

УДК [004.78:33](075.8)

СИНТЕЗ СКОРИНГОВОЙ МОДЕЛИ МЕТОДОМ СИСТЕМНО-КОГНИТИВНОГО АНАЛИЗА

Лебедев Е.А., – аспирант

Кубанский государственный аграрный университет

В статье рассматривается актуальная проблема прогнозирования рисков кредитования физических лиц, и предлагаются пути решения поставленной задачи. Рассматриваются возможные варианты формализации предметной области, и формируется обучающая выборка, на основании которой в дальнейшем проводится синтез скоринговой модели методом системно-когнитивного анализа. Также в статье рассматриваются возможные варианты оптимизации модели с целью улучшения ее качества: т.е. повышения достоверности прогнозирования и снижения различного рода ошибок..

Ключевые слова: ПРОГНОЗИРОВАНИЕ РИСКОВ КРЕДИТОВАНИЯ ФИЗИЧЕСКИХ ЛИЦ ФОРМАЛИЗАЦИЯ ПРЕДМЕТНОЙ ОБЛАСТИ СКОРИНГОВАЯ МОДЕЛЬ МЕТОД СИСТЕМНО-КОГНИТИВНОГО АНАЛИЗА

Кредитно-финансовая система является одной из важнейших структур рыночной экономики, так как от темпов ее развития напрямую зависят темпы развития экономики в целом. Банки выполняют важную роль посредника, перераспределяя финансовые потоки из тех отраслей экономики, которые имеют избытки финансовых средств, в отрасли с потребностью в дополнительном финансировании. Особую роль в этом процессе играет кредитование.

Первые кредитные продукты в современной России появились в начале 90-х годов и в основном они были направлены на удовлетворение потребностей крупного бизнеса. Кредитование населения в условиях постоянно растущей инфляции, высокого уровня безработицы и сокращения рабочих мест, считалось делом не перспективным. Единственным полноценным игроком на рынке кредитования населения долгое время оставался Сбербанк России. Это привело к дисбалансу между

рынками кредитования физических и юридических лиц, последствия которого заметны до сих пор.

В настоящее время можно говорить о том, что рынок кредитования юридических лиц в России прошел фазу становления, крупные клиенты поделены между банками и резкого увеличения объема рынка в ближайшее время ждать не приходится.

Обратную картину можно наблюдать на рынке потребительского кредитования. Рост экономики привел к повышению благосостояния граждан, увеличению покупательной способности населения. Не смотря на наметившуюся тенденцию к снижению процентных ставок по кредитам предоставляемым населению, кредитование физических лиц остается более выгодным способом размещения свободных денежных средств, нежели кредитование юридических лиц. Так стоимость кредитных продуктов предлагаемых населению Сбербанком России на 01.05.2007 год составляет от 15% до 17% годовых, что значительно выше, чем стоимость кредитов предлагаемых клиентам – юридическим лицам – от 9% до 15%. Аналогичную картину можно наблюдать и в других банках.

Коммерческие банки активно включились в борьбу за клиента, предлагая новые все более привлекательные условия кредитования. Усилия банков не прошли даром. Количество желающих взять займы под проценты с каждым годом растет. По данным Банка России на 01.01.2006 года населению выдано кредитов на 1179,3 млрд. руб., что значительно больше, чем на 01.01.2005 года - 618,9 млрд. руб.

Участвуя в активных операциях, банки принимают на себя всевозможные риски. В случае с кредитованием – это риски не возврата заемных средств. Принятие рисков - основа банковского дела, но успех имеют только тот, кто принимает разумные риски, контролируемые и находящиеся в пределах финансовых возможностей банка. Конкуренция на рынке кредитования населения заставила банки вести более

агрессивную кредитную политику, чем прежде, направленную на увеличение кредитного портфеля за счет привлечения в короткие сроки широкого круга заемщиков. Эта задача была решена за счет упрощения процедуры кредитования (сокращен перечень необходимых документов для получения ссуды, ликвидирован институт поручительства). Привлекая клиентов, таким образом, банки приняли на себя дополнительные кредитные риски, которые, реализовавшись, привели к росту просроченной ссудной задолженности. По данным Банка России на 01.01.2006 г. просроченная задолженность по кредитам, выданным физическим лицам составила 22 млрд. руб., что на 39% больше чем на 01.01.2005 г. – 8,6 млрд. руб. По мнению экспертов если ситуации на рынке кредитование в ближайшее время не изменится, следующим кризисом в России может быть кризис банковской системы.

