

### **Литература**

1. Нейлор К. Как построить свою экспертную систему. М., Энергоатомиздат, 1991. 286 с.
2. Романов, В.П. Интеллектуальные информационные системы в экономике. Под ред. Н.П. Тихомирова. М., Экзамен, 2003. 496 с.
- 3 Рыбина Г.В., Душкин Р.В., Демидов Д.В. Модели и методы обработки недостоверных знаний в инструментальном комплексе АТ-ТЕХНОЛОГИЯ. – В сб.: Труды II-го Междунар. НПС. Коломна, 2003, с. 401-407.
4. Жирабок А. Н. Нечеткие множества и их использование при принятии решений. – "Соросовский образовательный журнал", 2001, т. 7.
5. Кореневский Н.А., Крупчатников Р.А., Горбатенко С.А. Синтез нечетких сетевых моделей, обучаемых по структуре данных для медицинских экспертных систем. – "Медицинская техника", 2008, № 2, с. 18-24.

*Статья поступила 19.05.2016*

**К.т.н. Т.В. Зайцева, Р.М. Устинов, О.П. Пусная  
(НИУ «БелГУ»)**

**T.V. Zaitseva, R.M. Ustinov, O.P. Pusnaya**

**КОМПЬЮТЕРНАЯ РЕАЛИЗАЦИЯ АЛГОРИТМА ОБРАБОТКИ  
СТАТИСТИЧЕСКИХ ДАННЫХ С УЧЕТОМ ВЕРОЯТНОСТНОЙ  
НЕОПРЕДЕЛЕННОСТИ КЛАССИФИКАЦИИ**

**COMPUTER REALIZATION OF ALGORITHM OF PROCESSING  
OF THE STATISTICAL DATA TAKING INTO ACCOUNT  
LIKELIHOOD UNCERTAINTY OF CLASSIFICATION**

*В статье рассмотрена стратегия принятия решения с учетом вероятностной неопределенности классификации параметров. Представлена программная реализация алгоритма обработки статистических данных с учетом априорных условных вероятностей*

*The article discusses the strategy decision, taking into account the probabilistic uncertainty of classification parameters. Presented a software implementation of the algorithm of statistical data, taking into account a posteriori conditional probabilities*

*Ключевые слова: статистические данные, априорные условные вероятности, алгоритм*

## ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ НЕЙРОННЫЕ СЕТИ

*Key words statistical data, posterior conditional probabilities, algorithm*

Одной из важных задач в технологии инженерии знаний является построение продукционных правил. Существуют различные подходы к созданию продукционных правил для экспертных систем ЭС). Например, в статье «Компьютерная технология генерации правил для гибридных продукционно-фреймовых экспертных систем» [1] представлен алгоритм полного перебора всех параметров и полного покрытия предметной области. У данного подхода есть свои достоинства и недостатки. К достоинствам следует отнести то, что рассматриваются все возможные сочетания параметром и каждому набору ставится в соответствие решающее правило. Однако, некоторые сочетания параметров просто не имеют смысла для данной предметной области и рассматривать их нет никакой необходимости. Также к недостаткам следует отнести то, что при увеличении количества параметров и их значений количество правил возрастает. Например, есть два влияющих фактора  $\Phi_1$  и  $\Phi_2$ , каждый из которых имеет два значения, тогда количество решающих правил равно 4, а если факторов пять и каждый из них принимает 4 значения, то количество правил – 120. А если факторов десять?

В [2] описан один из методов Data Mining – метод деревьев решений, обеспечивающий автоматизацию построения правил на основе статистических данных с выбором параметра с минимальным значением энтропии. Рассмотрим пример, иллюстрирующий алгоритм построения продукционных правил на основе статистических данных с целью поддержки принятия решений о выдаче кредита.

