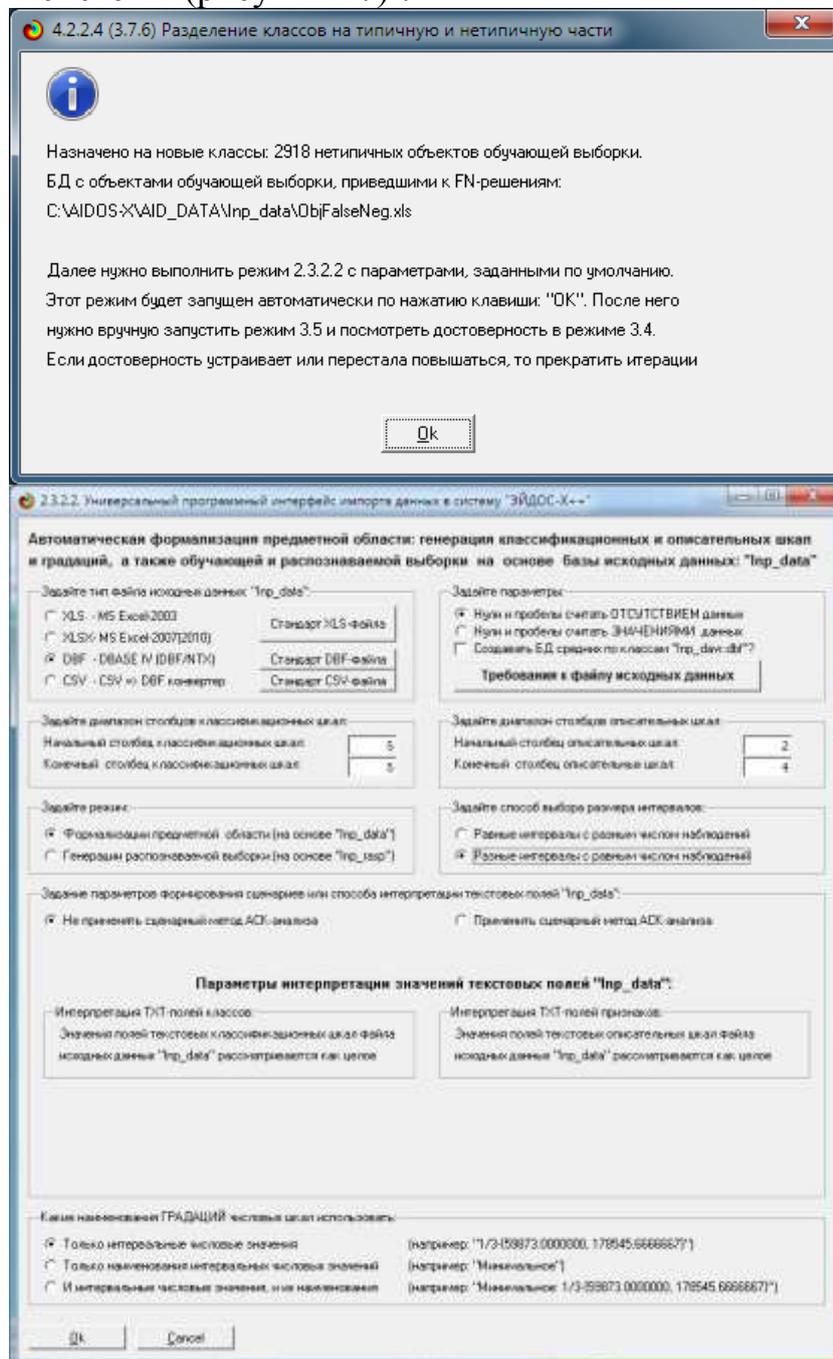


Рисунок 16. Экранные формы режима разделения классов на типичную и нетипичную части

<sup>8</sup> Мы эти рисунки, полученные на 1-й итерации, не комментируем, т.к. их наименования есть в наименовании экранных форм и они полностью аналогичны полученным в исходной модели.

2.3.2.2. Задание размерности модели системы "ЭЙДОС-X++"

ЗАДАНИЕ В ДИАЛОГЕ РАЗМЕРНОСТИ МОДЕЛИ  
Суммарное количество градаций классификационных и описательных шкал (4 x 8)

Тип шкалы	Количество классификационных шкал	Количество градаций классификационных	Среднее количество градаций на класс шкала	Количество описательных шкал	Количество градаций описательных шкал	Среднее количество градаций на опис. шкалу
Числовые	0	0	0,00	2	6	3,00
Текстовые	1	4	4,00	1	2	2,00
ВСЕГО:	1	4	4,00	3	8	2,67

Задайте число интервалов (градаций) в шкале:

В описательных шкалах:

Пересчитать шкалы и градации      Параметры числ. шкал и градаций      Выйти на создание модели

---

2.3.2.2. Параметры классификационных и описательных шкал и градаций

ПАРАМЕТРЫ ШКАЛ И ГРАДАЦИЙ С АДАПТИВНОЙ УСТАВКОЙ В ПРОЦЕССЕ РАБОТЫ КОЛИЧЕСТВОМ ВЫБРАННОЙ ПО ГРАДАЦИЯМ С КОДЕСНОЙ ОШИБКОЙ ОПРЕДЕЛЕНИЯ ЧИСЛА НЕБЛИЖАЮЩЕЙ КЪ ИНТЕРВАЛУ ГРАДАЦИИ ПРИ ПЕРЕХОДЕ К СЛЕДУЮЩЕЙ ГРАДАЦИИ

КЛАССИФИКАЦИОННАЯ ШКАЛА: код: { 1 }, имен.: "DEFAULT", тип/число градаций в шкале: "Равное число событий в интервалах"/4

- 1 Имен. градаций: 1/4-0c
- 2 Имен. градаций: 2/4-0c\_01
- 3 Имен. градаций: 3/4-0c
- 4 Имен. градаций: 4/4-0c\_01

---

ОПИСАТЕЛЬНАЯ ШКАЛА: код: { 1 }, имен.: "EMPLOYED", тип шкалы/число градаций в шкале: "Равное число событий в интервалах"/2

- 1 Имен. градаций: 1/2-0c
- 2 Имен. градаций: 2/2-0c

---

ОПИСАТЕЛЬНАЯ ШКАЛА: код: { 2 }, имен.: "BANK BALANCE", тип шкалы/число градаций в шкале: "Равное число событий в интервалах"

- 3 Имен. градаций: 1/3-1 0,2400000, 0,738818000000, размер интервала= 0,738818000000, расч./факт. число приближений на градацию: 3167/32
- 4 Имен. градаций: 2/3-1 0,738818000000, 1,277636000000, размер интервала= 0,538818000000, расч./факт. число приближений на градацию: 3167/32
- 5 Имен. градаций: 3/3-1 1,277636000000, 1,816454000000, размер интервала= 0,538818000000, расч./факт. число приближений на градацию: 3167/32

---

ОПИСАТЕЛЬНАЯ ШКАЛА: код: { 3 }, имен.: "ANNUAL SALARY", тип шкалы/число градаций в шкале: "Равное число событий в интервалах"

- 4 Имен. градаций: 1/3-1 0,263440000000, 0,780320000000, размер интервала= 0,516880000000, расч./факт. число приближений на градацию: 3333

---

2.3.2.2. Процесс импорта данных из внешней БД "Inp\_data" в систему "ЭЙДОС-X++"

Стадии исполнения процесса:

- 1/3: Формирование классификационных и описательных шкал и градаций на основе БД "Inp\_data"- Готово
- 2/3: Генерация обучающей выборки и базы событий "EventsKO" на основе внешней БД "Inp\_data"- Готово
- 3/3: Переиндексация всех баз данных нового приложения- Готово

**ПРОЦЕСС ФОРМАЛИЗАЦИИ ПРЕДМЕТНОЙ ОБЛАСТИ ЗАВЕРШЕН УСПЕШНО !!!**

Прогноз времени исполнения:

Начало: 14:16:10      Окончание: 14:16:17

100%

Прошло: 0:00:07      Осталось: 0:00:00

Ok

---

4.2.2.4 (3.7.6) Разделение классов на типичную и нетипичную части

Далее необходимо в режиме 3.5 создать и верифицировать модели: Abs, Prc1, Prc2, Inf1, Inf2, Inf3, Inf4, Inf5, Inf6, Inf7.

Затем в режиме 3.4 необходимо определить модель и интегральный критерий, при которых достигается максимальная достоверность.

Режим 4.2.2.4 (3.7.6) можно повторять до достижения необходимого достаточно высокого уровня достоверности моделей.

Если достоверность модели достаточно высока, то в ней корректно можно решать задачи идентификации и прогнозирования (4.1.2), принятия решений (4.4.8) и исследования объекта моделирования путем исследования его модели (режимы: 4.4.9, 4.4.10, 4.4.11, 4.4.12, 4.2.1, 4.2.2.1, 4.2.2.2, 4.2.3, 4.3.2.1, 4.3.2.2, 4.3.2.3, 4.5, 3.7.5, 3.7.4., 3.7.3, 3.7.9 и т.д.)

Ok

### 3.5. Синтез и верификация моделей

Стадии исполнения процесса:

- Шаг 1-й из 11: Копирование обучающей выборки в распознаваемую - Готово
- Шаг 2-й из 11: Синтез стат модели "ABS" (расчет матрицы абсолютных частот) - Готово
- Шаг 3-й из 11: Синтез стат.моделей "PRC1" и "PRC2" (учл.безусл.% распр.) - Готово
- Шаг 4-й из 11: Синтез системно-когнитивных моделей: INF1-INF7 - Готово
- НАЧАЛО ЦИКЛА ПО ЧАСТНЫМ И ИНТЕГРАЛЬНЫМ КРИТЕРИЯМ - ИСПОЛНЕНИЕ**
- Шаг 5-й из 11: Задание модели "INF7" в качестве текущей - Готово
- Шаг 6-й из 11: Пакетное распознавание в модели "INF7" - Готово
- Шаг 7-й из 11: Измерение достоверности модели: "Inf7" - Интегральный критерий: "Сумма знаний" - Готово
- КОНЕЦ ЦИКЛА ПО ЧАСТНЫМ И ИНТЕГРАЛЬНЫМ КРИТЕРИЯМ - ГОТОВО**
- Шаг 8-й из 11: Объединение БД DostRsp# в БД DostRasp - Готово
- Шаг 9-й из 11: Печать сводной формы по результатам верификации моделей - Готово
- Шаг 10-й из 11: Создание формы: "Достоверность идент.классов в различных моделях" - Готово
- Шаг 11-й из 11: "Присвоение заданной модели: Inf1 статуса текущей" - Готово

**Синтез и верификация статистических и системно-когнитивных моделей успешно завершены !!!**

Прогноз времени исполнения:

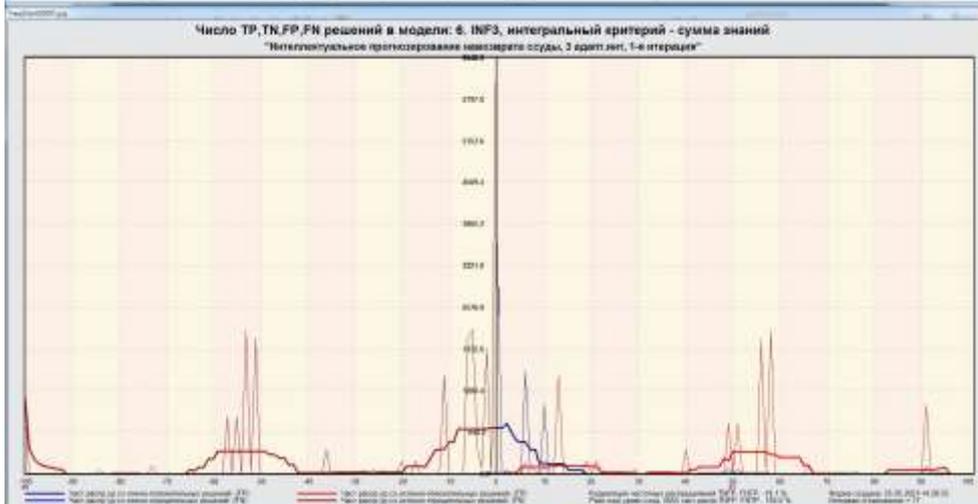
Начало: 14:17:18 Окончание: 14:25:34

100%

Прошло: 0:08:16 Осталось: 0:00:00

3.6. Сводная форма по результатам верификации моделей. Текущая модель: INF7

Идентификационный номер модели	Имя модели	Число признаков	Число классов	Число объектов	Число объектов на класс	Число объектов на класс (сбалансированно)	Число объектов на класс (нормализованно)	Число объектов на класс (нормализованно, сбалансированно)	Число объектов на класс (нормализованно, сбалансированно, сбалансированно)	Число объектов на класс (нормализованно, сбалансированно, сбалансированно, сбалансированно)	Число объектов на класс (нормализованно, сбалансированно, сбалансированно, сбалансированно, сбалансированно)	Число объектов на класс (нормализованно, сбалансированно, сбалансированно, сбалансированно, сбалансированно, сбалансированно)	Число объектов на класс (нормализованно, сбалансированно, сбалансированно, сбалансированно, сбалансированно, сбалансированно, сбалансированно)	Число объектов на класс (нормализованно, сбалансированно, сбалансированно, сбалансированно, сбалансированно, сбалансированно, сбалансированно, сбалансированно)	Число объектов на класс (нормализованно, сбалансированно, сбалансированно, сбалансированно, сбалансированно, сбалансированно, сбалансированно, сбалансированно, сбалансированно)	Число объектов на класс (нормализованно, сбалансированно, сбалансированно, сбалансированно, сбалансированно, сбалансированно, сбалансированно, сбалансированно, сбалансированно, сбалансированно)	Число объектов на класс (нормализованно, сбалансированно, сбалансированно, сбалансированно, сбалансированно, сбалансированно, сбалансированно, сбалансированно, сбалансированно, сбалансированно, сбалансированно)	Число объектов на класс (нормализованно, сбалансированно, сбалансированно)		
1. ABS	матрица абсолютных частот	10000	2	10000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000
2. PRC1	матрица относительных частот	10000	2	10000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000
3. INF1	интегральный критерий	10000	2	10000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000
4. INF2	интегральный критерий	10000	2	10000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000
5. INF3	интегральный критерий	10000	2	10000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000
6. INF4	интегральный критерий	10000	2	10000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000
7. INF5	интегральный критерий	10000	2	10000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000
8. INF6	интегральный критерий	10000	2	10000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000
9. INF7	интегральный критерий	10000	2	10000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000



5.5. Модель: "1. ABS - частный критерий: количество встреч сочетаний "Класс-признак" у объектов обучающей выборки"								
Код признака	Наименование описательной шкалы и градации	1. DEFAULTED? 1/4 NO	2. DEFAULTED? 2/4 NO_FN	3. DEFAULTED? 3/4 YES	4. DEFAULTED? 4/4 YES_FN	Сумма	Среднее	Средн. квадрат. откл.
1	EMPLOYED-1/2 No	1638	1199	126	1	2944	736.00	796.79
2	EMPLOYED-2/2 Yes	8140	1710	195	8	7096	1764.00	2376.00
3	BANK BALANCE, \$-1/3 (0.2400000, 7799.1600000)	3167				3167	791.75	1683.50
4	BANK BALANCE, \$-2/3 (7799.1600000, 12737.0400000)	3158				3167	791.75	1577.51
5	BANK BALANCE, \$-3/3 (12737.0400000, 31851.8400000)		2643	324		3167	791.75	1376.00
6	ANNUAL SALARY, \$-1/3 (9263.6400000, 304886.6400000)	1939	1260	133	1	3333	833.25	828.76
7	ANNUAL SALARY, \$-2/3 (304886.6400000, 487205.0400000)	2379	856	94	4	3333	833.25	1099.09
8	ANNUAL SALARY, \$-3/3 (487205.0400000, 862650.7600000)	2440	793	97	4	3334	833.50	1127.38
	Сумма числа признаков	19841	5661	972	27	29501		
	Среднее	2480	1063	122	3		821.91	
	Среднеквадратичное отклонение	1474	927	106	4			828.91
	Сумма числа объектов обучающей выборки	4758	2908	324	9	10000		

5.5. Модель: "3. PRC2 - частный критерий: условная вероятность i-го признака у объектов j-го класса"								
Код признака	Наименование описательной шкалы и градации	1. DEFAULT ... 1/4 NO	2. DEFAULT ... 2/4 NO_FN	3. DEFAULT ... 3/4 YES	4. DEFAULT ... 4/4 YES_FN	Безусл. вероятн.	Среднее	Средн. квадрат. откл.
1	EMPLOYED-1/2 No	23.942	41.217	38.889	11.111	29.440	28.790	14.053
2	EMPLOYED-2/2 Yes	76.058	58.783	61.111	88.889	70.560	71.210	14.053
3	BANK BALANCE, \$-1/3 (0.2400000, 7799.1600000)	46.863				31.670	11.716	23.481
4	BANK BALANCE, \$-2/3 (7799.1600000, 12737.0400000)	46.730			100.000	31.670	36.682	47.614
5	BANK BALANCE, \$-3/3 (12737.0400000, 31851.8400000)		97.731	100.000		31.670	49.433	57.088
6	ANNUAL SALARY, \$-1/3 (9263.6400000, 304886.6400000)	28.692	43.314	41.049	11.111	33.330	31.042	14.759
7	ANNUAL SALARY, \$-2/3 (304886.6400000, 487205.0400000)	35.203	29.426	29.012	44.444	33.330	34.521	7.194
8	ANNUAL SALARY, \$-3/3 (487205.0400000, 862650.7600000)	36.105	27.260	29.938	44.444	33.340	34.437	7.631
	Сумма	293.593	297.731	300.000	300.000	1191.324		
	Среднее	36.699	37.216	37.500	37.500		37.229	
	Среднеквадратичное отклонение	21.809	31.871	32.558	39.368			30.481

5.5. Модель: "4. INF1 - частный критерий: количество знаний по А.Харкевичу; вероятности из PRC1"								
Код признака	Наименование описательной шкалы и градации	1. DEFAULTED? 1/4 NO	2. DEFAULTED? 2/4 NO_FN	3. DEFAULTED? 3/4 YES	4. DEFAULTED? 4/4 YES_FN	Сумма	Среднее	Средн. квадрат. откл.
1	EMPLOYED-1/2 No	-0.039	0.064	0.091	-0.193	-0.117	-0.029	0.118
2	EMPLOYED-2/2 Yes	0.016	-0.037	-0.031	0.042	-0.011	-0.003	0.038
3	BANK BALANCE, \$-1/3 (0.2400000, 7799.1600000)	0.077				0.077	0.019	0.039
4	BANK BALANCE, \$-2/3 (7799.1600000, 12737.0400000)	0.077			0.220	0.297	0.074	0.104
5	BANK BALANCE, \$-3/3 (12737.0400000, 31851.8400000)		0.237	0.220		0.437	0.109	0.128
6	ANNUAL SALARY, \$-1/3 (9263.6400000, 304886.6400000)	-0.028	0.049	0.037	-0.217	-0.159	-0.040	0.122
7	ANNUAL SALARY, \$-2/3 (304886.6400000, 487205.0400000)	0.012	-0.026	-0.030	0.059	0.008	0.002	0.039
8	ANNUAL SALARY, \$-3/3 (487205.0400000, 862650.7600000)	0.016	-0.041	-0.024	0.053	0.004	0.001	0.042
	Сумма	0.130	0.226	0.223	-0.042	0.538		
	Среднее	0.016	0.028	0.028	-0.005		0.037	
	Среднеквадратичное отклонение	0.043	0.085	0.082	0.141			0.090

5.5. Модель: "6. INF3 - частный критерий: Хи-квадрат: равенности между фактическими и ожидаемыми абсолютными"								
Код признака	Наименование описательной шкалы и градации	1. DEFAULTED? 1/4 NO	2. DEFAULTED? 2/4 NO_FN	3. DEFAULTED? 3/4 YES	4. DEFAULTED? 4/4 YES_FN	Сумма	Среднее	Средн. квадрат. откл.
1	EMPLOYED-1/2 No	-361.997	334.691	29.001	-1.694	0.000	0.000	285.134
2	EMPLOYED-2/2 Yes	394.483	-341.524	-34.481	1.842	0.000	0.000	309.565
3	BANK BALANCE, \$-1/3 (0.2400000, 7799.1600000)	1037.023	-929.778	-104.346	-2.899	0.000	0.000	806.391
4	BANK BALANCE, \$-2/3 (7799.1600000, 12737.0400000)	1029.023	-929.778	-104.346	6.101	0.000	0.000	802.547
5	BANK BALANCE, \$-3/3 (12737.0400000, 31851.8400000)	-2129.977	1913.222	219.654	-2.899	0.000	0.000	1657.358
6	ANNUAL SALARY, \$-1/3 (9263.6400000, 304886.6400000)	-302.621	291.497	23.184	-2.050	0.000	0.000	238.995
7	ANNUAL SALARY, \$-2/3 (304886.6400000, 487205.0400000)	137.379	-122.513	-15.216	0.950	0.000	0.000	104.667
8	ANNUAL SALARY, \$-3/3 (487205.0400000, 862650.7600000)	197.707	-185.807	-12.849	0.949	0.000	0.000	156.820
	Сумма	0.000	0.000	0.000	0.000	0.000	0.000	
	Среднее	0.000	0.000	0.000	0.000		0.000	
	Среднеквадратичное отклонение	1007.489	907.184	102.125	3.054			646.060

Рисунок 17. Экранные формы режима разделения классов на типичную и нетипичную части 1-й итерации

В данном интеллектуальном приложении по F-критерию Ван Ризбергена наиболее достоверной является СК-модель INF3 с интегральным критерием «Сумма знаний».