Уменьшение кредитных рисков – актуальная задача, стоящая перед коммерческими банками.

Существуют различные подходы к определению кредитного риска для физического лица, начиная с субъективных оценок специалистов банка, основанных на личном опыте и на впечатлении о конкретном клиенте, и заканчивая автоматизированными системами оценки риска, созданными с использованием математических моделей. Каждая кредитная организация сама определяет, какими методами пользоваться. Опыт зарубежных банков показывает, что методы, основанные на математических моделях, являются более устойчивыми и действенными.

Модели оценки кредитного риска, использующие математические алгоритмы называются скоринговыми моделями. Скоринг физических лиц представляет собой сложную математическую систему оценки, основанную на различных характеристиках клиентов, таких как личный доход, возраст, семейное положение, профессия и многих других. Они являются входными переменными модели, классифицирующей

потенциальных заемщиков. В результате анализа переменных, поступающих на вход скоринговой системы, на выходе системы скоринга получается интегрированный показатель, который и оценивает степень кредитоспособности заемщика по ранговой шкале: «хороший» заемщик или «плохой» заемщик.

Широкое распространение на Западе получила модель, известная как FICO Score. Она была разработана компанией Fair Isaac и используется многими банками за рубежом. Эта модель пока не является стандартом, но, поскольку, она считается наиболее полной моделью, она стала неотъемлемой частью практически любого процесса предоставления кредита.

В России скоринговые системы только начинают внедряться. Из-за возросшего спроса на потребительские кредиты и незначительной суммы каждого кредита большинство банков не могут себе позволить проводить оценку заемщика в индивидуальном порядке и все чаще прибегают к скоринговой оценке кредитоспособности клиента. Однако на рынке наблюдается дефицит отечественных скоринговых систем, что вынуждает отечественные банки пользоваться моделями, разработанными для Западных пользователей. Такие скоринговые модели не эффективны в условиях Российского рынка, яркой иллюстрацией чего, может служить резкий рост просроченной задолженности физических лиц. Для создания эффективной скоринговой модели необходима обучающая выборка (так называемое кредитное кладбище) – состоящее из кредитных историй по ранее выданным кредитам. Такое кредитное кладбище в нашей стране имеет только Сбербанк России.

Целью данной работы является применение развитых экономико-математических методов и инструментальных программных средств для исследования причинно-следственных зависимостей между

индивидуальными особенностями заемщика и его кредитоспособностью на основе архивных данных Сбербанка России.

В современных условиях отечественного рынка банковских услуг для определения кредитоспособности потенциальных заемщиков, возможно, использовать новый математический метод экономики – системно-когнитивный анализ (СК-анализ). Необходимо отметить, что этот универсальный метод хорошо теоретически обоснован, оснащен удобным программным инструментарием и успешно апробирован в ряде задач интеллектуальной обработки данных.

Специальным программным инструментарием СК-анализа, реализующим его математическую модель и методику численных расчетов, является универсальная когнитивная аналитическая система “Эйдос”, которая обеспечивает решение следующих задач:

1. Формализация предметной области.
2. Формирование обучающей выборки.
3. Синтез модели
4. Оптимизация.
5. Верификация модели.

Для синтеза модели были использованы данные из 400 кредитных досье заемщиков получивших кредит в Краснодарском отделении Сбербанка России №8619 в период с 2002 по 2006 гг. и имеющих кредитную историю.

1. Формализация предметной области.

Присвоение имен классификационным шкалам и градациям.

Под классификационными шкалами и градациями понимают справочник будущих состояний активного объекта управления. В нашем случае будущими состояниями модели является кредитная история

заемщика. В ходе работы над модель были испробованы различные варианты построения классов: дробление кредитной истории на части (от 3-х до 17-ти классов) по качеству совершаемых платежей, слияния полученных классов, по результатам кластерного анализа выполненного с помощью ПО “SPSS”. Ни один из способов не решал поставленную задачу. В результате было принято решение, о классификации предметной области основываясь на имеющихся в законодательстве и нормативных документах определений кредитной истории и ее разновидностях. Такое определение дано в Федеральном законе РФ от 30.12.2004 №218-ФЗ с учетом изменений от 21.07.2005 №110-ФЗ “О кредитных историях”, которое гласит, что **кредитная история** – это информация, которая характеризует исполнение заемщиком принятых на себя обязательств по договорам займа (кредита). Решение, какую кредитную историю считать “положительной”, а какую “отрицательной” банк принимает самостоятельно в зависимости от выбранной кредитной политики. Исходя из того, что для данной работы использованы материалы кредитных досье Сбербанка России, логично при классификации будущих состояний заемщиков руководствоваться определением “положительной” кредитной истории изложенным в Правилах кредитования физических лиц Сбербанком России и его филиалами от 30.05.2003 №229-Зр.