Рассмотрим задачу определения кредитонадежности заемщика. База данных, на основе которой должно осуществляться прогнозирование, содержит следующие ретроспективные данные о клиентах банка, являющиеся ее атрибутами: пол, возраст, наличие недвижимости, образование, социальное положение, среднемесячный доход, вернул ли клиент вовремя кредит и т.д. В принципе, условия выдачи кредита в разных банках являются различными, однако, все вышеперечисленные атрибуты присутствуют в явном или неявном виде. В таблице приведен фрагмент базы данных одного из региональных банков за 2006 год. Данные разбивались следующим образом:

1. Пол С1 0 - женщина, 1 - мужчина
2. Возраст С2 до 30 лет и после 30 лет
3. Цель кредита С3 0 - НКД, 1 - ком деятельность
4. Образование С4 0 - среднее, 1 - высшее
5. Срок работы С5 до 5 лет и больше 5 лет
6. Брак С6 0 - не в браке 1 - в браке
7. Прописка С7 0 - нет, 1 - есть
8. Цена собственности С8 до 1млн и больше 1млн
9. Ежемесячный доход С9 до 15тыс. и больше 15 тыс.
10. Кредитная история С10 1 - положительная, 0 – отрицательная

Задача состоит в том, чтобы на основании перечисленных выше данных (кроме последнего атрибута) определить, стоит ли выдавать кредит новому клиенту. Такая задача решается в два этапа: построение классификационной модели и ее использование. На этапе построения модели строится дерево классификации или создается набор неких правил. На этапе использования модели построенное дерево или путь от его корня к одной из вершин, являющейся набором правил для конкретного клиента, используется для ответа на поставленный вопрос "Выдавать ли кредит?". Правилом является логическая конструкция, представленная в виде "если : то :" .

Качество построенного дерева решения зависит от правильного выбора критерия расщепления. Традиционно дерево решений строится, начиная с первого атрибута (т.е. в данном примере с пола клиента), не учитывая характер и силу влияния каждого атрибута. Более эффективным является подход, основанный на учете вероятностной неопределенности классификации. Другими словами, событие, состоящее в установлении соответствия между значениями цепочки атрибутов и определенным классом, является случайным и характеризуется некоторой вероятностью. При использовании нескольких атрибутов в качестве первого атрибута для анализа выбирается тот, который обеспечивает максимальное снижение неопределенности классификации по отношению к исходному множеству (т.е. минимальное значение энтропии).

С соответствии с предложенной методикой построения дерева решений начинается с атрибута, который больше всего уменьшает неопределенность (в рассмотренном примере - это факт возврата кредита). Далее по формуле Байеса находятся апостериорные

Таблица 1

Фрагмент базы данных

	C0	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	Класс
Клиент	Пол	Возраст	Цель кредита	Образование	Срок работы	Брак	Прописка	Собственность	Доход	Кредит история		
1 Я1	0	28	0	1	2,6	1	1	0	2,8	1		K2
2 Я2	1	54	1	1	7,6	1	1	0,5	21,2	0		K1
3 Я3	1	23	0	1	0,2	1	1	0	16,6	1		K1
4 Я4	0	20	1	1	2,6	1	1	0,8	22,6	1		K1
5 Я5	0	20	0	1	1,4	0	1	0	4,7	1		K1
6 Я6	0	29	0	0	1,3	0	0	0,1	19,7	0		K1
7 Я7	1	26	1	1	0,5	1	1	0	22,5	0		K1
79 Мт17	1	41	0	1	3,5	1	1	0,8	20	0		K1
380 Д10	0	29	1	1	2,2	1	1	0,6	11,3	1		K1

## ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ НЕЙРОННЫЕ СЕТИ

условные вероятности, которые будут использованы для построения дерева решений, а в дальнейшем и правил-продукций.

Далее отдельно рассматриваются те данные, записи которых соответствуют положительному значению рассмотренного атрибута, и данные, а затем записи, соответствующие отрицательному значению. Аналогично выбирается следующий из критериев расщепления дерева решений и т.д. По полученному дереву решений строится система производственных правил, например:

ЕСЛИ ( $C10 = \text{Да}$ ) И ( $C5 \leq 5$ ) И ( $C8 > 200$ ) И ( $C3 = \text{КД}$ ),  
ТО клиент К1 с вероятностью 100%.