Из сравнения рисунков 11 и 17 мы видим, что в результате 1-й итерации достоверность этой модели :

– по критерию Ван Ризбергена выросла на 5,3%: с 0,708 до 0,746;

– по критерию L1 проф.Е.В.Луценко [24] выросла на **61,5%**: с 0,592 до 0,956, что вообще является отличным показателем.

Частотное распределение истинных и ложных положительных и отрицательных решений в зависимости от уровня сходства также стало гораздо более разумным, т.к. мы видим, что отмеченная аномалия исчезла.

Видно, что для отрицательных решений ложные отрицательные решения вообще отсутствуют. Это и есть результат работы режима разделения классов на типичную и нетипичную части, реализующего предлагаемый в данной работе интеллектуальный алгоритм (рисунок 3).

Из рисунка 17 видно также, что для положительных решений картина более сложная и включает 3 диапазона уровней сходства

1) при уровнях сходства от 0% до 10% количество ложных решений больше числа истинных;

2) при уровнях сходства от 10% до примерно 20% есть и истинные и ложные положительные решения, но число истинных решений больше числа ложных и доля истинных решений возрастает при увеличении уровня сходства;

3) при уровнях сходства выше 20% ложные положительные решения практически не встречаются.

На рисунке 18 приведен Help режима 3.4, в котором описаны меры достоверности моделей, применяемые в системе «Эйдос»/

Также мы видим, что процесс синтеза и верификации моделей продолжался заметно дольше: если исходные модели формировались и верифицировались 5 минут 34 секунды, о модели 1-й итерации уже 8 минут 16 секунд. Это связано с увеличением количества классов в два раза в связи с добавлением классов для нетипичных случаев (таблица 6).

Из рисунка 15 мы видим, что из 10000 наблюдений обучающей выборки довольно значительная часть: 2918 оказались нетипичными, т.е. не были отнесены к классам, к которым они в действительности относятся. Для этих наблюдений и были созданы собственные классы с теми же наименованиями, но с добавлением символов: \_FN (это обозначение ложно-отрицательных решений по Ван Ризбергену) (таблица 11).

Таблица 11 – Классификационные шкалы и градации 1-й итерации

Код	Наименование
1	DEFAULTED?-1/4-No
2	DEFAULTED?-2/4-No_FN
3	DEFAULTED?-3/4-Yes
4	DEFAULTED?-4/4-Yes_FN

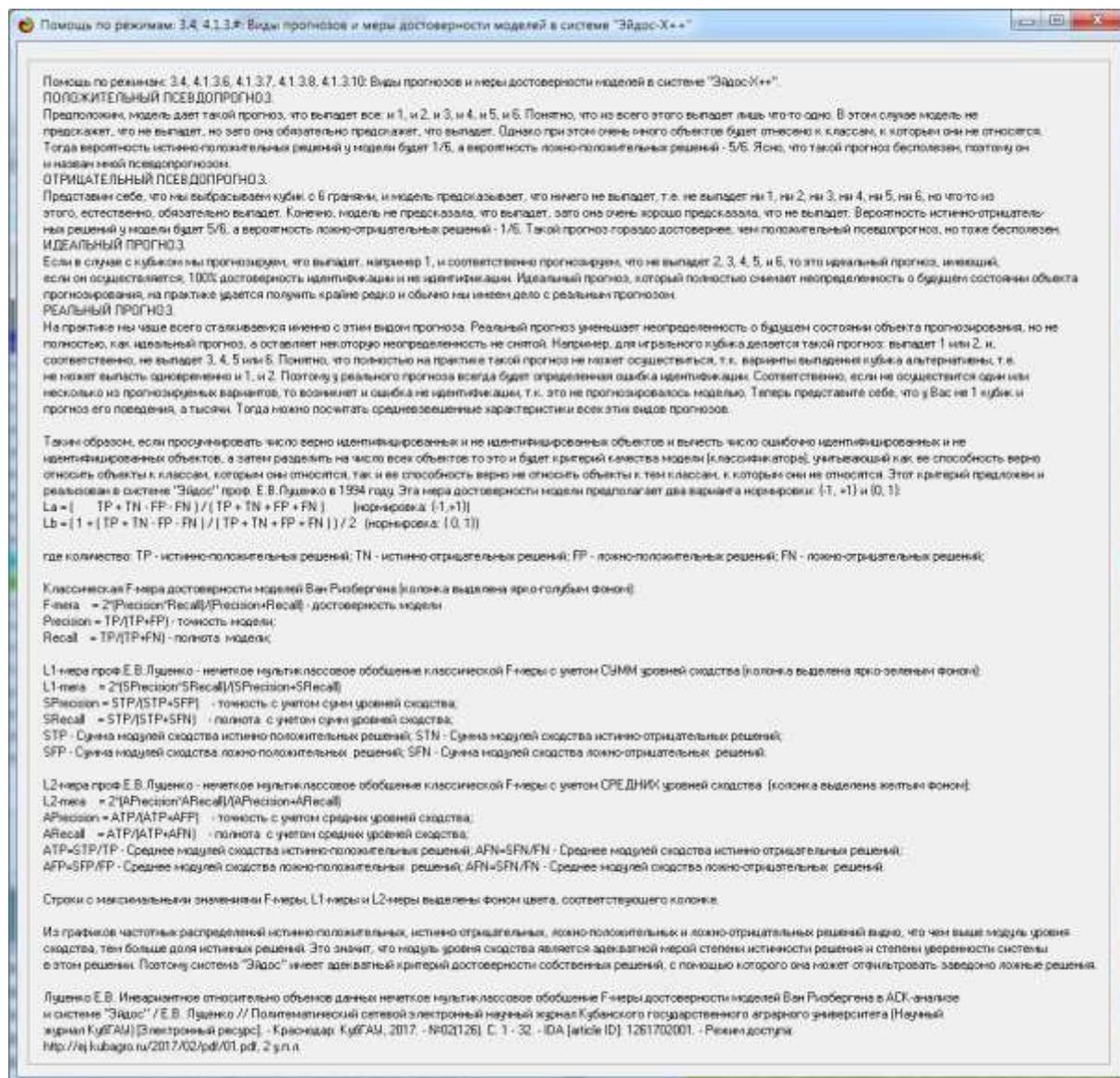


Рисунок 18. Help режима 3.4, в котором описаны меры достоверности моделей, применяемые в системе «Эйдос»

Важно иметь в виду, что корректируется (путем разделения классов на типичную и нетипичную части) всегда та модель, в которой проводилось распознавание. Если данный режим 3.7.6 запускается сразу после режима синтеза и верификации моделей 3.5 – то это будет СК-модель INF7, как последняя, в которой проводилась верификация в режиме 3.5. Если мы в режиме 3.4 выбрали другую модель для улучшения системы как наиболее достоверную, то надо ее задать в качестве текущей в режиме 5.6, а затем выполнить **распознавание** в этой модели в режиме 4.1.2.

Поэтому далее выполняем режим 5.6 (рисунки 2 и 19), а затем режим **распознавания** 4.1.2 в текущей модели INF3 (рисунок 20):

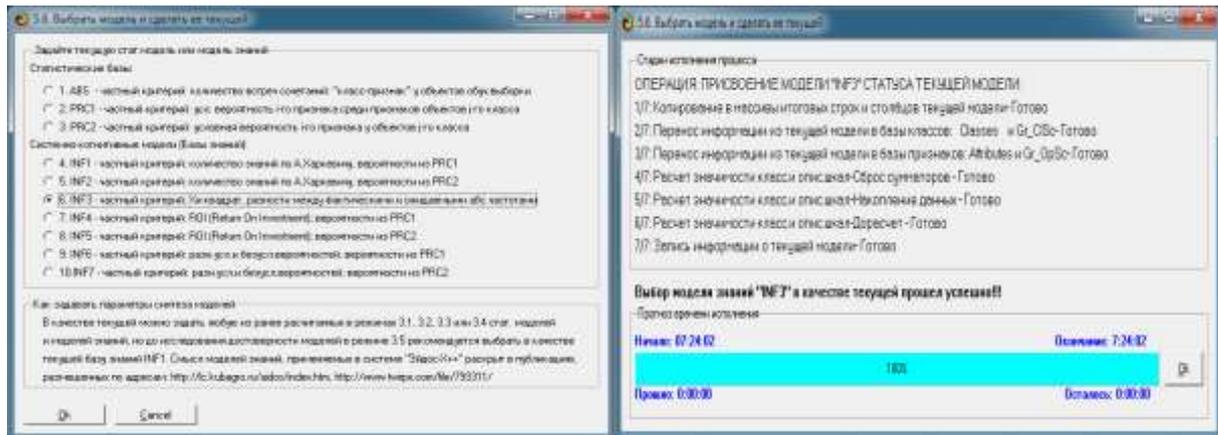


Рисунок 19. Экранные формы придания наиболее достоверной СК-модели Inf3 статуса текущей модели

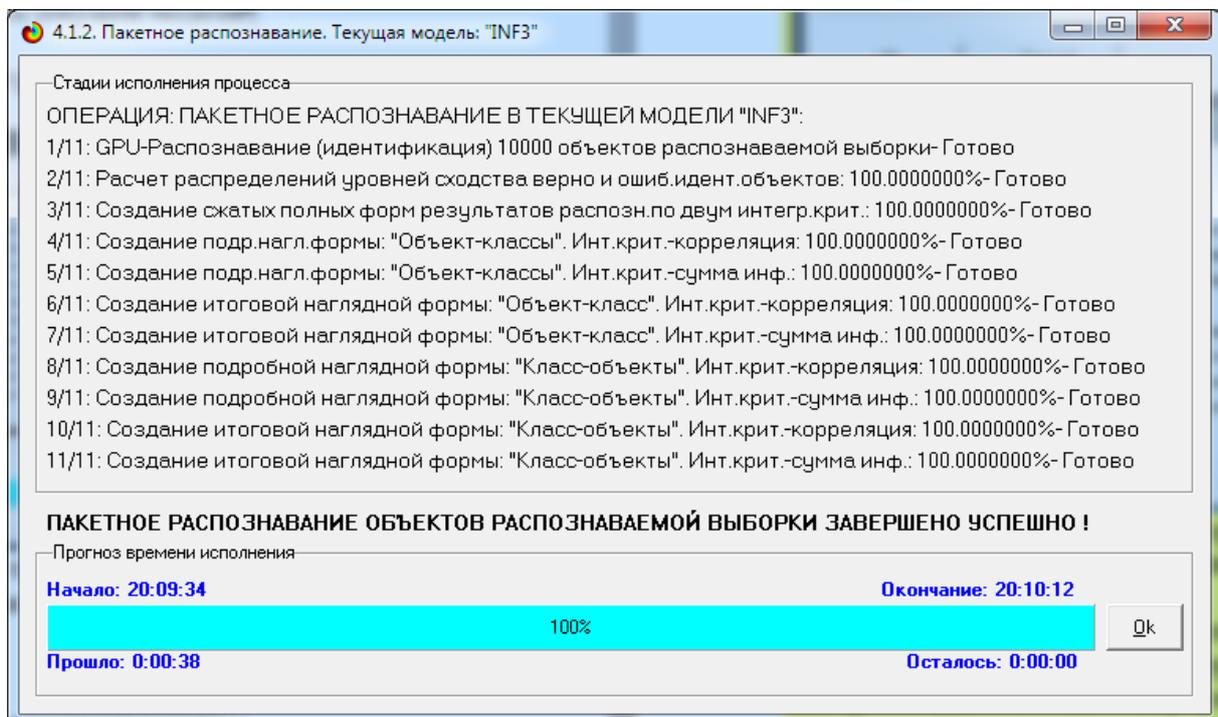


Рисунок 20. Экранная форма пакетного распознавания в текущей СК-модели Inf3

Из сравнения рисунков 14 и 20 видно, что на 1-й итерации распознавание 10000 объектов распознаваемой выборки длилось уже не 8, а 38 секунд. Это связано с увеличением числа классов с двух до 4 за счет создания классов для нетипичных случаев.

**4.3.3.3. Вторая итерация. Синтез и верификация статистических и системно-когнитивных моделей. Оценка достоверности моделей. Задание текущей модели и решение в ней всех поставленных задач**

Для 2-й итерации запускаем еще раз режим 3.7.6 (рисунок 21).

В результате работы данного режима мы обнаруживаем, что новых нетипичных объектов в обучающей выборки не найдено и, соответственно, новых классов не создано. Поэтому модели 2-й итерация ничем не отличается от моделей 1-й итерации, которая и является последней. То, что так и получится, было видно и до 2-й итерации уже из формы по достоверности моделей 1-й итерации, приведенной на рисунке 17. Из этой формы видно, что в 1-й итерации уже нет ложно-отрицательных решений.

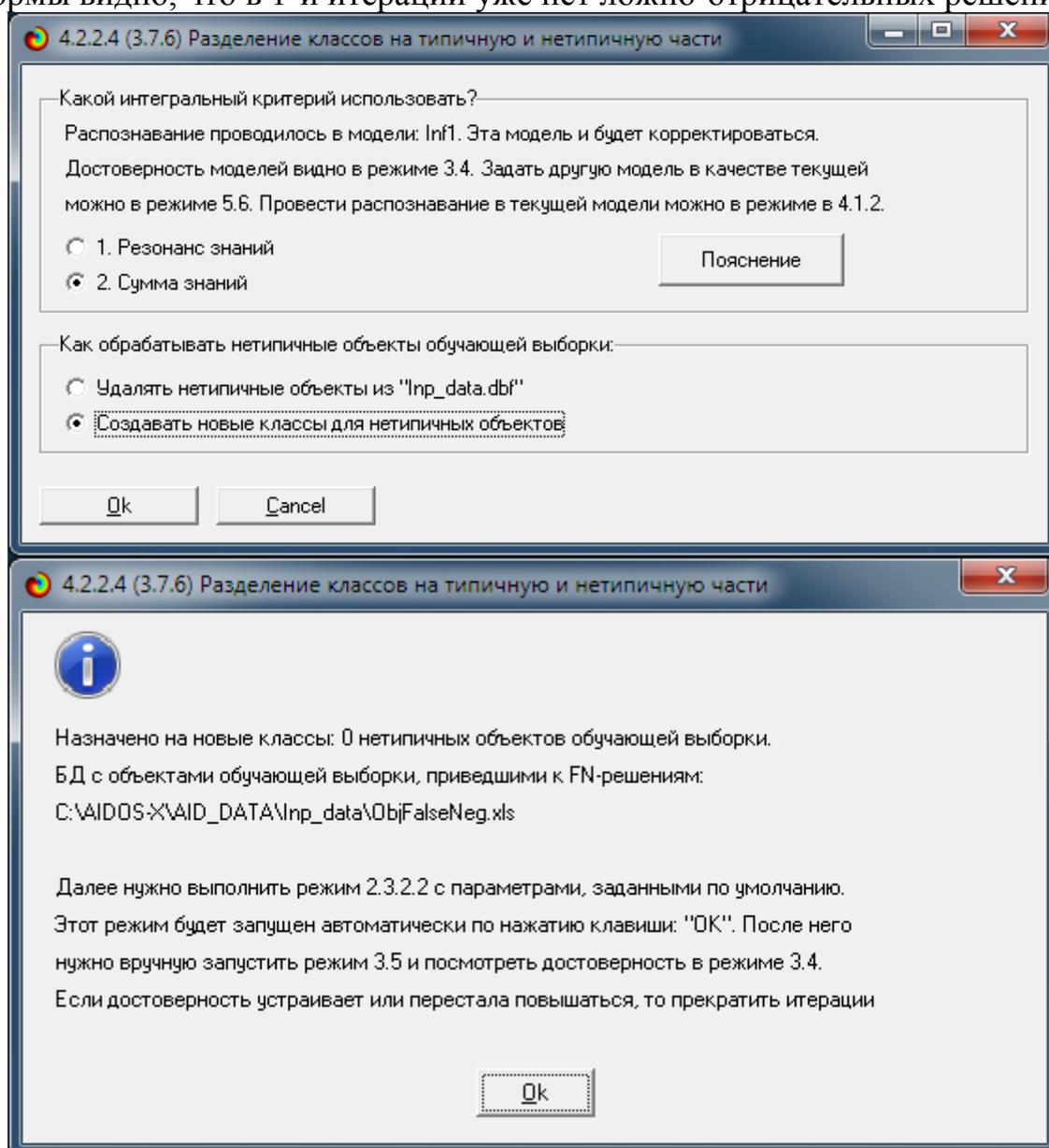


Рисунок 21. Экранные формы режима разделения классов на типичную и нетипичную части 2-й итерации

Поэтому на этом этапе выходим из интеллектуального алгоритма разделения классов на типичную и нетипичную части (рисунок 3), создаем все модели в режиме 3.5, выбираем наиболее достоверную модель INF3 в качестве текущей в режиме 5.6 и решаем в ней все последующие задачи. Этому и посвящены последующие разделы данной работы.

#### 4.3.4. Задача 4: решение различных задач в наиболее достоверной модели

##### 4.3.4.1. Подзадача 4.1. Прогнозирование (диагностика, классификация, распознавание, идентификация)

Решим задачу системной идентификации 10000 ссудополучателей обучающей выборки по их характеристикам, т.е. задачу прогнозирования риска невозврата ссуды. Эту задачу решим в наиболее достоверной СК-модели INF3 на графическом процессоре (GPU) (рисунок 22).

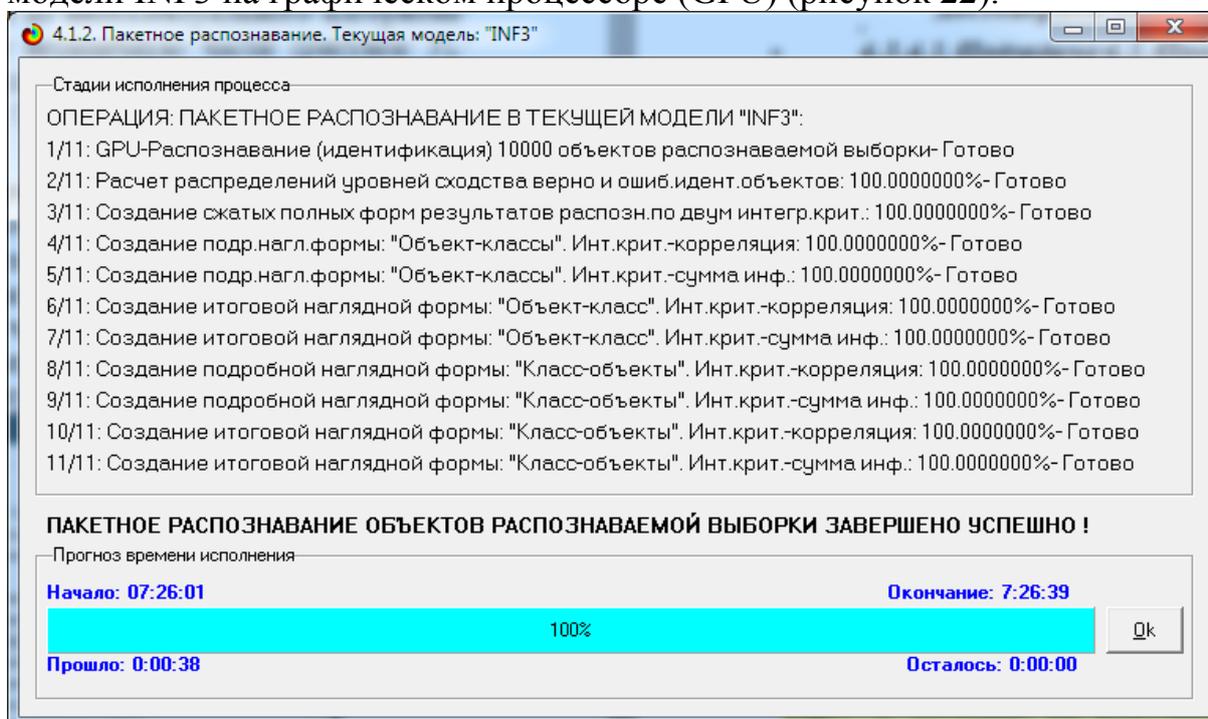


Рисунок 22. Экранные формы, которые отображают процесс решения задачи системной идентификации в текущей модели

Из рисунка 22 видно, что процесс идентификации занял 38 секунд.