Положительная кредитная история в банке – одновременное выполнение следующих условий:

– платежи по основному долгу и/или процентам за пользование кредитом осуществляются заемщиком своевременно и в полном объеме. При этом допускается наличие не более 3-х случаев просрочки, каждая из которых не более 5 календарных дней, в любом годовом интервале за анализируемый период времени;

– отсутствие негативной информации о Заемщике в базе данных по Заемщикам – физическим лицам (в части предоставления Заемщиком

поддельных документов и/или недостоверной информации, неисполненных обязательств по кредитным договорам, договорам поручительств, выданных Заемщиком в обеспечение исполнения обязательств за третьих лиц, повлекших проведение Банком претензионно-исковой работы и др.).

Исходя из вышеизложенного определения, было принято решение о формировании двух классов заемщиков с “положительной” и “отрицательной” кредитной историей.

Присвоение имен описательным шкалам и градациям

Под описательными шкалами и градациями понимаются справочник факторов – признаков, влияющих на поведение активного объекта управления. Для решения задачи формализации предметной области решено остановиться на 17 описательных шкалах и 412 градациях. Описательные шкалы представлены в таблице 1. Так как количество градаций слишком велико, в рамках данной статьи градации расшифровываться не будут.

Признаки можно условно разделить на три группы: физиологические, социальные, финансовые (характеризующие материальное положение заемщика). К физиологическим относятся такие признаки, как пол и возраст заемщика, к социальным – место рождения, семейное положение, наличие иждивенцев, округ проживания, продолжительность проживания на последнем и предпоследнем месте, образование; к финансовым – место работы, сфера деятельности работодателя, организационно-правовая форма работодателя, должность, стаж на последнем месте работы, доходы, коэффициент долговой нагрузки, наличие собственности. Большинство признаков используемых в

данной работе в той или иной форме применялись ранее для создания различных скоринговых моделей.

**ТАБЛИЦА 1 – ОПИСАТЕЛЬНЫЕ ШКАЛЫ И ГРАДАЦИИ ИСПОЛЬЗУЕМЫЕ
ДЛЯ ФОРМАЛИЗАЦИИ ПРЕДМЕТНОЙ ОБЛАСТИ**

№ п.п.	Наименование описательной шкалы	Кол-во градаций описательной шкалы
1.	Пол	2
2.	Возраст	58
3.	Место рождения	5
4.	Семейное положение	8
5.	Наличие иждивенцев	4
6.	Округ проживания	4
7.	Продолжительность проживания на последнем месте	21
8.	Продолжительность проживания на предпоследнем месте	9
9.	Образование	5
10.	Место работы	188
11.	Сфера деятельности работодателя	39
12.	Организационно-правовая форма работодателя	5
13.	Должность	14
14.	Стаж на последнем месте работы	21
15.	Доходы	15
16.	Коэффициент долговой нагрузки	11
17.	Наличие собственности	3

Из-за высокой информативности широкое применение в скоринговых моделях получили такие признаки как пол и возраст заемщика. Также часто анализируется семейное положение и количество иждивенцев у заемщика, должность и доход. В данной работе автором используется оригинальный вторичный, т.е. расчетный признак – коэффициент долговой нагрузки, который показывает, какую часть доходов заемщик тратит на обслуживание ссудной задолженности. Формула расчета коэффициента долговой нагрузки выведена из формул

применяемых в Сбербанке России для расчета Платежеспособности заемщика (1.1) и расчета Максимальной суммы кредита. (1.2)

$$P = T \times D \times K, \quad (1.1)$$

где:

P – платежеспособность заемщика на момент обращения в банк за кредитом;

T - срок кредитования в целых месяцах;

D – среднемесячный доход;

K – коэффициент в зависимости от величины чистого дохода (0,7-0,9).

$$S = \frac{P}{1 + \frac{(T + 1) \times \%}{2 \times 12 \times 100}}, \quad (1.2)$$

где:

S – максимальный размер предоставляемого кредита;

% – годовая процентная ставка по кредиту.