Отметим, что 99% этого времени заняла не сама идентификация на GPU, а создание 10 выходных форм на основе результатов этого прогнозирования. Эти формы отражают результаты прогнозирования в различных разрезах и обобщениях:

Приведем две из этих 10 форм: 4.1.3.1 и 4.1.3.2 (рисунок 23).

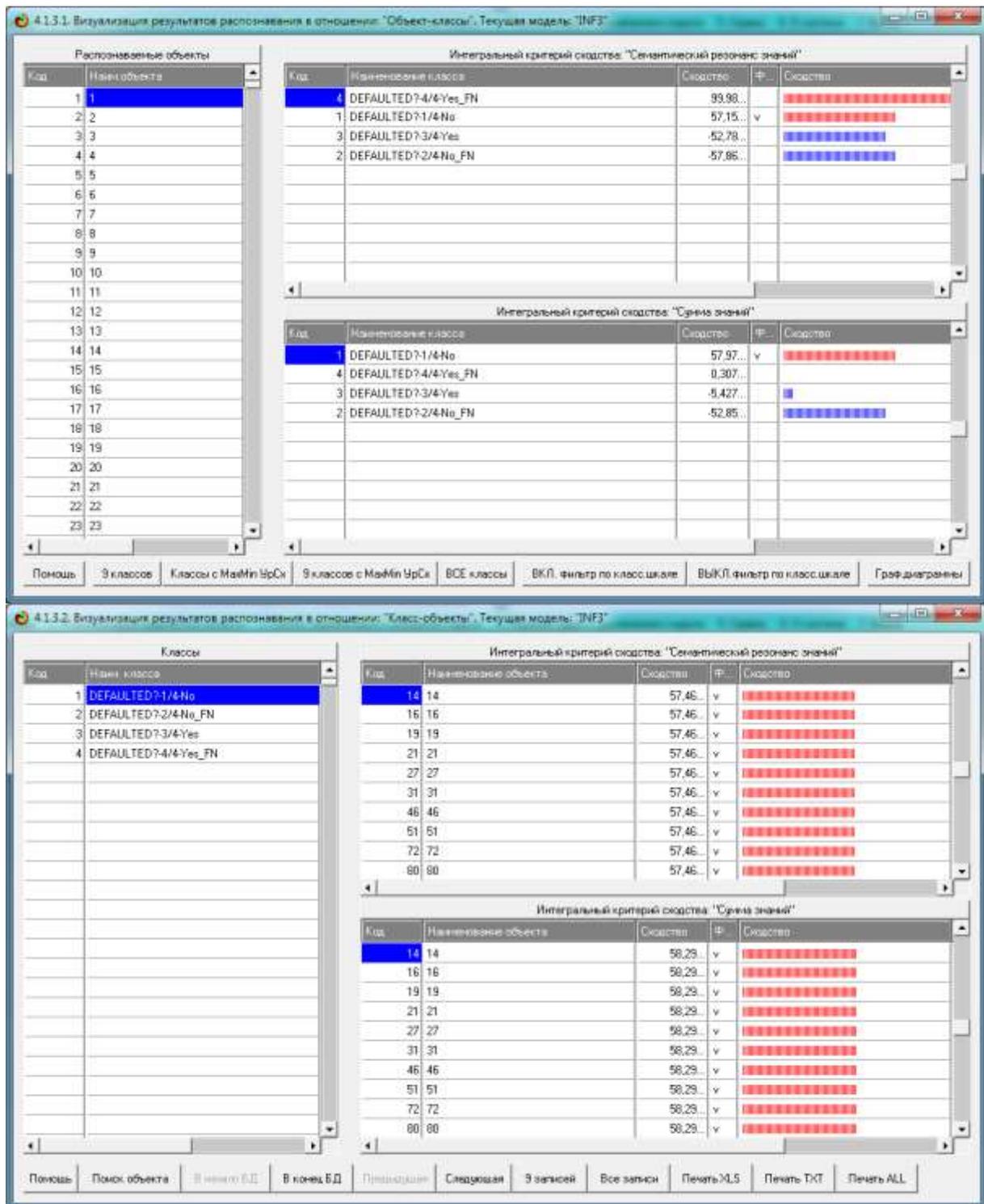


Рисунок 23. Выходные формы по результатам прогнозирования риска невозврата ссуды по характеристикам получателей

Символ «√» стоит против тех результатов идентификации, которые подтвердились на опыте, т.е. соответствуют факту. Из рисунка 23 видно, что результаты идентификации являются отличными, естественно при учете информации из рисунка 17 о том, что *достоверные прогнозы в данной модели имеют уровень сходства выше 10% по интегральному*

критерию «Сумма знаний» (нижнее правое окно в экранных формах на рисунке 19), т.е., по сути, результаты с более низки уровнем сходства надо просто игнорировать.

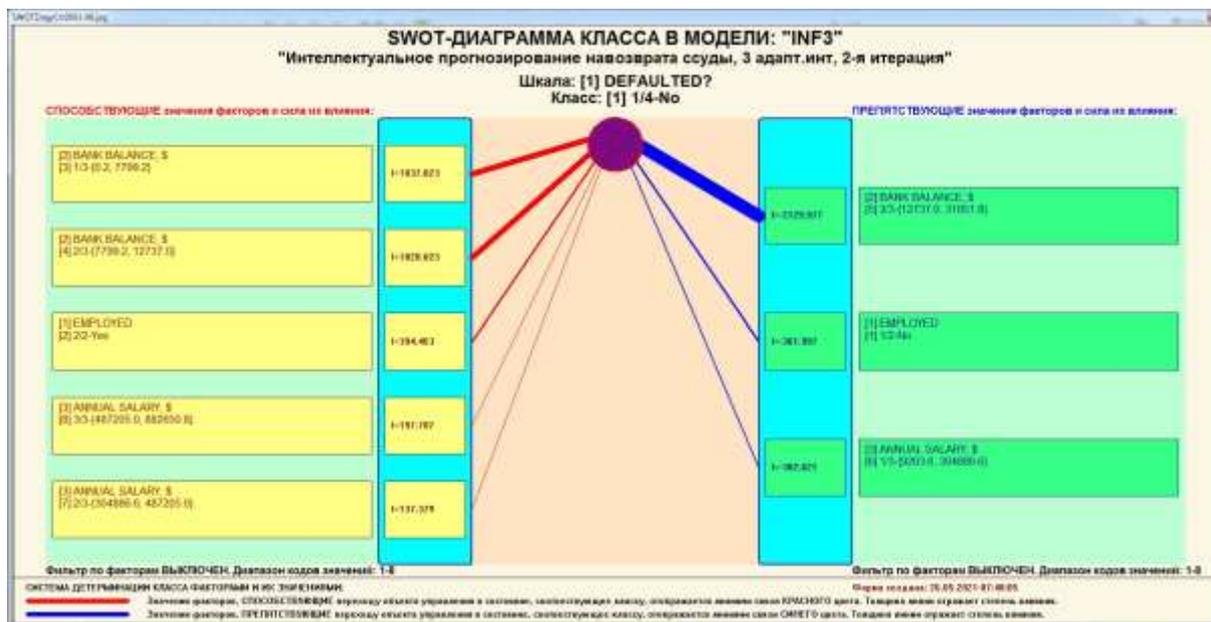
#### 4.3.4.2. Подзадача 4.2. Поддержка принятия решений в простейшем варианте (SWOT-анализ)

При принятии решений определяется сила и направление влияния значений факторов на принадлежность состояний объекта моделирования к тем или иным классам, соответствующим различным будущим состояниям. В простейшем варианте принятие решений это, по сути, решение задачи SWOT-анализа [25].

Применительно к задаче, решаемой в данной работе, SWOT-анализ показывает степень влияния различных значений характеристик ссудополучателей на риск невозврата полученной ими ссуды.

В системе «Эйдос» в режиме 4.4.8 поддерживается решение этой задачи. При этом **выявляется система детерминации заданного класса**, т.е. система значений факторов, обуславливающих переход объекта моделирования и управления в состояние, соответствующее данному классу, а также препятствующих этому переходу.

На рисунках 24 приведены SWOT-диаграммы наглядно отражающие силу и направление влияния различных значений характеристик ссудополучателей на риск невозврата полученной ими ссуды. Экранные формы, приведенные на рисунках 24, содержат все необходимые пояснения и интуитивно понятны.



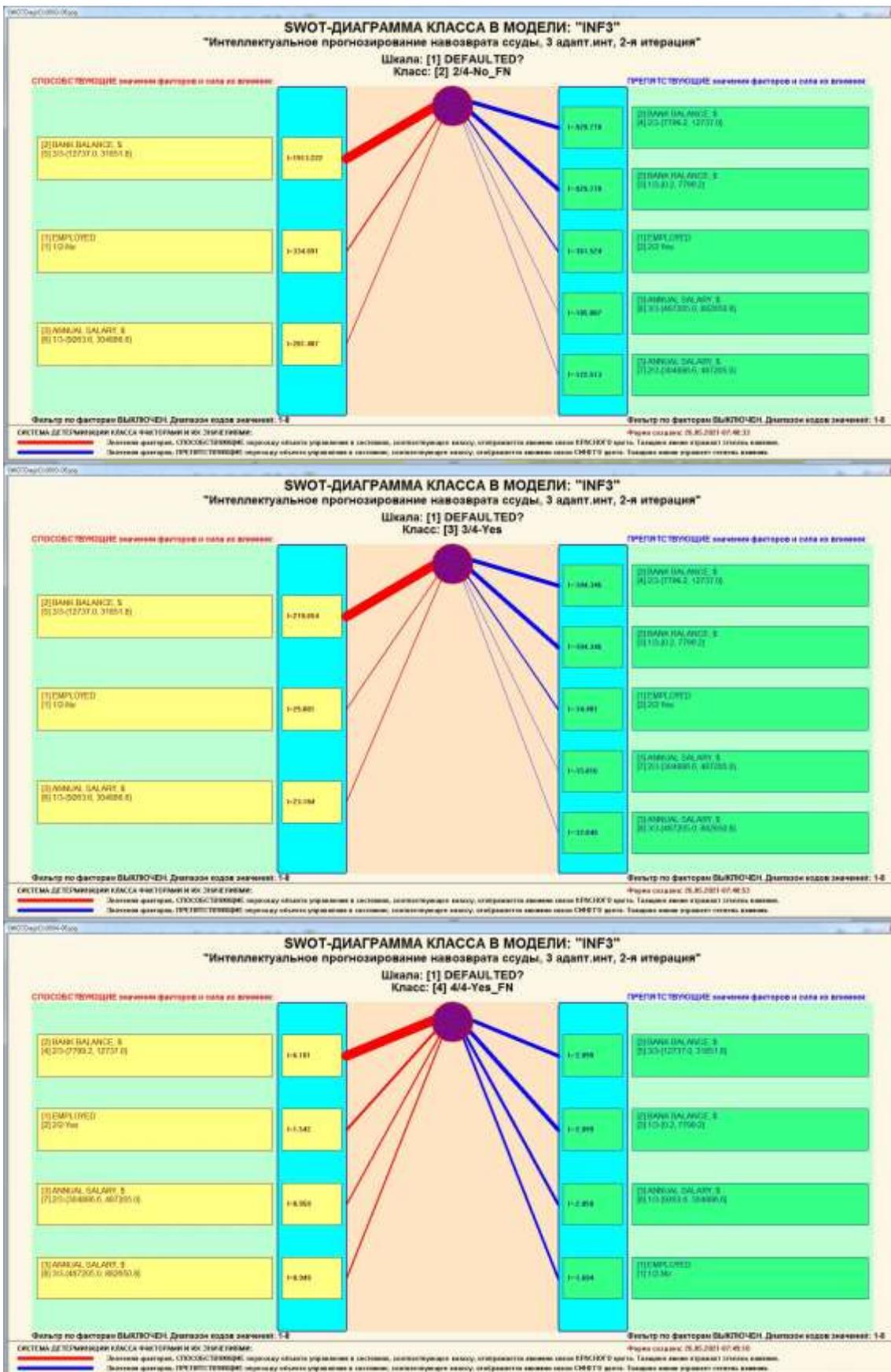


Рисунок 24. SWOT-диаграммы детерминации риска невозврата ссуды характеристиками получателей

Эти SWOT-диаграммы наглядно отражают силу и направление влияния различных значений характеристик ссудополучателей на риск невозврата полученной ими ссуды.

Отметим также, что система «Эйдос» обеспечивала решение этой задачи *всегда*, т.е. даже в самых ранних DOS-версиях и в реализациях системы «Эйдос» на других языках и типах компьютеров. Например, первый акт внедрения системы «Эйдос», где об этом упоминается в явном виде, датируется 1987 годом, а первый подобный расчет относится к 1981 году. Но тогда SWOT-диаграммы назывались позитивным и негативным информационными портретами классов.

У Т В Е Р Ж Д А Ю

Заведующий Краснодарским сектором ИСИ АН СССР, к.ф.н. А. А. Хагуров 1987г.

У Т В Е Р Ж Д А Ю

Директор Северо-Кавказского филиала ВНИЦ "АИУС-агроресурсы", к.э.н. Э. М. Трахов 1987г.

А К Т

Настоящий акт составлен комиссией в составе: Кириченко М.М., Ляшко Г.А., Самсонов Г.А., Коренец В.И., Луценко Е.В. в том, что в соответствии с договором о научно-техническом сотрудничестве между Северо-Кавказским филиалом ВНИЦ "АИУС-агроресурсы" и Краснодарским сектором Института социологических исследований АН СССР Северо-Кавказским филиалом ВНИЦ "АИУС-агроресурсы" выполнены следующие работы:

- осуществлена постановка задачи: "Обработка на ЭВМ социологических анкет Крайагропрома";
- разработаны математическая модель и программное обеспечение подсистемы распознавания образов, позволяющие решать данную задачу в среде персональной технологической системы ВЕГА-М;
- на профессиональной персональной ЭВМ "Искра-226" осуществлены расчеты по задаче в объеме:

Входная информация составила 425 анкет по 9-ти предприятиям.  
Выходная информация - 4 вида выходных форм объемом 90 листов формата А3 и 20 листов формата А4 содержит:

- процентное распределение ответов в разрезе по социальным типам корреспондентов;
- распределение информативностей признаков (в битах) для распознавания социальных типов корреспондентов;
- позитивные и негативные информационные портреты 30-ти социальных типов на языке 212 признаков;
- обобщенная характеристика информативности признаков для выбора такого минимального набора признаков, который содержит максимум информации о распознаваемых объектах (оптимизация анкет).

Работы выполнены на высоком научно-методическом уровне и в срок.

От ИСИ АН СССР:

Мл. научный сотрудник

Кириченко М.М. Кириченко  
19.05. 1987г.

Мл. научный сотрудник

Ляшко Г.А. Ляшко  
19.05. 1987г.

От СКФ ВНИЦ "АИУС-агроресурсы":

Зав. отделом аэрокосмических и тематических изысканий №4, к.э.н. Самсонов Г.А. Самсонов  
19.05. 1987г.

Главный конструктор проекта Коренец В.И. Коренец  
19.05.87. 1987г.

Главный конструктор проекта Луценко Е.В. Луценко  
19.05.87. 1987г.

Информация о системе значений факторов, обуславливающих переход объекта моделирования в различные будущие состояния, соответствующие классам, может быть приведена не только в диаграммах, показанных на рисунках 20, но и во многих других табличных и графических выходных формах, которые в данной работе не приводятся только из-за ограничений на ее объем. В частности в этих формах может быть выведена значительно более полная информация (в т. ч. вообще вся имеющая в модели). Подобная подробная информация содержится в базах данных, расположенных по пути: \Aidos-X\AID\_DATA\A0000001\System\SWOTcls####Inf3.DBF, где: «####» – код класса с ведущими нулями. Эти базы открываются в MS Excel.

На рисунке 25 приведены примеры нескольких инвертированных SWOT-диаграмм (предложены автором [25]), отражающих силу и направление влияния различных характеристик ссудополучателя на риск невозврата полученной им ссуды.

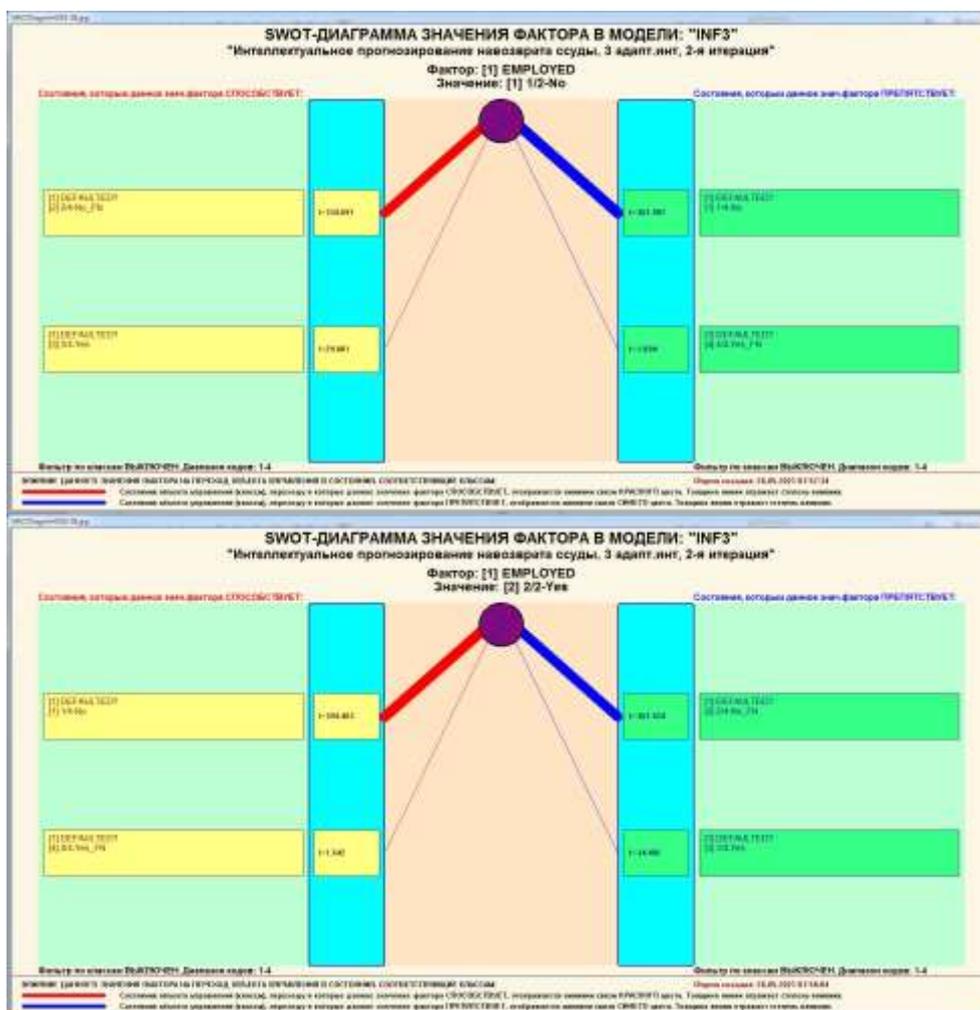


Рисунок 25. Примеры SWOT-диаграмм, отражающих силу и направление влияния различных характеристик ссудополучателя на риск невозврата полученной им ссуды

Из инвертированных SWOT-диаграмм, приведенных на рисунке 21, видно, как влияет безработный ссудополучатель или нет на риск невозврата полученной им ссуды.

Отметим, что аналогичные инвертированные SWOT-диаграммы могут быть получены для всех характеристик ссудополучателей и здесь они не приводятся только из-за ограничений на объем работы. Но они могут быть получены любым желающим, если он скачает систему «Эйдос» с сайта ее автора и разработчика проф.Е.В.Луценко по ссылке: <http://lc.kubagro.ru/aidos/Aidos-X.htm>, установит ее на своем компьютере, а затем в режиме 1.3 установит интеллектуальное облачное Эйдос-приложение №293, просчитает модели в режиме 3.5 и перейдет в режим 4.4.9.