Для вычисления коэффициента долговой нагрузки изменим значение некоторых переменных в формулах (1.1) и (1.2.) Так примем переменную K за неизвестную, значение переменной S изменим на сумму полученного кредита, а D на среднемесячный доход на 1 члена семьи заемщика. Формула для расчета коэффициента примет вид:

$$K = \frac{S \times \left(1 + \frac{(T + 1) \times \%}{2 \times 12 \times 100} \right)}{T \times D}, \quad (1.3)$$

где:

K – коэффициент долговой нагрузки;

D – среднемесячный доход на 1 члена семьи заемщика;

S – сумма полученного кредита.

2. Формирование обучающей выборки

Разработав описательные и классификационные шкалы, переходим к формированию обучающей выборки, которая включает в себя информацию о факторах влияющих на состояние объекта управления и о состоянии объекта. Информация в обучающей выборке шифруется согласно справочникам классов и признаков и принимает вид показанный в таблице 2.

ТАБЛИЦА 2 – ОБУЧАЮЩАЯ ВЫБОРКА (ФРАГМЕНТ)

№	Класс.	Описательные шкалы																
	шкалы	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	2	1	7	64	66	77	78	87	0	115	268	311	344	355	364	390	403	412
2	2	1	14	63	68	74	78	92	0	116	270	311	344	350	366	391	406	410
3	2	2	14	64	73	74	81	82	0	116	119	305	347	357	367	390	405	412
4	2	1	11	64	68	77	79	91	0	116	211	315	344	354	364	392	407	410
5	2	1	40	61	72	74	81	102	0	116	269	311	344	355	364	385	405	410
6	1	2	42	61	73	77	79	96	0	114	0	0	0	362	370	384	402	410
7	1	1	10	64	66	77	81	89	0	114	263	311	344	355	364	388	402	410
8	2	1	11	61	68	74	81	92	0	114	145	339	347	355	365	389	405	412
9	2	1	11	64	68	75	81	91	0	114	188	311	345	354	364	390	407	410
10	1	1	8	64	66	77	80	88	0	115	299	342	344	351	364	396	403	412

3. Синтез модели

С помощью системы “Эйдос” проведем синтез модели, который включает в себя расчет матрицы абсолютных частот, поиск и исключение

из дальнейшего анализа артефактов, расчет матрицы информативностей, расчет матрицы условных процентных распределений.

4. Оптимизация

Оптимизируем полученную модель с помощью удаления признаков, по которым имеется недостаточно данных. За пороговое значение встреч признаков в модели примем 5%. Удаление признаков решает сразу несколько задач. Во-первых, с уменьшением количества признаков упрощается задача анализа, т.к. отпадает необходимость анализировать влияние удаленных признаков на состояние объекта управления, во-вторых улучшается качество модели, т.к. редко встречающиеся признаки являются источниками шумов. После оптимизации количество градаций описательных шкал уменьшилось с 412 до 197.

5. Верификация модели

Скопируем полученную обучающую выборку в распознаваемую и проведем пакетное распознавание, после чего измерим адекватность информационной модели (измерение внутренней валидности). Из 400 анкет выборки, верно идентифицировалось 84,3% анкет, верно не идентифицировались 65,6% анкет, ошибочно не идентифицировались 15,7%, ошибочно идентифицировались 34,4%. Анализируя полученные данные можно предположить, что не все заемщики представленные в выборке сходны по своим признакам в разрезе классов, т.е. классы неоднородны. Так, не смотря на принадлежность заемщиков к одному из существующих классов, 15,7% анкет не были идентифицировано.

Для решения задачи 100% идентификации анкет предложено повторить этап оптимизации модели. Суть оптимизации состоит в сохранении существующих классов состоящих из верно идентифицирующихся типичных анкет заемщиков и добавлении новых классов состоящих из не идентифицирующихся нетипичных (рис.1),