В заключение отметим, что SWOT-анализ является широко известным и общепризнанным методом стратегического планирования. Однако это не мешает тому, что он подвергается критике, часто вполне справедливой, обоснованной и хорошо аргументированной. В результате критического рассмотрения SWOT-анализа выявлено довольно много его слабых сторон (недостатков), источником которых чаще всего является необходимость привлечения экспертов, в частности для оценки силы и направления влияния факторов. Ясно, что эксперты это делают неформализуемым путем (интуитивно), на основе своего опыта и профессиональной компетенции. Но возможности экспертов имеют свои ограничения и часто по различным причинам они не могут и не хотят это сделать. Таким образом, возникает проблема проведения SWOT-анализа без привлечения экспертов. Эта проблема может решаться путем автоматизации функций экспертов, т.е. путем измерения силы и направления влияния факторов непосредственно на основе эмпирических данных. Подобная технология разработана давно, ей уже более 30 лет, но, к сожалению, она сравнительно малоизвестна – это интеллектуальная система «Эйдос».

#### **4.3.4.3. Подзадача 4.2. Развитый алгоритм принятия решений**

В предыдущем разделе кратко описан лишь очень упрощенный вариант принятия решений. В развитой форме, полностью поддерживаемой системой «Эйдос», он приведен в данном разделе (рисунок 26).

Подробное пояснение данного алгоритма (который в принципе и так вполне понятен) не входит в задачи данной работы и дано в других работах автора, например [20], а также в видеозанятиях:

– в Пермском национальном университете:

<https://bigbluebutton.pstu.ru/b/w3y-2ir-ukd-bqn>

– в Кубанском государственном университете и Кубанском государственном аграрном университете:

<https://disk.yandex.ru/d/knISAD5qzV83Ng?w=1>;

YouTube: [https://studio.youtube.com/channel/UC\\_QF84d8SCaWxsnXnexNFzg](https://studio.youtube.com/channel/UC_QF84d8SCaWxsnXnexNFzg)

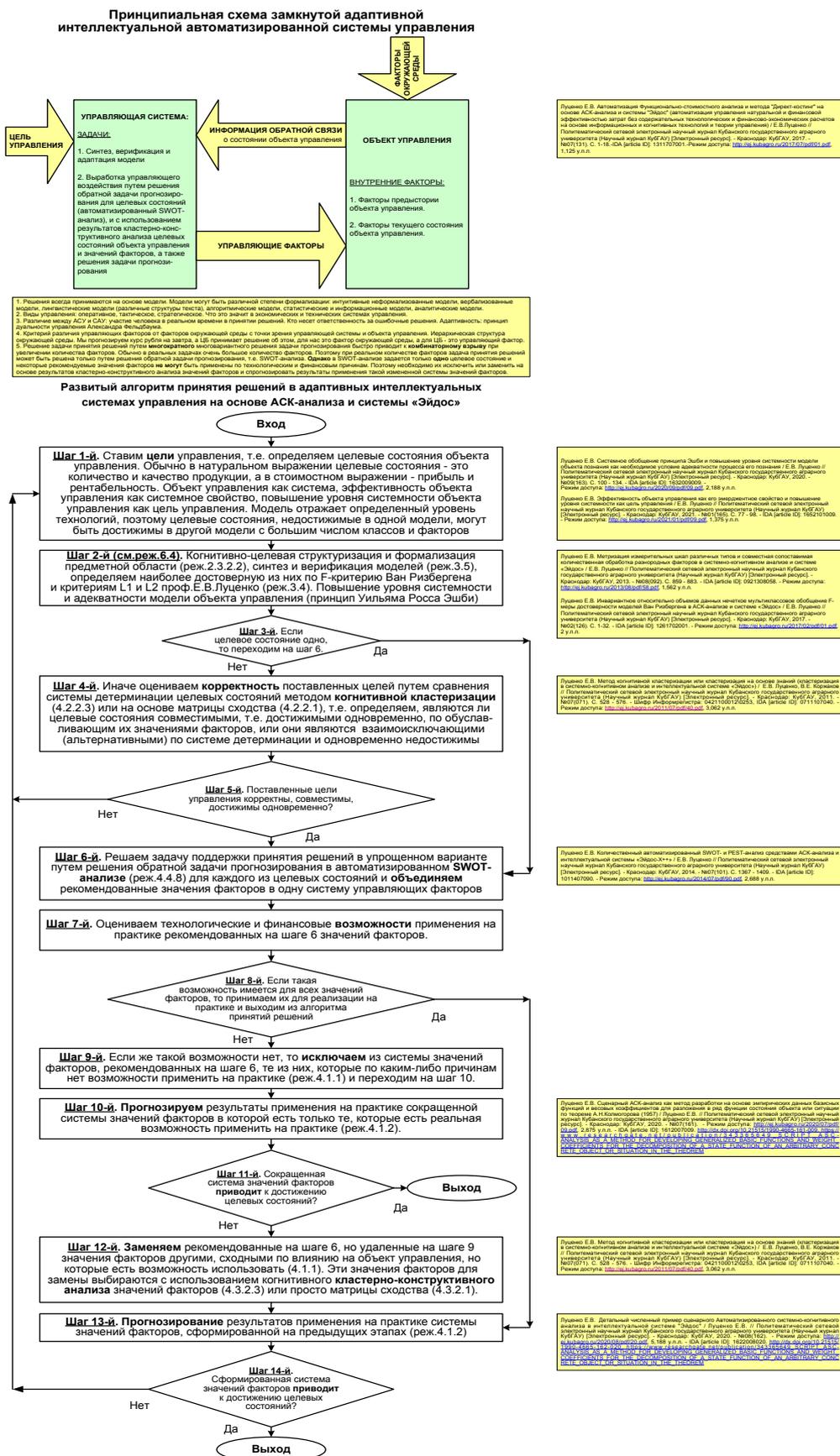


Рисунок 26. Развитый алгоритм принятия решений в интеллектуальных системах управления на основе АСК-анализа и системы «Эйдос»

#### 4.3.4.4. Подзадача 4.3. Исследование моделируемой предметной области путем исследования ее модели

Если модель предметной области достоверна, то исследование модели можно считать исследованием самого моделируемого объекта, т.е. результаты исследования модели корректно относить к самому объекту моделирования, «переносить на него».

В системе «Эйдос» есть довольно много возможностей для такого исследования, но в данной работе из-за ограничений на ее объем мы рассмотрим лишь некоторые из них: когнитивные диаграммы классов и значений факторов, агломеративная когнитивная кластеризация классов и значений факторов, нелокальные нейроны и нейронные сети, 3d-интегральные когнитивные карты, когнитивные функции), исследование силы и направления влияния факторов и степени детерминированности классов, обуславливающими их значениями факторов..

##### 4.3.4.4.1. Когнитивные диаграммы классов

Эти диаграммы отражают сходство/различие классов. Мы получаем их в режимах 4.2.2.1 и 4.2.2.2 (рисунок 27).

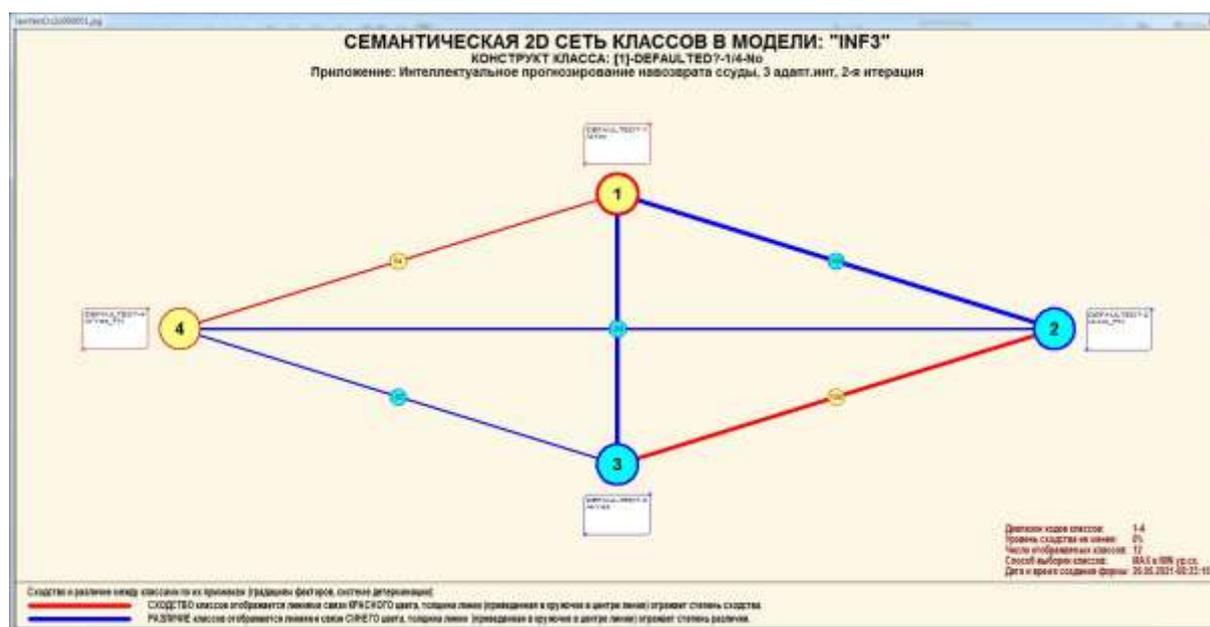


Рисунок 27. Когнитивная диаграмма классов, отражающая сходство/различие рисков невозврата ссуды по связанным с ними значениям характеристик ссудополучателей

Отметим также, что на когнитивной диаграмме, приведенной на рисунке 23, показаны **количественные** оценки сходства/различия рисков невозврата ссуды по связанным с ними значениям характеристик ссудополучателей. Важно, что эти результаты сравнения получены с применением системно-когнитивной модели, созданной *непосредственно на основе эмпирических данных*, а не как традиционно делается на основе

экспертных оценок неформализуемым путем на основе опыта, интуиции и профессиональной компетенции.

В системе «Эйдос» есть возможность при необходимости управлять параметрами формирования и вывода изображения, приведенного на рисунке 27. Для этого используется диалоговое окно, приведенное на рисунке 28.

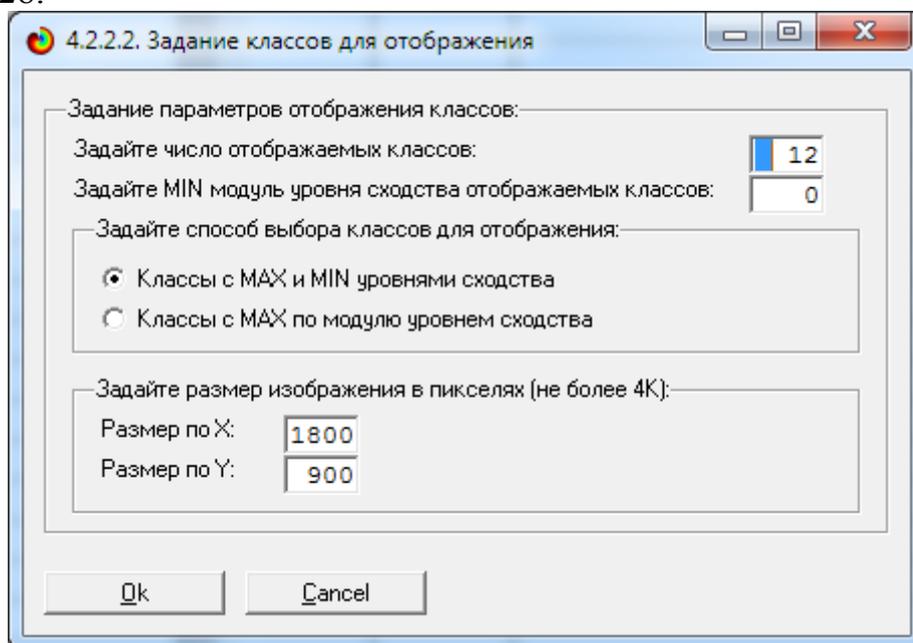


Рисунок 28. Диалоговое окно управления параметрами формирования и вывода изображения когнитивной диаграммы классов

#### 4.3.4.4.2. Агломеративная когнитивная кластеризация классов

Информация о сходстве/различии классов, содержащаяся в матрице сходства, может быть визуализирована не только в форме, когнитивных диаграмм, пример которой приведен на рисунке 27, но и в форме агломеративных дендрограмм, полученных в результате *когнитивной кластеризации* (рисунок 29) [26]. На рисунке 30 мы видим график изменения межкластерных расстояний:

Из рисунка 29 мы видим, что определенные классы, соответствующие рискам невозврата ссуды, сходны по детерминирующей их системе характеристик ссудополучателей, а другие сильно отличаются.

Из дендрограммы когнитивной агломеративной кластеризации классов, приведенной на рисунке 29, мы видим также, что все классы образуют два противоположных кластера, являющихся полюсами конструктора, по системе значений обуславливающих их характеристик.

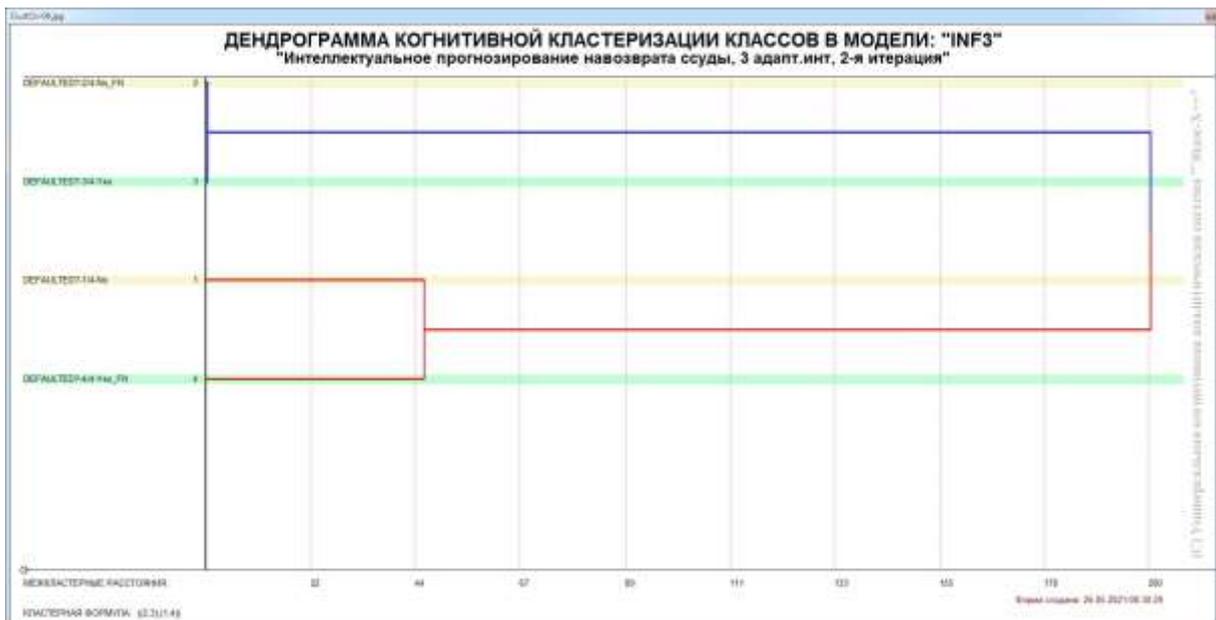


Рисунок 29. Дендрограмма когнитивной агломеративной кластеризации, отражающая сходство/различие рисков невозврата ссуды по системе их детерминации характеристиками ссудополучателей

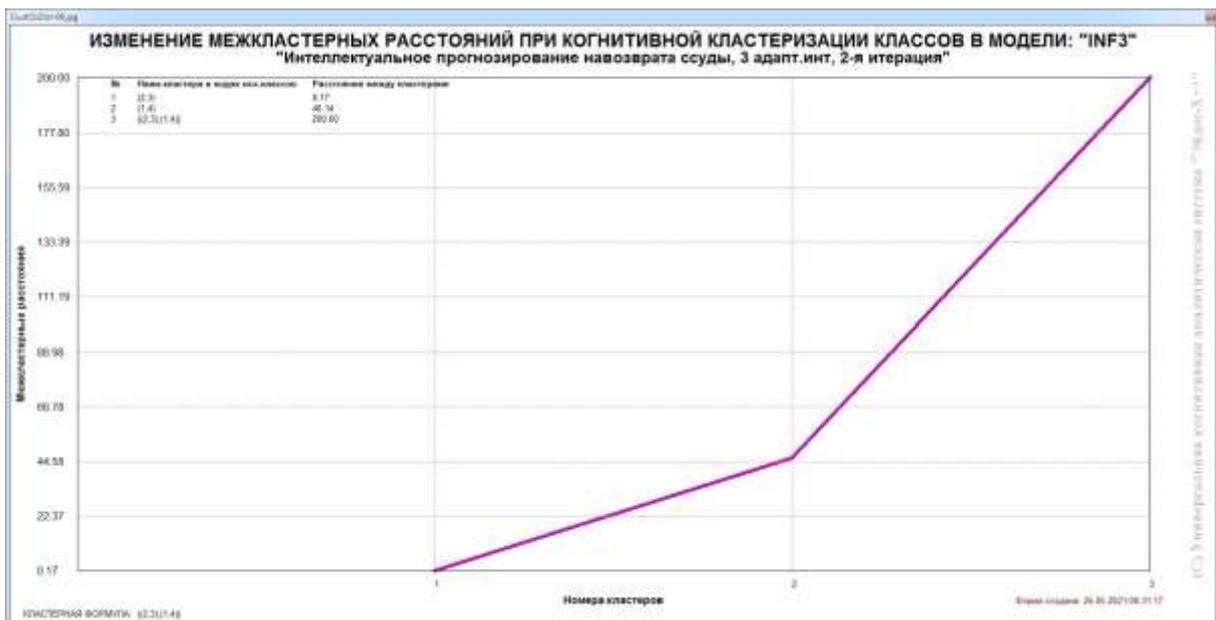


Рисунок 30. График изменения межкластерных расстояний

#### 4.3.4.4.3. Когнитивные диаграммы значений факторов

Эти диаграммы отражают сходство/различие значений характеристик ссудополучателей по их смыслу, т.е. по содержащейся в них информации о риске невозврата ссуды.

Эти диаграммы мы получаем в режимах 4.3.2.1 и 4.3.2.2 (рисунок 31).

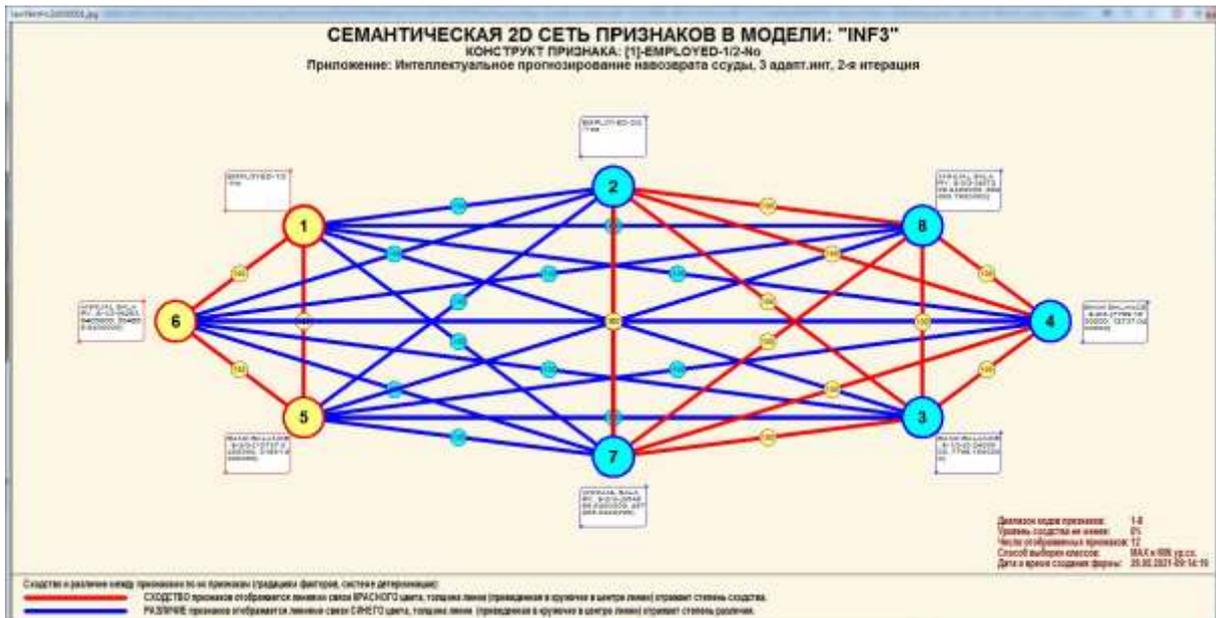


Рисунок 31. Сходство/различие характеристик ссудополучателей по их влиянию на риск невозврата ссуды

Из рисунка 31 видно, что все значения факторов образуют два крупных кластера, противоположных по их смыслу. Эти кластеры образуют полюса конструкта.

Отметим, что на когнитивной диаграмме, приведенной на рисунке 31, показаны **количественные** оценки сходства/различия значений факторов, полученные с применением системно-когнитивной модели, созданной *непосредственно на основе эмпирических данных*, а не как традиционно делается на основе экспертных оценок неформализуемым путем на основе опыта, интуиции и профессиональной компетенции.

Диаграмма, приведенная на рисунке 31, получена при параметрах, приведенных на рисунке 32.

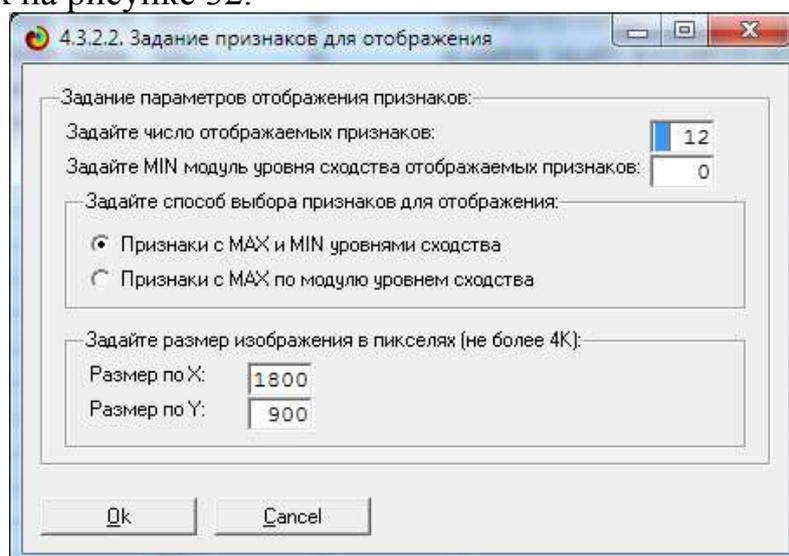


Рисунок 32. Параметры отображения когнитивной диаграммы, приведенной на рисунке 27

#### 4.3.4.4.4. Агломеративная когнитивная кластеризация значений факторов

На рисунке 33 приведена агломеративная дендрограмма когнитивной кластеризации значений факторов [26], полученная на основе той же матрицы сходства признаков по их смыслу, что и в когнитивных диаграммах, пример которой приведен на рисунке 31.

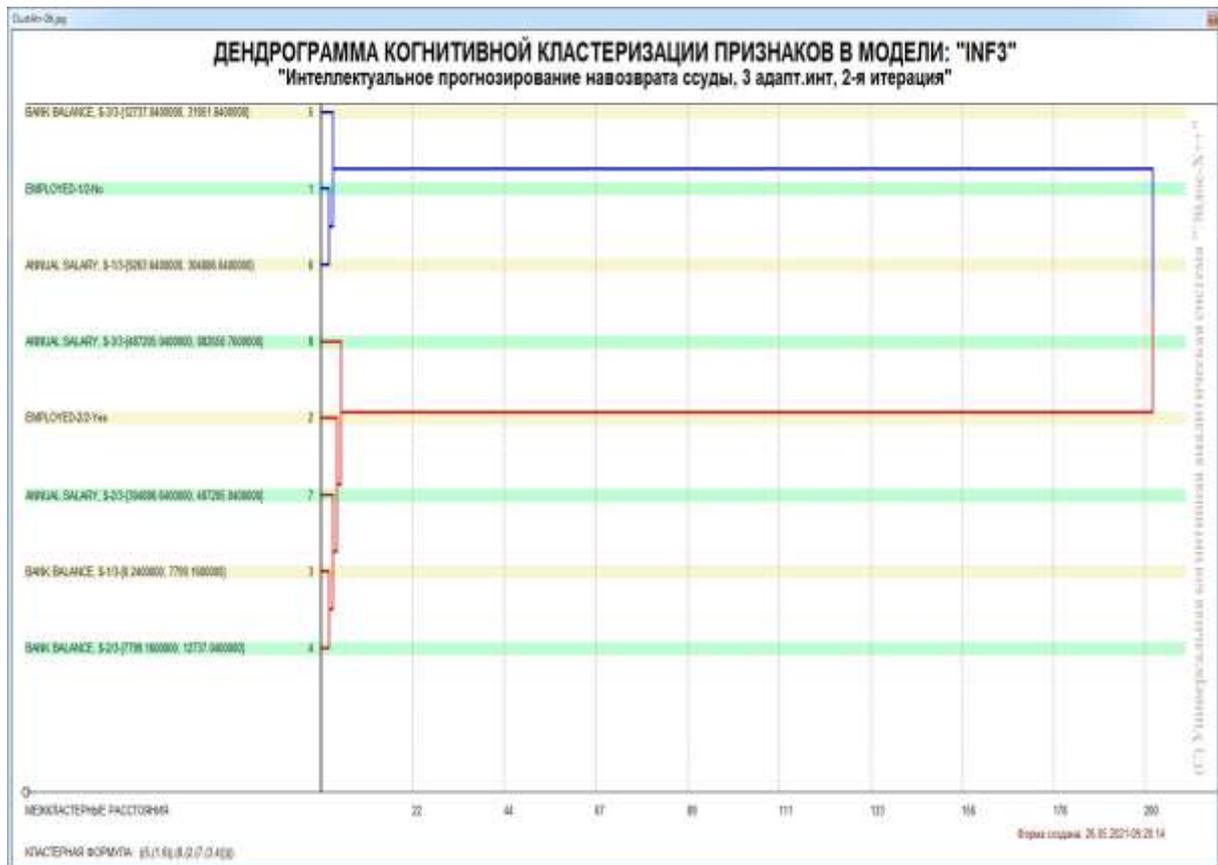


Рисунок 33. Дендрограмма агломеративной когнитивной кластеризации характеристик ссудополучателей по их влиянию на риск невозврата ссуды

Из дендрограммы на рисунке 33 мы видим, что все значения факторов образуют 2 четко выраженных кластера, объединенных в полюса конструкта (показаны синими и красными цветами).

Хорошо видна группировка характеристик ссудополучателей по смыслу этих характеристик, т.е. по содержащейся в них информации о риске невозврата ссуды.

*Значения факторов на полюсах конструкта факторов (рисунок 33) обуславливают переход объекта моделирования в состояния, соответствующие классам, представленным на полюсах конструкта классов (рисунки 27 и 29).*

На рисунке 34 приведен график межкластерных расстояний значений признаков.

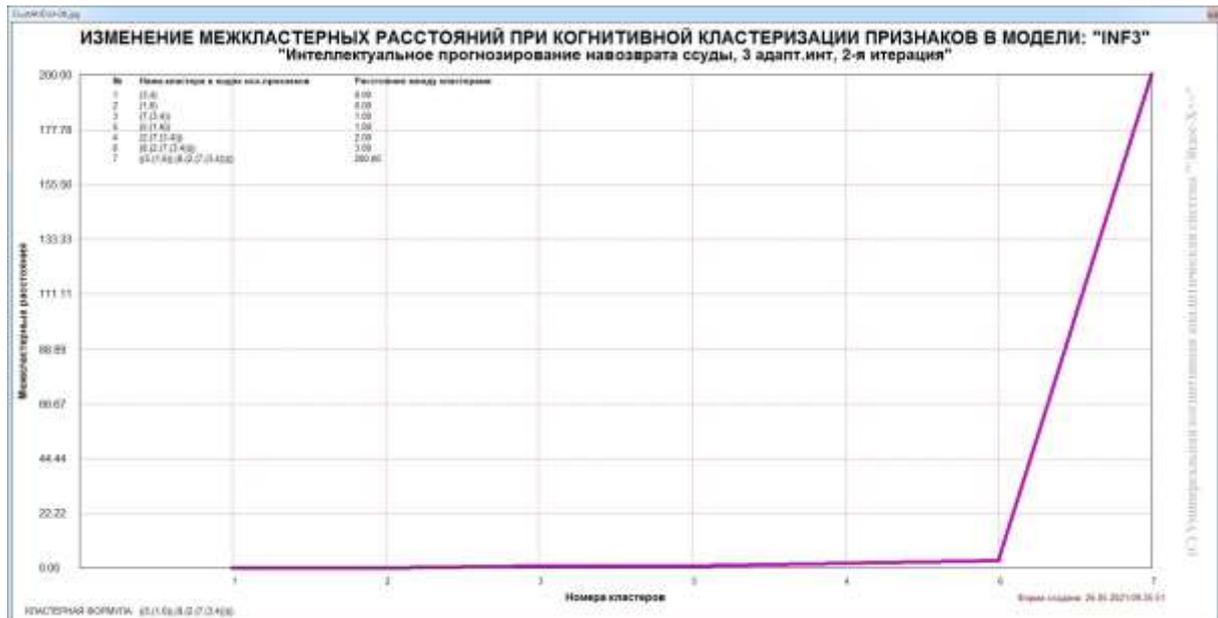


Рисунок 34. График изменения межкластерных расстояний при когнитивной кластеризации значений факторов

#### 4.3.4.4.5. Нелокальные нейроны и нелокальные нейронные сети

На рисунке 35 приведён пример нелокального нейрона, а на рисунке 36 – фрагмент одного слоя нелокальной нейронной сети [27]:

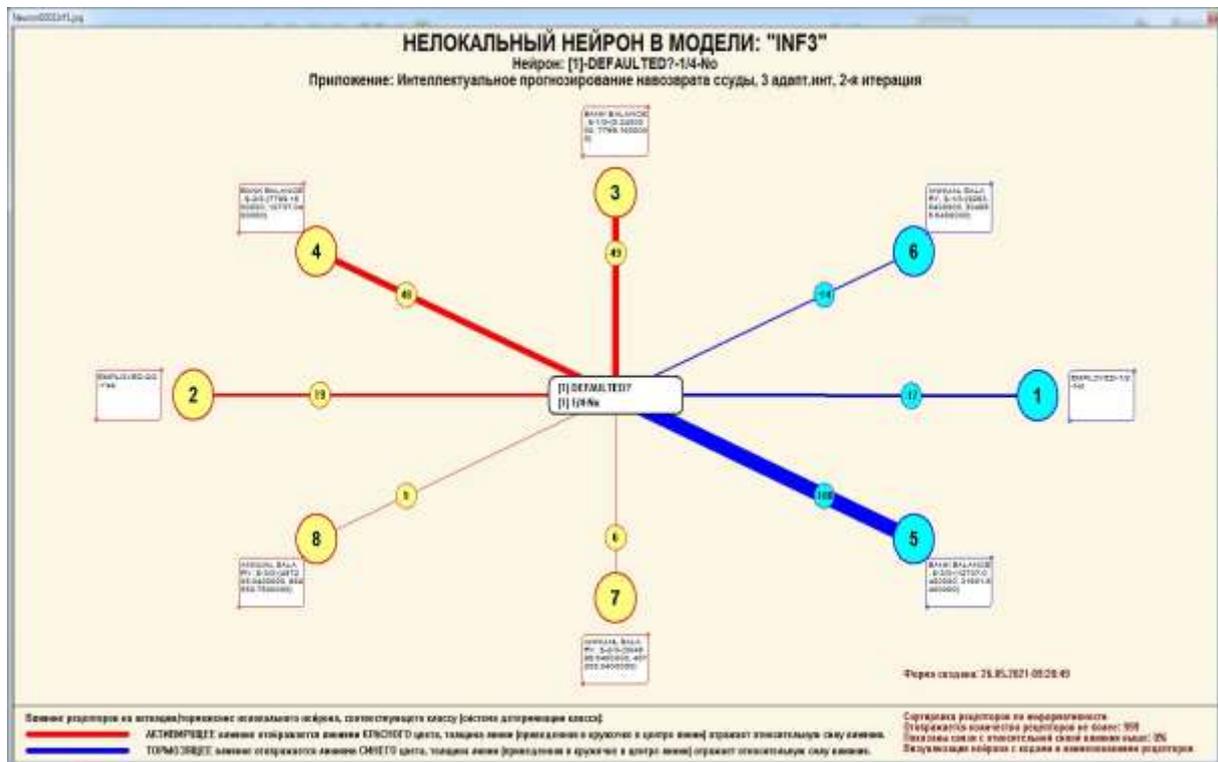


Рисунок 35. Пример нелокального нейрона, отражающего силу и направление влияния значений характеристик ссудополучателей на риск невозврата полученной ими ссуды

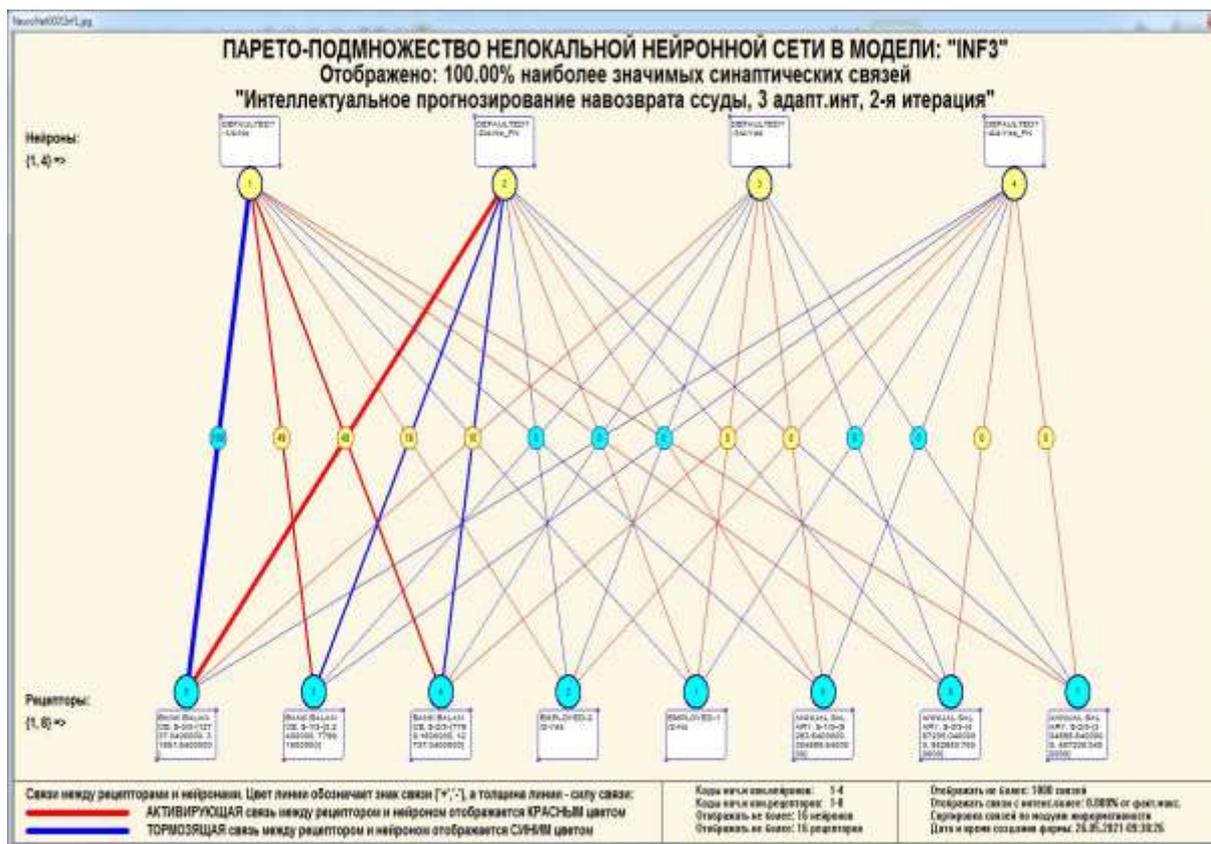


Рисунок 36. Один слой нелокальной нейронной сети, отражающий силу и направление влияния характеристик ссудополучателей на риск невозврата полученной ими ссуды

В приведенном фрагменте слоя нейронной сети нейроны соответствуют классу (рisku невозврата ссуды), а рецепторы – характеристикам ссудополучателей. Нейроны расположены слева направо в порядке убывания силы детерминации, т.е. слева находятся результаты, наиболее жестко обусловленные обуславливающими их значениями факторов, а справа – менее жестко обусловленные.

Модель знаний системы «Эйдос» относится к *нечетким декларативным* гибридным моделям и объединяет в себе некоторые особенности нейросетевой и фреймовой моделей представления знаний. Классы в этой модели соответствуют нейронам и фреймам, а признаки рецепторам и шпациям (описательные шкалы – слотам).

От фреймовой модели представления знаний модель системы «Эйдос» отличается своей эффективной и простой программной реализацией, полученной за счет того, что разные фреймы отличаются друг от друга не набором слотов и шпаций, а лишь информацией в них. Поэтому в системе «Эйдос» при увеличении числа фреймов само количество баз данных не увеличивается, а увеличивается лишь их размерность.

От нейросетевой модели представления знаний модель системы «Эйдос» отличается тем, что:

- 1) весовые коэффициенты на рецепторах не подбираются итерационным методом обратного распространения ошибки, а считаются прямым счетом на основе хорошо теоретически обоснованной модели, основанной на теории информации (это напоминает байесовские сети);
- 2) весовые коэффициенты имеют хорошо теоретически обоснованную содержательную интерпретацию, основанную на теории информации;
- 3) нейросеть является нелокальной, как сейчас говорят «полносвязной».

#### 4.3.4.4.6. 3d-интегральные когнитивные карты

На рисунке 37 приведена 3d-интегральная когнитивная карта в СК-модели Inf3.

3d-интегральная когнитивная карта является отображением на одном рисунке когнитивных диаграмм классов и значений факторов, отображенных соответственно на рисунках 27 и 31, и одного слоя нейронной сети, приведенного на рисунке 36.

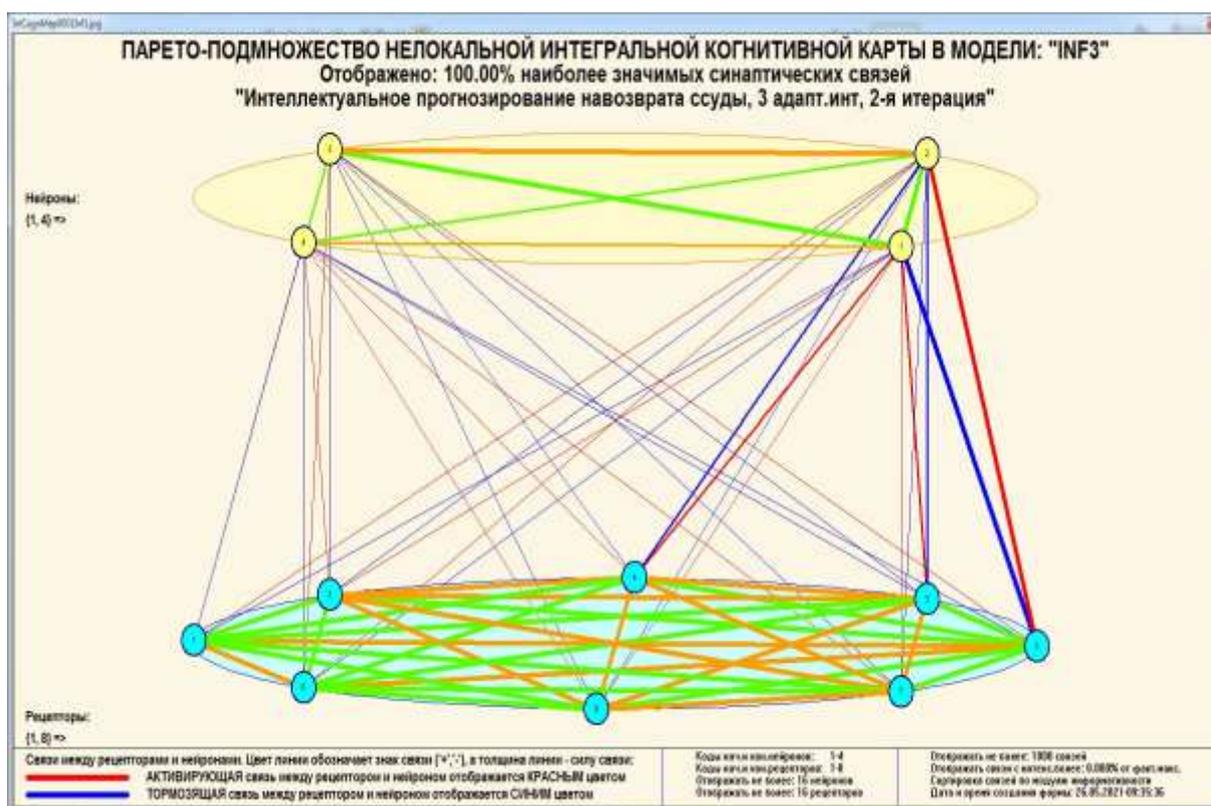


Рисунок 37. 3d-интегральная когнитивная карта в СК-модели Inf3

#### 4.3.4.4.7. Когнитивные функции

Вместо описания того, что представляют собой когнитивные функции, приведем help соответствующего режима системы «Эйдос» (рисунок 38) и сошлемся на работы, в которых описан этот подход [22]<sup>9</sup>.