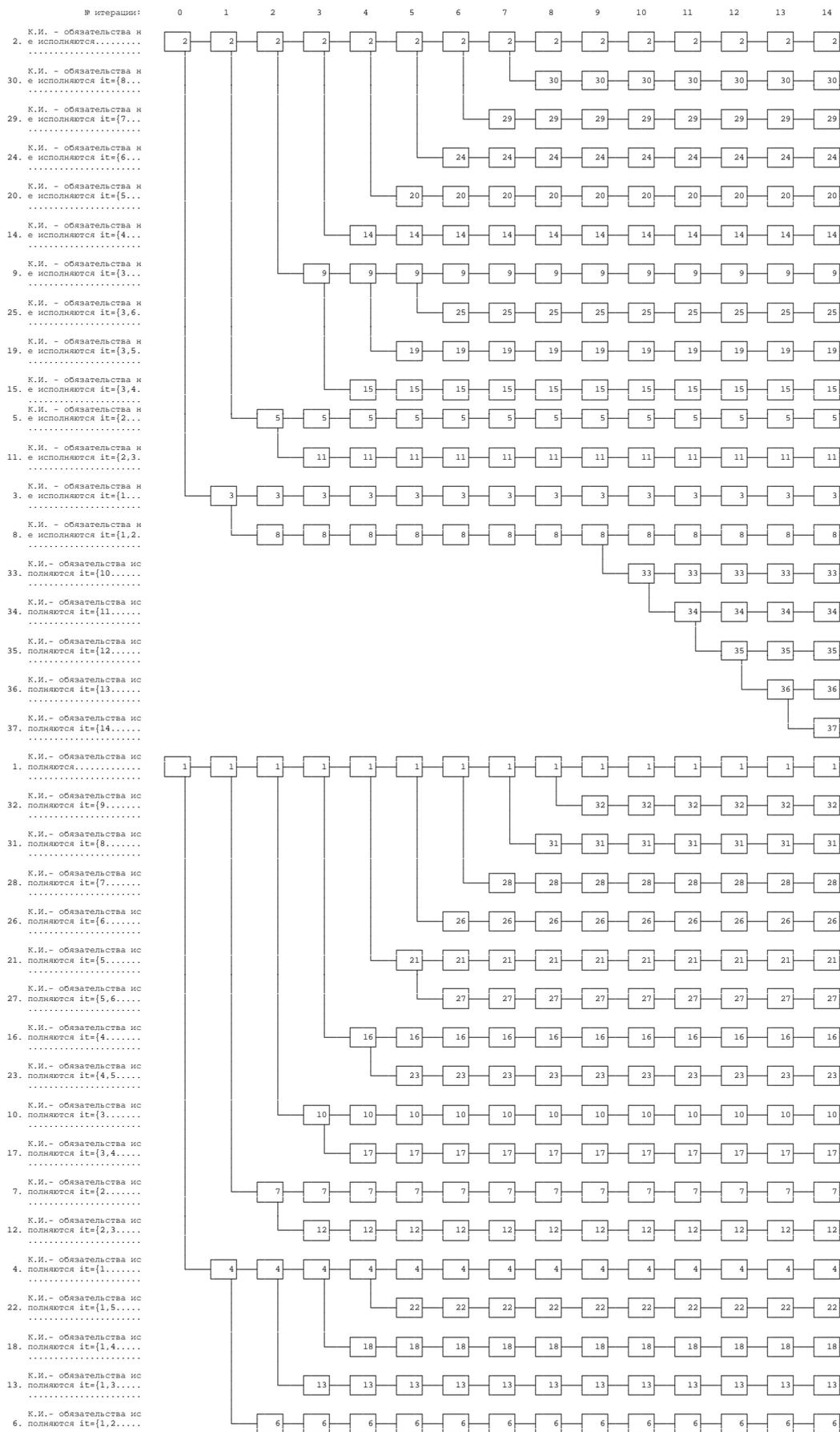


Рисунок 1. Дерево разделения классов

анкет из старых классов модели. Данная процедура должна быть проделана до полной идентификации распознаваемой выборки. После каждого разделения классов для измерения внутренней валидности создается новая итерация модели. Процесс разделения классов показан на рисунке 1. Для решения поставленной задачи процесс оптимизации (разделения классов) был повторен 14 раз результатом чего стало увеличение количества классов с 2-х до 37-и. Результаты оптимизации показаны на рисунке 2. Полученный результат является приемлемым для решения задачи прогнозирования будущих состояний объекта управления, т.к. позволяет производить верную идентификацию заемщиков входящих в обучающую выборку со 100% вероятностью. Также удовлетворительным можно считать процент ошибочной идентификации, который составляет 17,3%.