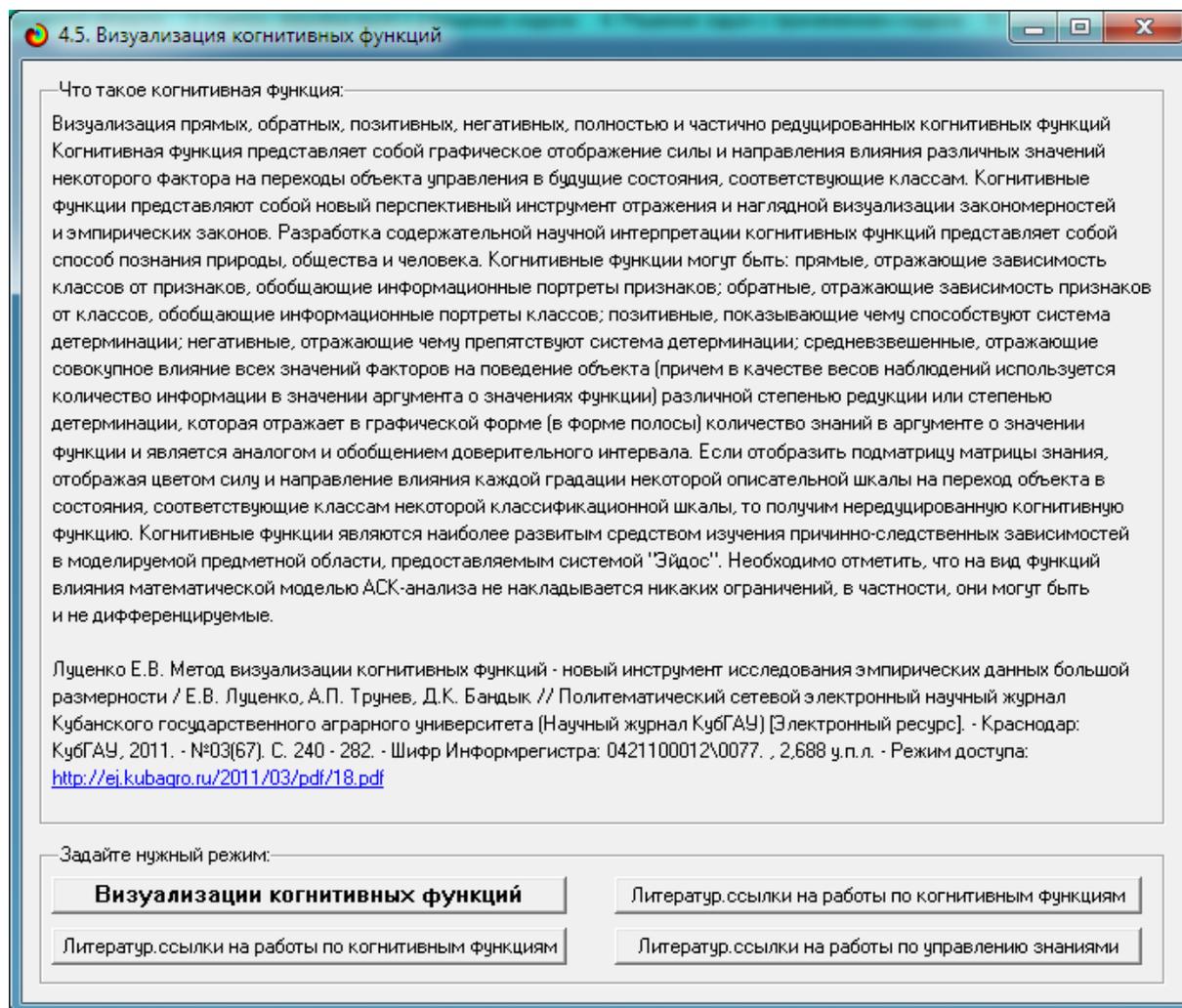


Рисунок 38. Help режима визуализации когнитивных функций

Когнитивная функция представляет собой графическое отображение силы и направления влияния различных значений некоторого фактора (признаков) на переходы объекта управления в будущие состояния, соответствующие классам. Классы являются градациями классификационных шкал.

Когнитивные функции представляют собой новый перспективный инструмент отражения и наглядной визуализации эмпирических закономерностей и эмпирических законов. Разработка содержательной

<sup>9</sup> Подборка публикаций проф.Е.В.Луценко & С<sup>o</sup> по когнитивным функциям: [http://lc.kubagro.ru/aidos/Works\\_on\\_cognitive\\_functions.htm](http://lc.kubagro.ru/aidos/Works_on_cognitive_functions.htm).

научной интерпретации когнитивных функций представляет собой способ познания природы, общества и человека [23]<sup>10</sup>.

Когнитивные функции могут быть: прямые, отражающие зависимость классов от признаков, обобщающие информационные портреты признаков; обратные, отражающие зависимость признаков от классов, обобщающие информационные портреты классов; позитивные, показывающие чему способствуют система детерминации (обозначены белой линией); негативные, отражающие чему препятствуют система детерминации (обозначены черной линией); средневзвешенные, отражающие совокупное влияние всех значений факторов на поведение объекта (причем в качестве весов наблюдений используется количество информации в значении аргумента о значениях функции) различной степенью редукции или степенью детерминации, которая отражает в графической форме (в форме полосы разной толщины) количество знаний в аргументе о значении функции и является аналогом и обобщением доверительного интервала.

Если отобразить подматрицу матрицы знания, отображая цветом силу и направление влияния каждой градации некоторой описательной шкалы на переход объекта в состояния, соответствующие классам некоторой классификационной шкалы, то получим нередуцированную когнитивную функцию.

Когнитивные функции являются наиболее развитым средством изучения причинно-следственных зависимостей в моделируемой предметной области, предоставляемым системой "Эйдос".

Необходимо отметить, что *на вид функций влияния математической моделью АСК-анализа не накладывается никаких ограничений*, в частности, они могут быть и не дифференцируемые.

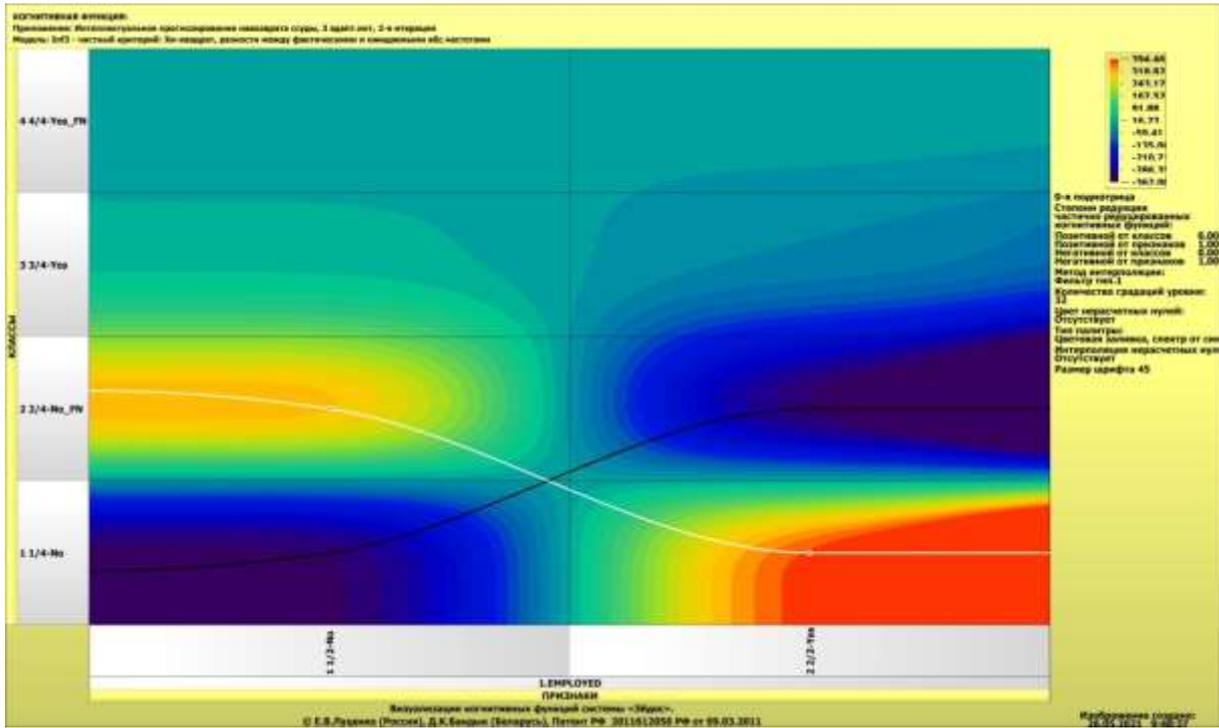
На рисунках 39 приведены когнитивные функции, наглядно отражающие силу и направление влияния значений (т.е. степени выраженности) различных характеристик ссудополучателей на риск невозврат полученных ими ссуд (класс).

Из когнитивных функций, приведенных на рисунке 39, хорошо видно, что *зависимости между характеристиками ссудополучателей и риском невозврата полученной ими ссуды имеют ярко выраженный и вполне очевидный и предсказуемый характер*.

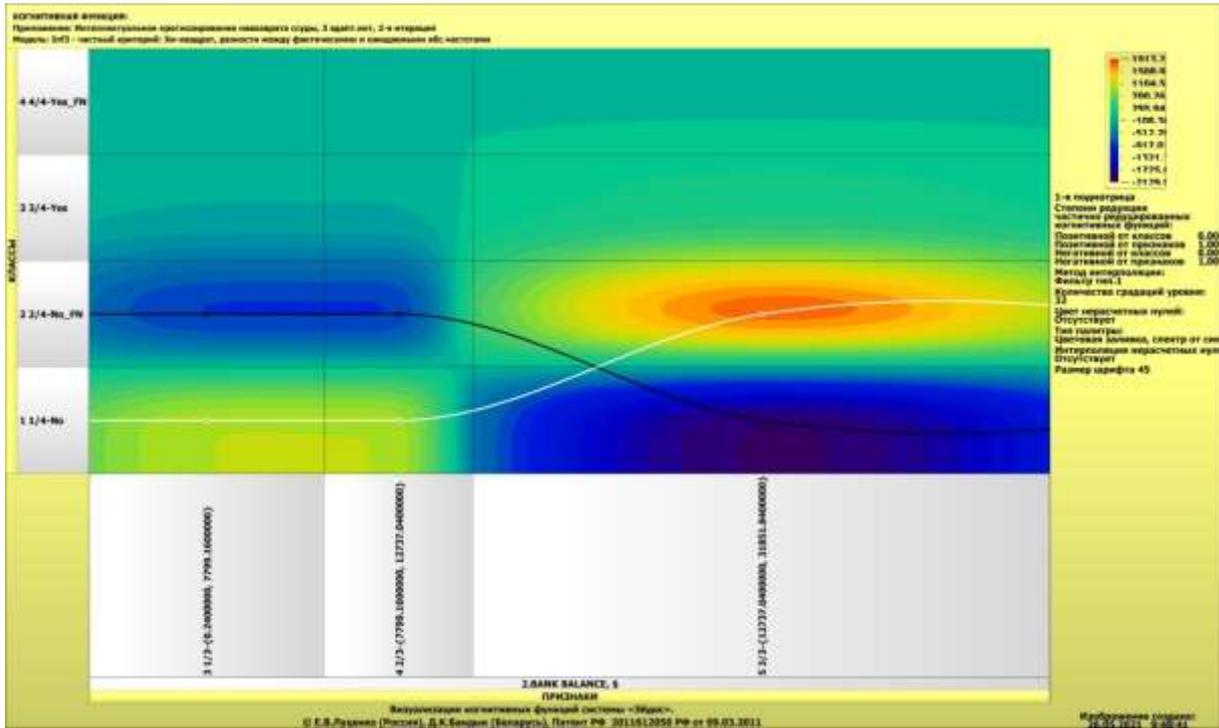
Но есть и несколько интересных неожиданных моментов, требующих специальной содержательной интерпретации. Эта содержательная интерпретация является делом специалистов по субсидированию и не входит в задачи данной работы.

---

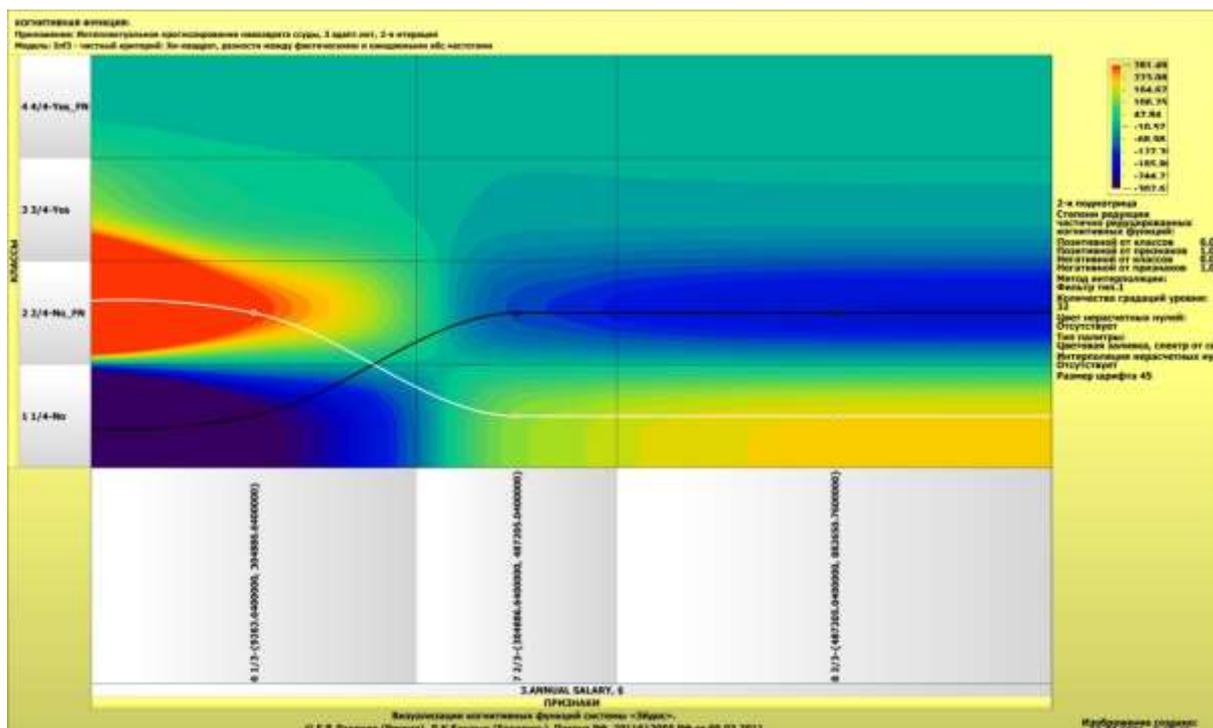
<sup>10</sup> Работы проф.Е.В.Луценко & С<sup>о</sup> по выявлению, представлению и использованию знаний, логике и методологии научного познания:  
[http://lc.kubagro.ru/aidos/Works\\_on\\_identification\\_presentation\\_and\\_use\\_of\\_knowledge.htm](http://lc.kubagro.ru/aidos/Works_on_identification_presentation_and_use_of_knowledge.htm)



А)



Б)



В)

Рисунок 39. Примеры когнитивных функций в СК-модели INF3, отражающих силу и направление влияния значений характеристик ссудополучателей на риск невозврата полученной ими ссуды

#### 4.3.4.4.8. Сила и направление влияния значений факторов на принадлежность к классам

На рисунках 15 приведены некоторые статистические и системно-когнитивные модели, отражающие моделируемую предметную область.

Строки матриц моделей соответствуют значениям факторов, т.е. значениям характеристик ссудополучателей (градации описательных шкал).

Колонки матриц моделей соответствуют различным классам, отражающим риск невозврата ссуды (градации классификационных шкал).

Числовые значения в ячейках матриц моделей, находящихся на пересечении строк и колонок, отражают направление (знак) и силу влияния конкретной характеристики, соответствующей строке, на конкретное значение класса – риска невозврата ссуды для ссудополучателем с такой характеристикой.

Если какая-то характеристика слабо влияет на класс риска невозврата ссуды, то в соответствующей строке матрицы модели будут малые по модулю значения разных знаков, если же влияние сильное – то и значения будут большие по модулю разных знаков.

Если какая-либо характеристика способствует определённому риску невозврат ссуды, то в соответствующей этому результату ячейке матрицы

модели будут положительные значения, если же понижает – то и значения будут отрицательные.

Из этого следует, что суммарную силу влияния той или иной характеристики ссудополучателя на класс (т.е. ценность данного значения характеристики для решения задачи прогнозирования риска невозврата ссуды и других задач) можно количественно оценивать **степенью вариабельности значений** в строке матрицы модели, соответствующей этой характеристике.

Существует много мер вариабельности значений: это и среднее модулей отклонения от среднего, и дисперсия, и среднеквадратичное отклонение и другие. В АСК-анализе и системе «Эйдос» для этой цели принято использовать среднеквадратичное отклонение. Численно оно равно стандартному отклонению и вычисляется по той же формуле, но мы предпочитаем не использовать термин «стандартное отклонение», т.к. он предполагает нормальность распределения исследуемых последовательностей чисел, а значит и проверку соответствующих статистических гипотез.

Самая правая колонка в матрицах моделей на рисунках 15 содержит количественную оценку вариабельности значений строки модели (среднеквадратичное отклонение), которая и представляет собой ценность характеристики, соответствующего строке, для решения задачи прогнозирования риска невозврата ссуды и других задач, рассмотренных в данной работе.

Если рассортировать матрицу модели по этой самой правой колонке в порядке убывания, а потом просуммировать значения в ней нарастающим итогом, то получим логистическую Парето-кривую, отражающую зависимость ценности модели от числа наиболее ценных признаков в ней (рисунок 40, таблица 12).