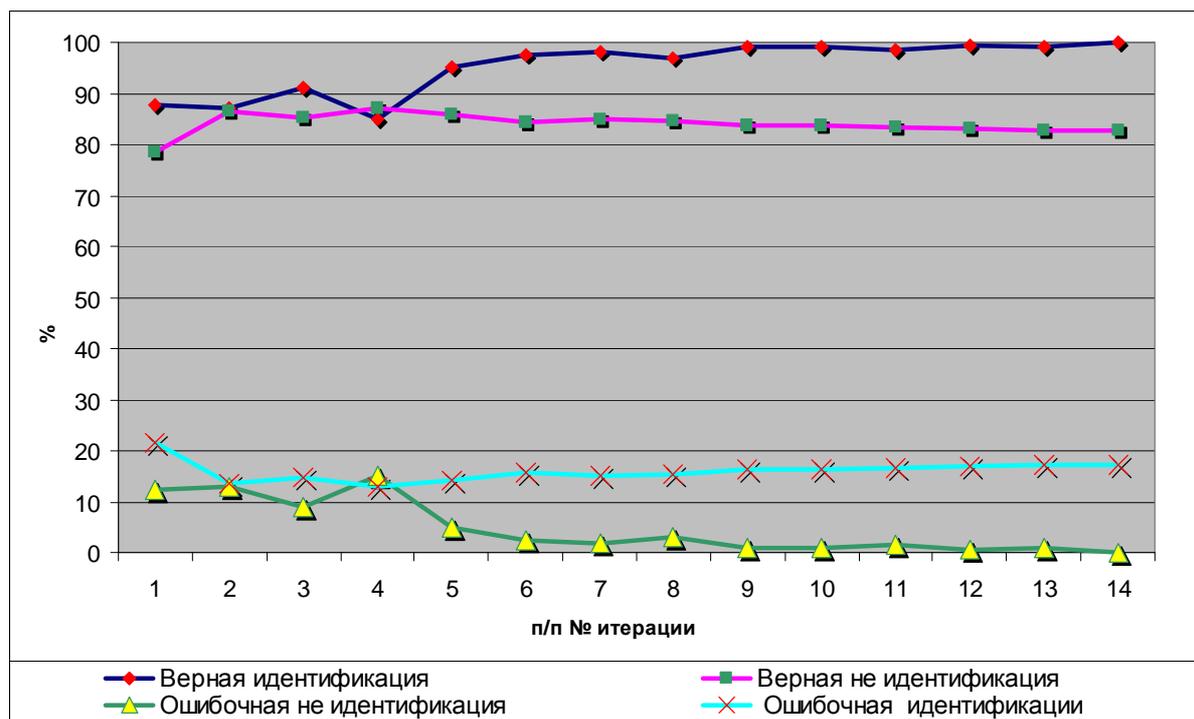


Рисунок 2. Изменение адекватности информационной модели в зависимости от итерации

Таким образом, можно сделать обоснованный вывод о том, что скоринговая семантическая информационная модель, созданная методом <http://ej.kubagro.ru/2007/05/pdf/14.pdf>

системно-когнитивного анализа, может быть с успехом применена для прогнозирования кредитной истории заемщиков не входящих в обучающую выборку. Полученные результаты будут опубликованы в следующих статьях.

Литература

1. Луценко Е.В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. –605с.

2. Лебедев Е.А. Оценка рисков кредитования физических лиц (проблема исследования, ее актуальность, идея решения) / Лебедев Е.А. // Научный журнал КубГАУ [Электронный ресурс]. - Краснодар: КубГАУ, 2006. - № 01(17). - Режим доступа: <http://ej.kubagro.ru/2006/01/13/p13.asp>.

3. Луценко Е.В. Теоретические основы и технология адаптивного семантического анализа в поддержке принятия решений (на примере универсальной автоматизированной системы распознавания образов "ЭЙДОС-5.1"). – Краснодар: КЮИ МВД РФ, 1996. – 280 с.

4. Луценко Е.В. Интеллектуальные информационные системы: Учебное пособие для студентов специальности: 351400 "Прикладная информатика (по отраслям)". – Краснодар: КубГАУ. 2004. – 633 с.

5. Луценко Е.В., Лебедев Е.А. Определение кредитоспособности физических лиц и риски их кредитования. – М.: Финансы и кредит, ноябрь 2006 – № 32(236).

6. Лебедев Е.А. Прогнозирование рисков кредитования физических лиц с применением системно-когнитивного анализа. Научное обеспечение агропромышленного комплекса: материалы 7-й региональной научно-практической конференции молодых ученых. - Краснодар: КубГАУ, 2005 - 450 с.