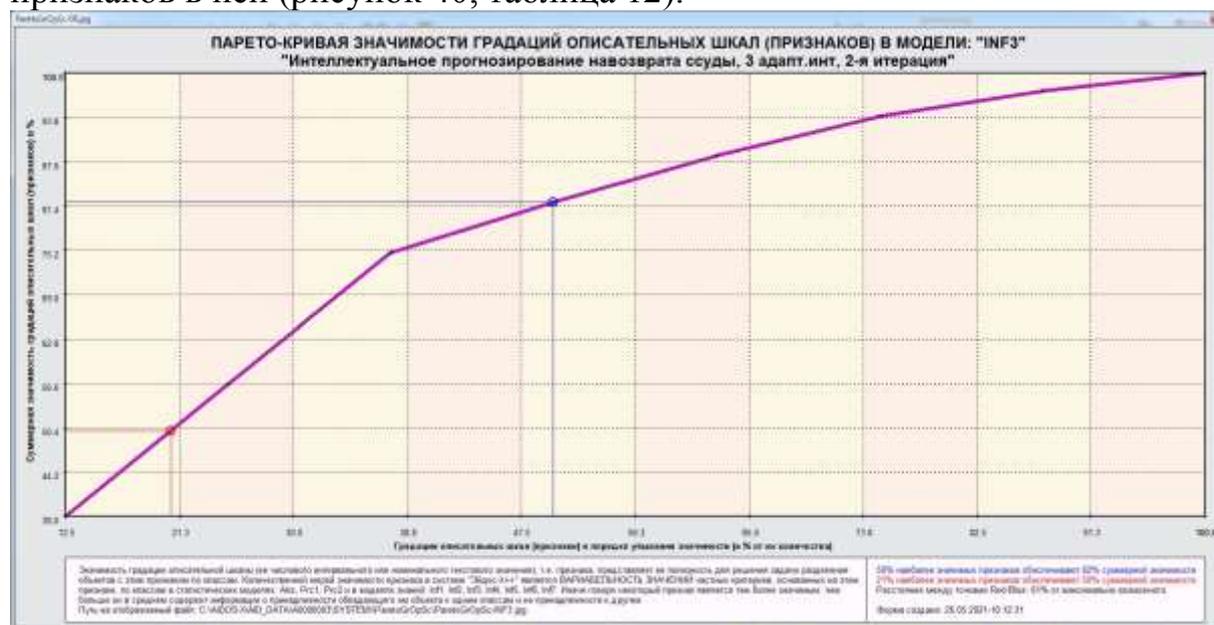


Рисунок 40. Парето-кривая значимости градаций описательных шкал

Таблица 12 – Парето-таблица значимости градаций описательных шкал, т.е. сила влияния значений характеристик ссудополучателя на риск невозврата ссуды в СК-модели INF3

№	Код	Наименование	Код шкалы	Значимость, %	Значимость нараст. итогом, %
1	5	BANK BALANCE, \$-3/3-{12737.0400000, 31851.8400000}	2	37,990	37,990
2	3	BANK BALANCE, \$-1/3-{0.2400000, 7799.1600000}	2	18,478	56,468
3	4	BANK BALANCE, \$-2/3-{7799.1600000, 12737.0400000}	2	18,390	74,858
4	2	EMPLOYED-2/2-Yes	1	7,094	81,952
5	1	EMPLOYED-1/2-No	1	6,534	88,486
6	6	ANNUAL SALARY, \$-1/3-{9263.6400000, 304886.6400000}	3	5,477	93,962
7	8	ANNUAL SALARY, \$-3/3-{487205.0400000, 882650.7600000}	3	3,594	97,556
8	7	ANNUAL SALARY, \$-2/3-{304886.6400000, 487205.0400000}	3	2,444	100,000

Данные, приведенные на рисунке 35 и в таблице 12, находится в XLS-файлах, созданных в режиме 3.7.5. Информация об этом содержится в экранной форме на рисунке 41:

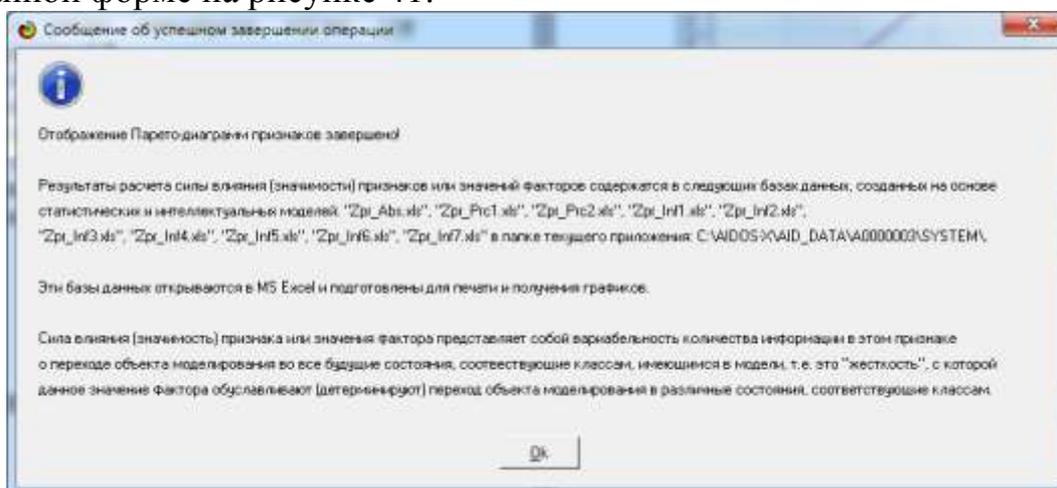


Рисунок 41. Информация о XLS-файлах

Из рисунка 40 и таблицы 12 видно, что 50% наиболее ценных для решения задачи прогнозирования рисков невозврата ссуд обуславливают 82% суммарной ценности, а 50% суммарной ценности обеспечиваются 21% наиболее ценных характеристик ссудополучателей.

Обращаем внимание на то, что наиболее ценной для решения задачи прогнозирования риска невозврата ссуды является максимальный баланс в банке ссудополучателя, а наименее ценным – средний заработок за год.

Из таблицы 12 видно, что наиболее сильное влияние на риск невозврата ссуды оказывают следующие характеристики ссудополучателя:

- BANK BALANCE, \$-3/3-{12737.0400000, 31851.8400000};
- BANK BALANCE, \$-1/3-{0.2400000, 7799.1600000};

а наиболее низкое:

- ANNUAL SALARY, \$-3/3-{487205.0400000, 882650.7600000};
- ANNUAL SALARY, \$-2/3-{304886.6400000, 487205.0400000}.

При этом сила влияния наиболее и наименее значимых значений факторов классов отличается в 15 раз, что очень существенно.

Ценность же характеристик ссудополучателей (всей описательной шкалы или фактора), для решения этих задач можно количественно оценивать как среднее от ценности значений этого параметра (таблица 13). Это можно сделать в режиме 3.7.4 (рисунок 42):

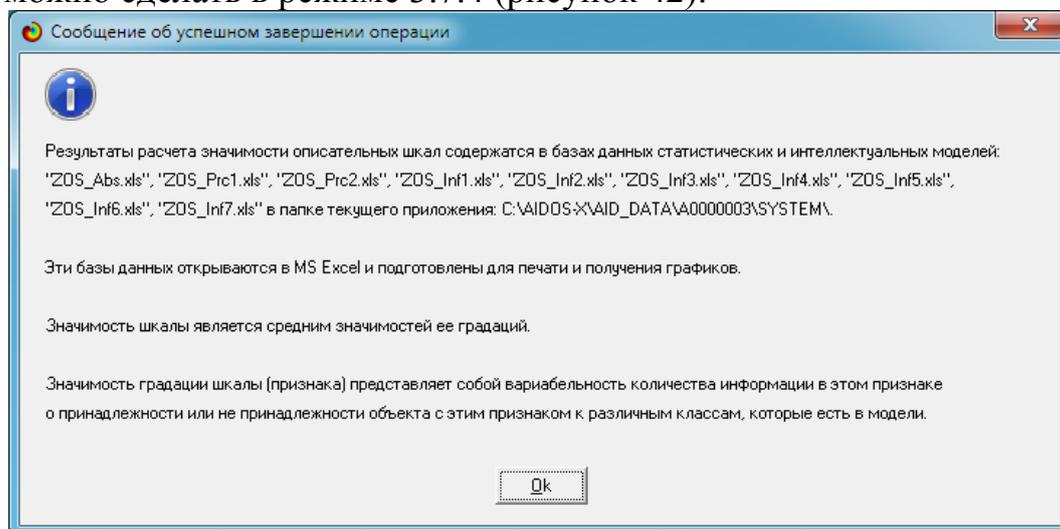


Рисунок 42. Информация о XLS-файлах

Таблица 13 – Парето-таблица значимости описательных шкал, т.е. сила влияния характеристик ссудополучателя на риск невозврата ссуды в СК-модели INF3

№	Код	Название описательной шкалы	Значимость, %	Значимость нарастающим итогом
1	2	BANK BALANCE, \$	70,0829613	70,0829613
2	1	EMPLOYED	19,1372199	89,2201812
3	3	ANNUAL SALARY, \$	10,7798188	100,0000000

Из таблицы 8 видно, что наиболее сильное влияние на прогнозирование риска невозврата ссуды оказывает баланс счета в банке, который дает более 2/3 суммарной значимости, в три с лишним раза меньшее влияние на это оказывает то, является ли ссудополучатель работающим или безработным, и еще в два раза меньшее влияние оказывает заработная плата за год.

При этом сила влияния наиболее и наименее значимых факторов классов отличается в 7 раз, что очень существенно.

#### 4.3.4.4.9. Степень детерминированности классов значениями обуславливающих их факторов

Степень детерминированности (обусловленности) класса в системе «Эйдос» количественно оценивается *степенью варибельности значений факторов* (градаций описательных шкал) в колонке матрицы модели, соответствующей данному классу (таблица 14).

В данной работе у нас классами являются риски невозврата ссуд, а значениями градаций описательных шкал – характеристики ссудополучателей.

На рисунке 43 мы видим Парето-кривую степени детерминированности классов значениями характеристик нарастающим итогом.



Рисунок 43. Парето-кривая степени детерминированности классов

Эта информация есть и в табличной форме (рисунок 44, таблица 14):

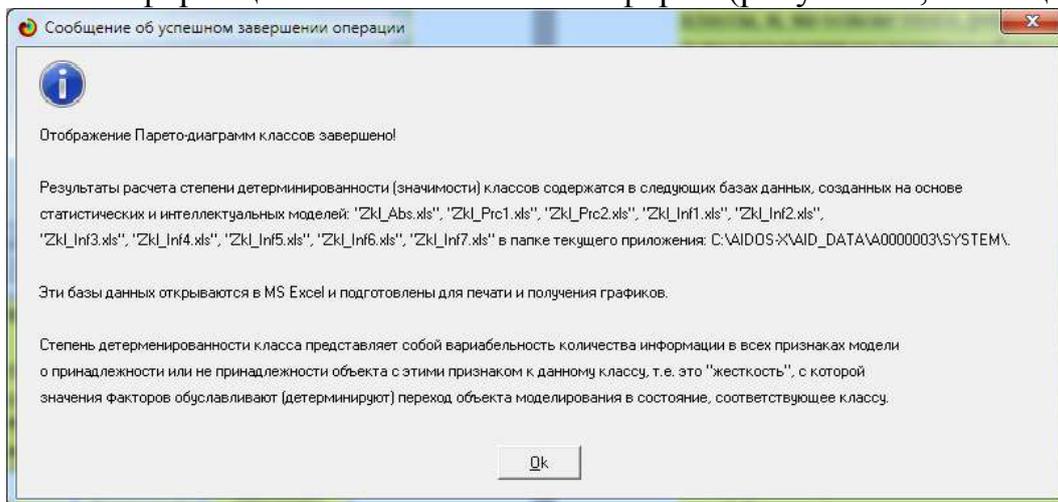


Рисунок 44. Информация о XLS-файлах

Таблица 14 – Парето-таблица степеней детерминированности (обусловленности) классов в СК-модели INF3

№	Код класса	Название класса	Значимость, %	Значимость нарастающим итогом
1	1	DEFAULTED?-1/4-No	49,879	49,879
2	2	DEFAULTED?-2/4-No_FN	44,913	94,793
3	3	DEFAULTED?-3/4-Yes	5,056	99,849
4	4	DEFAULTED?-4/4-Yes_FN	0,151	100,000

Из таблицы 9 мы видим, что значения характеристик ссудополучателей наиболее сильно детерминируют (обуславливают) такие

- DEFAULTED?-1/4-No;
- DEFAULTED?-2/4-No\_FN;

а наименее сильно:

- DEFAULTED?-3/4-Yes;
- DEFAULTED?-4/4-Yes\_FN.

При этом степень детерминированности наиболее и наименее детерминированных классов отличается в 6622 раз, что очень существенно.

Чем выше степень детерминированности класса риска невозврата ссуды характеристиками ссудополучателя, тем легче определить этот класс по этим характеристикам.

#### **4.3.5. Выводы**

Как показывает анализ результатов численного эксперимента предложенное и реализованное в системе «Эйдос» решение поставленных задач является вполне эффективным, что позволяет обоснованно утверждать, что цель работы достигнута, поставленная проблема решена.

В результате проделанной работы, с помощью системы «Эйдос» были созданы 3 статистические и 7 системно-когнитивных моделей, в которых непосредственно на основе эмпирических данных сформированы обобщенные образы классов по различным рискам невозврата ссуды, изучено влияние характеристик ссудополучателей на эти классы, и, на основе этого, решены задачи идентификации и прогнозирования, классификации и исследования моделируемой предметной области путем исследования ее модели.

Со всеми моделями, созданными в данной статье, можно ознакомиться установив облачное Эйдос-приложение №293 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](http://lc.kubagro.ru/aidos/_Aidos-X.htm).

#### **4.4. Выводы**

По результатам данного раздела делается обоснованный вывод о том, что поставленные в работе задачи решены, цель достигнута, поставленная проблема успешно решена, эффективность внедрения подтверждена модельным численным примером, основанным на реальных данных портала Kaggle.

На основании этого делается обоснованный вывод том, что научное решение, ранее доведенное до статуса новаций, в данном разделе до инновационного уровня, т.е. полностью готово для широкого применения на практике.

## **Заключение**

Объект исследования – прогнозирование риска невозврата ссуды.

Предмет исследования – прогнозирование рисков невозврата ссуды с применением интеллектуального итерационного алгоритма учета нетипичных случаев.

Проблема – возникающая при прогнозировании риска невозврата ссуды – ложно-отрицательные решения, возникающие, когда модель не относит ситуацию к классу (обобщенной категории), к которой она в действительности относится.

Целью данной работы является решение поставленной проблемы путем разработки алгоритма, обеспечивающего формирование обобщенных образов классов не только на основе типичных относящихся к ним наблюдений, но и на основе исключений, т.е. нетипичных случаев.

Путем декомпозиции цели получим последовательность **задач**, решение которых является этапами достижения цели: задача 1: когнитивная структуризация предметной области; задача 2: подготовка исходных данных и формализация предметной области; задача 3: синтез и верификация статистических и системно-когнитивных моделей. Улучшение наиболее достоверной модели путем применения итерационного алгоритма разделения обобщенных образов классов на типичную и нетипичную части; задача 4: решение различных задач в наиболее достоверной модели: подзадача 4.1. Прогнозирование (диагностика, классификация, распознавание, идентификация); подзадача 4.2. Поддержка принятия решений; подзадача 4.3. Исследование моделируемой предметной области путем исследования ее модели.

Все поставленные задачи были решены.

Задача 1. В качестве классификационной шкалы выбран риск невозврата ссуды, в качестве описательных шкал: характеристики ссудополучателя работает ли он или нет, какой у него счет в банке и годовой доход.

Задача 2. Разработаны градации классификационных и описательных шкал и с их помощью закодированы исходные данные, в результате чего получена обучающая выборка.

Задача 3. Осуществлен синтез и верификация 3 статистических и 7 системно-когнитивных моделей, выбрана наиболее достоверная из них. Ей оказалась СК-модель INF3. Затем было повышено качество наиболее достоверной модели путем применения итерационного алгоритма разделения обобщенных образов классов на типичную и нетипичную части. В результате ее достоверность по L1 критерию повысилась на 61%.

Задача 4. В наиболее достоверной модели: были решены следующие подзадачи: подзадача 4.1. Прогнозирование (диагностика, классификация, распознавание, идентификация); подзадача 4.2. Поддержка принятия решений; подзадача 4.3. Исследование моделируемой предметной области

путем исследования ее модели: когнитивные диаграммы классов и значений факторов, агломеративная когнитивная кластеризация классов и значений факторов, нелокальные нейроны и нейронные сети, 3d-интегральные когнитивные карты, когнитивные функции), исследование силы и направления влияния факторов и степени детерминированности классов, обуславливающими их значениями факторов.

На основании всего этого можно сделать обоснованный вывод о том, что поставленные в работе задачи решены, цель достигнута, поставленная проблема успешно решена. Достигнутая в работе цель и решенная проблема соответствуют поставленным.

### Литература

1. Lutsenko E.V. Methods of writing scientific papers, logic and the manner in which scientific statements // February 2021, DOI: [10.13140/RG.2.2.23546.41920](https://doi.org/10.13140/RG.2.2.23546.41920), License [CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/), <https://www.researchgate.net/publication/349039044>
2. Луценко Е. В. Научная публицистика : учебник / Е. В. Луценко. – Краснодар : ВЦСКИ «Эйдос», 2020. – 188 с. DOI 10.13140/RG.2.2.36089.01126, <https://www.researchgate.net/publication/340264083>.
3. Луценко Е. В. Современные информационно-коммуникационные технологии в научно-исследовательской деятельности и образовании: учеб. пособие для асп. напр. подг.: 40.06.01 "Юриспруденция"/ Е. В. Луценко, Г. М. Меретуков, В. И. Лойко. – Краснодар: КубГАУ, 2019. – 256с. ISBN 978-5-907294-46-2, <https://www.researchgate.net/publication/339285519>.
4. Прокопьева А.А. Применение информационных технологий и математического моделирования в управлении банковскими рисками // Магист. дисс., Режим доступа: [https://dspace.spbu.ru/bitstream/11701/12281/1/Master\\_s\\_dissertation-7.pdf](https://dspace.spbu.ru/bitstream/11701/12281/1/Master_s_dissertation-7.pdf) (обращение: 23.05.2021).
5. Беляков А.В.. Резервы на возможные потери - экономическая сущность и учёт для целей налогообложения. Дата обращения: 27 февраля 2010. [Архивировано](#) 12 апреля 2012 года.
6. Курносенко А.А.. Особенности правового регулирования банковскими рисками в условиях рыночной экономики // Банковское право. — 2008. — Вып. 5.
7. Погорелова О.С.. Проблемы прогнозирования кредитных рисков // Банковское кредитование. — 2008. — Вып. 3.
8. Слуцкий А.А.. Концепция определения значения минимального резерва по ссудам // Банковское кредитование. — 2008. — Вып. 4.
9. Сухов А.В.. Управление кредитными рисками в России и Европе: сравнительный анализ // Управление в кредитной организации. — 2008. — Вып. 6.
10. Информационный бюллетень // Выпуск № 1(44), Москва 2013 год, Режим доступа: [http://www.ksp.mos.ru/upload/documents/inf\\_bulletin/KSP-144.pdf](http://www.ksp.mos.ru/upload/documents/inf_bulletin/KSP-144.pdf), (обращение: 23.05.2021).
11. Луценко Е.В., Лебедев Е.А., Подсистема автоматического формирования двоичного дерева классов семантической информационной модели (Подсистема "Эйдос-Tree"). Пат. № 2008610096 РФ. Заяв. № 2007613721 РФ. Опубл. от 09.01.2008. – Режим доступа: <http://lc.kubagro.ru/aidos/2008610096.jpg>, 3,125 / 2,500 у.п.л.
12. Луценко Е.В. Системно-когнитивный анализ временных рядов на примере фондового рынка (когнитивная структуризация и формализация предметной области) / Е.В. Луценко, Е.А. Лебедев // Политематический сетевой электронный научный журнал

Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №07(051). С. 1 – 37. – Шифр Информрегистра: 0420900012\0073, IDA [article ID]: 0510907001. – Режим доступа: <http://ej.kubagro.ru/2009/07/pdf/01.pdf>, 2,312 у.п.л.

13. Луценко Е.В. Системно-когнитивный анализ временных рядов на примере фондового рынка (синтез и верификация семантической информационной модели) / Е.В. Луценко, Е.А. Лебедев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №07(051). С. 38 – 46. – Шифр Информрегистра: 0420900012\0072, IDA [article ID]: 0510907002. – Режим доступа: <http://ej.kubagro.ru/2009/07/pdf/02.pdf>, 0,562 у.п.л.

14. Луценко Е.В. Системно-когнитивный анализ временных рядов на примере фондового рынка (прогнозирование, принятие решений и исследование предметной области) / Е.В. Луценко, Е.А. Лебедев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №07(051). С. 47 – 82. – Шифр Информрегистра: 0420900012\0071, IDA [article ID]: 0510907003. – Режим доступа: <http://ej.kubagro.ru/2009/07/pdf/03.pdf>, 2,25 у.п.л.

15. Луценко Е.В. Повышение качества моделей «knowledge management» путем разделения классов на типичную и нетипичную части / Е.В. Луценко, Е.А. Лебедев, В.Н. Лаптев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №10(054). С. 78 – 93. – Шифр Информрегистра: 0420900012\0109, IDA [article ID]: 0540910005. – Режим доступа: <http://ej.kubagro.ru/2009/10/pdf/05.pdf>, 1 у.п.л.

16. Луценко, Е. В. Определение кредитоспособности физических лиц и риска их кредитования / Е. В. Луценко, Е. А. Лебедев // Финансы и кредит. – 2006. – № 32(236). – С. 75-83. <https://elibrary.ru/item.asp?id=9290406>

17. Луценко Е.В. Астросоциотипология и спектральный анализ личности по астросоциотипам с применением семантических информационных мультимodelей / Е.В. Луценко, А.П. Трунев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №01(035). С. 101 – 151. – Шифр Информрегистра: 0420800012\0002, IDA [article ID]: 0350801010. – Режим доступа: <http://ej.kubagro.ru/2008/01/pdf/10.pdf>, 3,188 у.п.л.

18. Луценко Е.В. Повышение адекватности спектрального анализа личности по астросоциотипам путем их разделения на типичную и нетипичную части / Е.В. Луценко, А.П. Трунев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №02(036). С. 153 – 174. – Шифр Информрегистра: 0420800012\0017, IDA [article ID]: 0360802010. – Режим доступа: <http://ej.kubagro.ru/2008/02/pdf/10.pdf>, 1,375 у.п.л.

19. Луценко Е.В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ, 2002. – 605 с. <http://elibrary.ru/item.asp?id=18632909>

20. Lutsenko E.V. SYSTEM ANALYSIS AND DECISION MAKING (Automated system-cognitive analysis and solving problems of identification, decision-making and

research of the simulated subject area) // August 2020, DOI: [10.13140/RG.2.2.27247.05289](https://doi.org/10.13140/RG.2.2.27247.05289), License [CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/), <https://www.researchgate.net/publication/343998862>

21. Луценко Е.В. Метризация измерительных шкал различных типов и совместная сопоставимая количественная обработка разнородных факторов в системнокогнитивном анализе и системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №08(092). С. 859 – 883. – IDA [article ID]: 0921308058. – Режим доступа: <http://ej.kubagro.ru/2013/08/pdf/58.pdf> 1,562 у.п.л.

22. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика. Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с. ISBN 978-5-94672-757-0. <http://elibrary.ru/item.asp?id=21358220/>.

23. Луценко Е.В. Проблемы и перспективы теории и методологии научного познания и автоматизированный системно-когнитивный анализ как автоматизированный метод научного познания, обеспечивающий содержательное феноменологическое моделирование / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2017. – №03(127). С. 1 – 60. – IDA [article ID]: 1271703001. – Режим доступа: <http://ej.kubagro.ru/2017/03/pdf/01.pdf>, 3,75 у.п.л.

24. Луценко Е.В. Инвариантное относительно объемов данных нечеткое мультиклассовое обобщение F-меры достоверности моделей Ван Ризбергена в АСКанализе и системе «Эйдос» / Е.В. Луценко // Политематический сетевой 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.

25. Луценко Е.В. Количественный автоматизированный SWOT- и PEST-анализ средствами АСК-анализа и интеллектуальной системы «Эйдос-X++» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №07(101). С. 1367 – 1409. – IDA [article ID]: 1011407090. – Режим доступа: <http://ej.kubagro.ru/2014/07/pdf/90.pdf> 2,688 у.п.л.

26. Луценко Е.В. Метод когнитивной кластеризации или кластеризация на основе знаний (кластеризация в системно-когнитивном анализе и интеллектуальной системе «Эйдос») / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №07(071). С. 528 – 576. – Шифр Информрегистра: 0421100012\0253, IDA [article ID]: 0711107040. – Режим доступа: <http://ej.kubagro.ru/2011/07/pdf/40.pdf> 3,062 у.п.л.

27. Луценко Е.В. Системная теория информации и нелокальные интерпретируемые нейронные сети прямого счета / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2003. – №01(001). С. 79 – 91. – IDA [article ID]: 0010301011. – Режим доступа: <http://ej.kubagro.ru/2003/01/pdf/11.pdf> 0,812 у.п.л.

28. Луценко Е.В. Моделирование сложных многофакторных нелинейных объектов управления на основе фрагментированных зашумленных эмпирических данных большой размерности в системно-когнитивном анализе и интеллектуальной системе «Эйдос-X++» / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой

электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №07(091). С. 164 – 188. – IDA [article ID]: 0911307012. – Режим доступа: <http://ej.kubagro.ru/2013/07/pdf/12.pdf> 1,562 у.п.л.

29. Луценко Е.В. Открытая масштабируемая интерактивная интеллектуальная online среда для обучения и научных исследований на базе АСК-анализа и системы «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2017. – №06(130). С. 1 – 55. – IDA [article ID]: 1301706001. – Режим доступа: <http://ej.kubagro.ru/2017/06/pdf/01.pdf>, 3,438 у.п.л. [http://lc.kubagro.ru/aidos/Presentation\\_Aidos-online.pdf](http://lc.kubagro.ru/aidos/Presentation_Aidos-online.pdf)

30. Луценко Е.В., Открытая масштабируемая интерактивная интеллектуальная online среда «Эйдос» («Эйдос-online»). Свид. РосПатента РФ на программу для ЭВМ, Заявка № 2017618053 от 07.08.2017, Гос.рег.№ 2017661153, зарегистр. 04.10.2017. – Режим доступа: <http://lc.kubagro.ru/aidos/2017661153.jpg> 2 у.п.л.

31. Lutsenko E.V. Theoretical foundations, mathematical model and software tools for Automated system-cognitive analysis // July 2020, DOI: [10.13140/RG.2.2.21918.15685](https://doi.org/10.13140/RG.2.2.21918.15685), License [CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/), <https://www.researchgate.net/publication/343057312>

32. Сайт проф.Е.В.Луценко: <http://lc.kubagro.ru/>

33. Блог Е.В.Луценко в RG <https://www.researchgate.net/profile/Eugene-Lutsenko>

34. Канал проф.Е.В.Луценко на YouTube: [https://youtube.com/channel/UC\\_QF84d8SCaWxsnXnexNFzg](https://youtube.com/channel/UC_QF84d8SCaWxsnXnexNFzg)

## **Приложение 1. Программная реализация в системе «Эйдос» интеллектуального итерационного алгоритма учета нетипичных случаев**

```

*****
***** Из файла исходных данных "Inp_data.dbf" стандарта программного интерфейса 2.3.2.2 удаляются объекты обучающей выборки
***** с отрицательным уровнем сходства с классом, к которому они относятся. В данном режиме используются результаты распознавания.
*****
FUNCTION F3_7_6()

LOCAL GetList := {}

*Razrab()
*Running(.F.)
*RETURN NIL

Running(.T.)

IF M_KodAdmAppls = 0 // Выйти из системы если нет авторизации
  LB_Warning(L("Вы не авторизовались в системе (режим 1.1) и не можете ей пользоваться!"))
  Running(.F.)
  RETURN NIL
ENDIF

*IF ApplChange("3.7.6()") // Если не запущен режим, работающий с БД, то перейти в папку выбранного приложения
* Running(.F.)
* RETURN NIL
*ENDIF

IF ApplChange("") // Если не запущен режим, работающий с БД, то перейти в папку выбранного приложения
  Running(.F.)
  RETURN NIL
ENDIF

**** Проверки на наличие необходимых баз данных и сообщения, если их нет

IF .NOT. FILE(Disk_dir + "\AID_DATA\Inp_data\Inp_data.dbf")
  LB_Warning(L('В папке: нет базы данных: "Inp_data.dbf"!'), L('3.7.6. Разделение классов на типичную и нетипичную части'))
  Running(.F.)
  RETURN NIL
ENDIF

IF .NOT. FILE(Disk_dir + "\_2_3_2_2.arx")
  LB_Warning(L("Необходимо создать модель в режиме 2.3.2.2."), L('3.7.6. Разделение классов на типичную и нетипичную части' ))
ELSE
  aSoftInt = DC_ARestore(M_PathAppl+"\_2_3_2_2.arx")
  aSoftInt[ 2] = 1 // Нули и пробелы считать отсутствием данных
  aSoftInt[15] = 1 // 1-равные интервалы, 2-адаптивные интервалы (с примерно равным количеством наблюдений, по Котельникову)
  // Если удалять классы, то можно использовать равные интервалы, а если добавлять, то можно и адаптивные
  aSoftInt[27] = 3 // Использовать Inp_data.dbf
  DC_ASave(aSoftInt , M_PathAppl+"\_2_3_2_2.arx")
  DC_ASave(aSoftInt , Disk_dir + "\_2_3_2_2.arx")
ENDIF

CloseAll() // Закрытие всех баз данных с ожиданием завершения операций

```

```

USE Inp_data EXCLUSIVE NEW

mFlagTXT = .T.
FOR mClSc = aSoftInt[3] TO aSoftInt[4]           // Цикл по классификационным шкалам
  IF FIELDTYPE(mClSc)="C"                       // Символьные столбцы
    mFlagTXT = .F.
    EXIT
  ENDIF
NEXT
IF mFlagTXT                                     // Нет текстовых классификационных шкал
  aMess := {}
  AADD(aMess, L('В файле исходных данных "Inp_data" нет текстовых классификационных шкал, '))
  AADD(aMess, L('а данный режим работает только с текстовыми классификационными шкалами. '))
  LB_Warning(aMess, L('3.7.6. Разделение классов на типичную и нетипичную части' ))
  Running(.F.)
  RETURN NIL
ENDIF

*MsgBox(M_PathAppl+"Rsp2i.dbf")
IF .NOT. FILE(M_PathAppl+"Rsp2i.dbf") .OR.;
  .NOT. FILE(M_PathAppl+"Rsp2k.dbf")
  aMess := {}
  AADD(aMess, L('В папке: ') + ALLTRIM(M_PathAppl) + L(' нет базы данных: "Rsp2i.dbf"!'))
  AADD(aMess, L('Необходимо выполнить режим 3.5, чтобы сформировать ее. '))
  LB_Warning(aMess, L('3.7.6. Разделение классов на типичную и нетипичную части' ))
  Running(.F.)
  RETURN NIL
ENDIF

*****

mIntKrit = 1
IF .NOT. FILE('_DelObj.txt')
  StrFile(ALLTRIM(STR(1)), '_mIntKrit.txt')      // Запись текстового файла _mIntKrit.txt
ENDIF
mIntKrit = VAL(FileStr('_mIntKrit.txt'))        // Загрузка текстового файла _mIntKrit.txt

mDelObj = 1
IF .NOT. FILE('_DelObj.txt')
  StrFile(ALLTRIM(STR(1)), '_DelObj.txt')      // Запись текстового файла _DelObj.txt
ENDIF
mDelObj = VAL(FileStr('_DelObj.txt'))          // Загрузка текстового файла _DelObj.txt

mIntKrit = 2
@ 1, 0 DCGROUP oGroup1  CAPTION L('Какой интегральный критерий использовать?') SIZE 80.0, 3.5
@ 1, 2 DCRADIO mIntKrit VALUE 1 PROMPT L('1. Резонанс знаний') PARENT oGroup1
@ 2, 2 DCRADIO mIntKrit VALUE 2 PROMPT L('2. Сумма знаний') PARENT oGroup1

@1.2,50 DCPUSHBUTTON CAPTION L("Пояснение") SIZE 15, 1.5 ACTION {||Help376(), DC_GetRefresh(GetList)} PARENT oGroup1

@ 5, 0 DCGROUP oGroup2  CAPTION L('Как обрабатывать нетипичные объекты обучающей выборки:') SIZE 80.0, 3.5
@ 1, 2 DCRADIO mDelObj VALUE 1 PROMPT L('Удалять нетипичные объекты из "Inp_data.dbf" ') PARENT oGroup2
@ 2, 2 DCRADIO mDelObj VALUE 2 PROMPT L('Создавать новые классы для нетипичных объектов ') PARENT oGroup2

DCREAD GUI;
  TO lExit ;

```

```

FIT;
ADDBUTTONS;
MODAL;
TITLE L("4.2.2.4 (3.7.6) Разделение классов на типичную и нетипичную части")

*****
IF lExit
  ** Button Ok
ELSE
  *****
  ***** БД, открытые перед запуском главного меню
  ***** Восстанавливать их после выхода из функций главного меню
  *****
  CCloseAll() // Закрытие всех баз данных с ожиданием завершения операций
  DIRCHANGE(Disk_dir) // Перейти в папку с исполнимым модулем системы
  USE PathGrAp EXCLUSIVE NEW
  USE Appls EXCLUSIVE NEW
  USE Users EXCLUSIVE NEW
  *****
  Running(.F.)
  RETURN NIL
ENDIF
*****

ERASE('_IntKrit.txt');StrFile(ALLTRIM(STR(mIntKrit)), '_IntKrit.txt') // Запись текстового файла _ObjErr.txt
ERASE('_DelObj.txt');StrFile(ALLTRIM(STR(mDelObj)), '_DelObj.txt') // Запись текстового файла _DelObj.txt
*****

aSoftInt = DC_ARestore(M_PathAppl+"\_2_3_2_2.arx")

// Если в объектах обучающей выборки, приведших к FN-решениям, удалять классы, то допустимы только равные интервалы,
// а если для таких решений классы добавлять, то можно использовать и адаптивные интервалы
IF mDelObj = 1 // Удалять нетипичные объекты из "Inp_data.dbf"
  aSoftInt[ 2] = 1 // Нули и пробелы считать отсутствием данных
  aSoftInt[15] = 1 // 1-равные интервалы, 2-адаптивные интервалы (с примерно равным количеством наблюдений, по Котельникову)
<<====#####
ENDIF
aSoftInt[27] = 3 // Использовать Inp_data.dbf
DC_ASave(aSoftInt , M_PathAppl+"\_2_3_2_2.arx")
DC_ASave(aSoftInt , Disk_dir +"\_2_3_2_2.arx")

CCloseAll() // Закрытие всех баз данных с ожиданием завершения операций
COPY FILE ("Inp_data.dbf") TO ("ObjFalseNeg.dbf")
COPY FILE ("Inp_data.dbf") TO ("InpDataSource.dbf")

CCloseAll() // Закрытие всех баз данных с ожиданием завершения операций
USE Inp_data EXCLUSIVE NEW
USE ObjFalseNeg EXCLUSIVE NEW;ZAP

DO CASE
  CASE mIntKrit=1
    USE Rsp2k EXCLUSIVE NEW
  CASE mIntKrit=2
    USE Rsp2i EXCLUSIVE NEW
ENDCASE

nMax = RECCOUNT()

```

```

Mess = L('4.2.2.4 (3.7.6). Разделение классов на типичную и нетипичную части')
@ 4,5 DCPROGRESS oProgress SIZE 70,1.1 MAXCOUNT nMax COLOR GRA_CLR_CYAN PERCENT EVERY 100
DCREAD GUI TITLE Mess PARENT @oDialog FIT EXIT
oDialog:show()
nTime = 0
DC_GetProgress(oProgress,0,nMax)

mNObjFN = 0

DBGOTOP()
DO WHILE .NOT. EOF()

    mKodObj = KOD_OBJ
    mKodClSc = KOD_CLSC

    mNameClsOld = ALLTRIM(NAME_CLS)
    mPos = RAT('-', mNameClsOld)
    mNameCls = SUBSTR(mNameClsOld, mPos+1, LEN(mNameClsOld)-mPos)

    IF LEN(ALLTRIM(Fakt)) > 0 .AND. IF(mIntKrit=1, KORR, SUM_INF) < 0

*       MsgBox('Kod_obj='+ALLTRIM(STR(KOD_OBJ))+ ' '+ALLTRIM(NAME_OBJ)+ ' '+ALLTRIM(STR(KOD_CLS))+ ' '+ALLTRIM(NAME_CLS)+ ' '+ALLTRIM(STR(KOD_CLSC))+ ' '+ALLTRIM(STR(KORR))+ ' '+ALLTRIM(STR(SUM_INF))+ ' '+ALLTRIM(FAKT))

        SELECT Inp_data
        DBGOTO(mKodObj)

*       MsgBox('Kod_obj='+ALLTRIM(STR(FIELDGET(1)))+ ' '+ALLTRIM(FIELDGET(2))+ ' '+ALLTRIM(STR(FIELDGET(3)))+ ' '+ALLTRIM(STR(FIELDGET(4)))+ ' '+ALLTRIM(FIELDGET(5)))

        mNumClSc = aSoftInt[3]+mKodClSc-1 // Номер колонки классификационной шкалы в БД Inp_data
        IF FIELDTYPE(mNumClSc) = "C" // Текстовые классификационные шкалы
            mNObjFN++
            IF mDelObj = 1 // Удалять объекты обучающей выборки
                FIELDPUT(mNumClSc, ' ' ) // Класс - отсутствие данных
            ELSE
                FIELDPUT(mNumClSc, mNameCls+'_FN' ) // Класс - отсутствие данных
            ENDIF
            aObj := {}
            FOR j=1 TO FCOUNT()
                AADD(aObj, FIELDGET(j))
            NEXT
            SELECT ObjFalseNeg // БД с объектами обучающей выборки, приведшими к ложно-отрицательным решениям
            APPEND BLANK
            FOR j=1 TO LEN(aObj)
                FIELDPUT(j, aObj[j])
            NEXT
        ENDIF

    ENDIF

DO CASE
CASE mIntKrit=1
SELECT Rsp2k
CASE mIntKrit=2
SELECT Rsp2i
ENDCASE

```

```

DC_GetProgress(oProgress, ++nTime, nMax)
DBSKIP(1)

ENDDO

*MsgBox('STOP')
DC_GetProgress(oProgress, nMax, nMax)
oDialog.Destroy()

**** После корректировки файла Inp_data.dbf в папке приложения записать его в ..\AID_DATA\Inp_data\

*MsgBox(M_ApplsPath+"\Inp_data\Inp_data.dbf")

ClosesAll() // Закрытие всех баз данных с ожиданием завершения операций
COPY FILE ("Inp_data.dbf") TO (M_ApplsPath+"\Inp_data\Inp_data.dbf")
COPY FILE ("ObjFalseNeg.dbf") TO (M_ApplsPath+"\Inp_data\ObjFalseNeg.dbf")
COPY FILE ("ObjFalseNeg.dbf") TO (M_ApplsPath+"\Inp_data\ObjFalseNeg.xls")

aMess := {}
IF mDelObj = 1
    AADD(aMess, L('Удалено:')+' '+ALLTRIM(STR(mNObjFN))+' '+L('нетипичных объектов обучающей выборки.'))
ENDIF
IF mDelObj = 2
    AADD(aMess, L('Назначено на новые классы:')+' '+ALLTRIM(STR(mNObjFN))+' '+L('нетипичных объектов обучающей выборки.'))
ENDIF
AADD(aMess, L('БД с объектами обучающей выборки, приведшими к FN-решениям:')+' '+M_ApplsPath+"\Inp_data\ObjFalseNeg.xls")
AADD(aMess, L(' '))
AADD(aMess, L('Далее нужно выполнить режим 2.3.2.2 с параметрами, заданными по умолчанию.))
LB_Warning(aMess, L('4.2.2.4 (3.7.6) Разделение классов на типичную и нетипичную части'))

*****
***** ЗАПИСАТЬ ПАРАМЕТРЫ ДЛЯ 2.3.2.2, ЧТОБЫ ЗАГРУЗКА ШЛА ИЗ INP_DATA.DBF
*****

IF FILE("_2_3_2_2.arx")

    aSoftInt = DC_ARestore(Disk_dir + "_2_3_2_2.arx")
    aSoftInt[2] = 1
    aSoftInt[27] = 3
    DC_ASave(aSoftInt, Disk_dir + "_2_3_2_2.arx")

ENDIF

*****
**** Запустить 2.3.2.2, 3.5
*****

F2_3_2_2("", "") // Запуск универсального программного интерфейса с внешними базами данных

* Возникает ошибка в отображении хода исполнения. Так и не смог разобраться
* F3_5('GPU', '', '', '') // Какая модель?
* F3_5('GPU', 'SintRec', '3.7.6', 'ALL') // Какая модель?
* F3_5('GPU', 'SintRec', '3.5', 'ALL')

*****
**** БД, открытые перед запуском главного меню
**** Восстанавливать их после выхода из функций главного меню

```

```

*****
CcloseAll()           // Закрытие всех баз данных с ожиданием завершения операций
DIRCHANGE(Disk_dir)  // Перейти в папку с исполнимым модулем системы
USE_PathGrAp EXCLUSIVE NEW
USE_Appls EXCLUSIVE NEW
USE_Users EXCLUSIVE NEW
*****

AADD(aMess, L(' '))
AADD(aMess, L('Далее необходимо в режиме 3.5 создать и верифицировать модели: Abs, Prc1, Prc2, Inf1, Inf2, Inf3, Inf4, Inf5, Inf6, Inf7. '))
AADD(aMess, L('Затем в режиме 3.4 необходимо определить модель и интегральный критерий, при которых достигается максимальная достоверность.'))
AADD(aMess, L(' '))
AADD(aMess, L('Режим 4.2.2.4 (3.7.6) можно повторять до достижения необходимого достаточно высокого уровня достоверности моделей. '))
AADD(aMess, L(' '))
AADD(aMess, L('Если достоверность модели достаточно высока, то в ней корректно можно решать задачи идентификации и прогнозирования (4.1.2),'))
AADD(aMess, L(' принятия решений (4.4.8) и исследования объекта моделирования путем исследования его модели (режимы: 4.4.9, 4.4.10, 4.4.11,'))
AADD(aMess, L(' 4.4.12, 4.2.1, 4.2.2.1, 4.2.2.2, 4.2.2.3, 4.2.3, 4.3.2.1, 4.3.2.2, 4.3.2.3, 4.5, 3.7.5, 3.7.4., 3.7.3, 3.7.9 и т.д.) '))
LB_Warning(aMess, L('4.2.2.4 (3.7.6) Разделение классов на типичную и нетипичную части'))

Running(.F.)
RETURN NIL

*****

```