ЕСЛИ ( $C10 = \text{Да}$ ) И ( $C5 \leq 5$ ) И ( $C8 > 200$ ) И ( $C3 = \text{НКД}$ ),  
ТО клиент К1 с вероятностью 86%; клиент К2 с  
вероятностью 14%. И т д.

Совокупность полученных правил-продукций после небольшой доработки преобразуется в законченную базу знаний и может быть использована в производственных или гибридных ЭС.

Программная поддержка алгоритма реализована в виде программных модулей:

1. получения дерева решений по статистическим данным;
2. создания производственных правил для ЭС.

В качестве инструментального средства для создания программной поддержки информационной технологии был выбран язык программирования C#.

Рассмотрим пользовательский интерфейс прототипа системы «Дерево решений». При запуске программной системы «Дерево решений» на экране монитора появляется диалоговая форма, представленная на рис. 1.

Меню содержит следующие пункты:

- Файл
  - Открыть базу
  - Сохранить отчёт
  - Сформировать правила
  - Печать отчёта
  - Закрыть базу
  - Выход
- Действия
  - Сделать шаг
  - Открыть отчёт

## ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ НЕЙРОННЫЕ СЕТИ

- Справка
  - Вызов справки
  - О программе

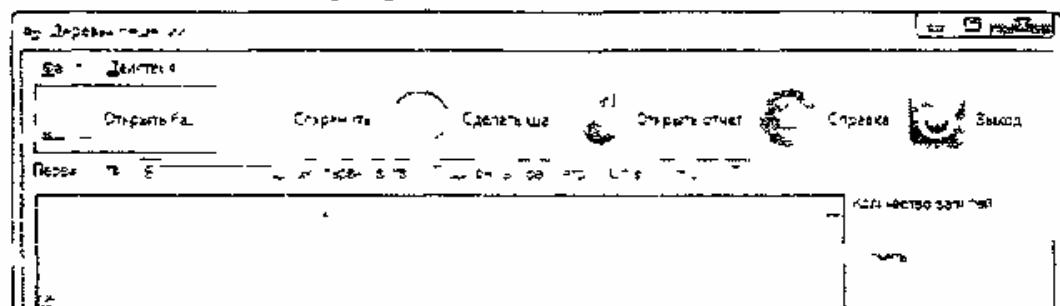


Рис 1  
Интерфейс программы

После загрузки базы данных, программная система принимает вид, отображенный на рис. 2.

Запрос	Помощь	Приоритет	Код	Статус	Клиент	Компания
601	0	1	1	1	-	-
602	0	1	2	0	1	1
603	0	1	2	1	1	1
604	0	1	2	1	1	1
605	0	1	2	1	1	1
606	0	1	2	1	1	1
607	0	1	2	1	1	1
608	0	1	2	1	1	1
609	0	1	2	1	1	1
610	0	1	2	1	1	1
611	0	1	2	1	1	1
612	0	1	2	1	1	1
613	0	1	2	1	1	1
614	0	1	2	1	1	1
615	0	1	2	1	1	1
616	0	1	2	1	1	1
617	0	1	2	1	1	1
618	0	1	2	1	1	1
619	0	1	2	1	1	1
620	0	1	2	1	1	1
621	0	1	2	1	1	1
622	0	1	2	1	1	1
623	0	1	2	1	1	1
624	0	1	2	1	1	1
625	0	1	2	1	1	1
626	0	1	2	1	1	1
627	0	1	2	1	1	1
628	0	1	2	1	1	1
629	0	1	2	1	1	1
630	0	1	2	1	1	1
631	0	1	2	1	1	1
632	0	1	2	1	1	1
633	0	1	2	1	1	1

Рис. 2  
Открытая база данных

В поле «Количество записей» выдается общее число записей в базе данных. Есть возможность упорядочить данные по одному из полей (атрибутов). Далее выбирая пункт меню или, используя кнопку на панели инструментов «Сделать шаг», вычисляем энтропии для каждого атрибута, находим минимальное значение энтропии и соответствующий ему номер атрибута (рис. 3). Построение дерева решений начнем именного с этого атрибута.

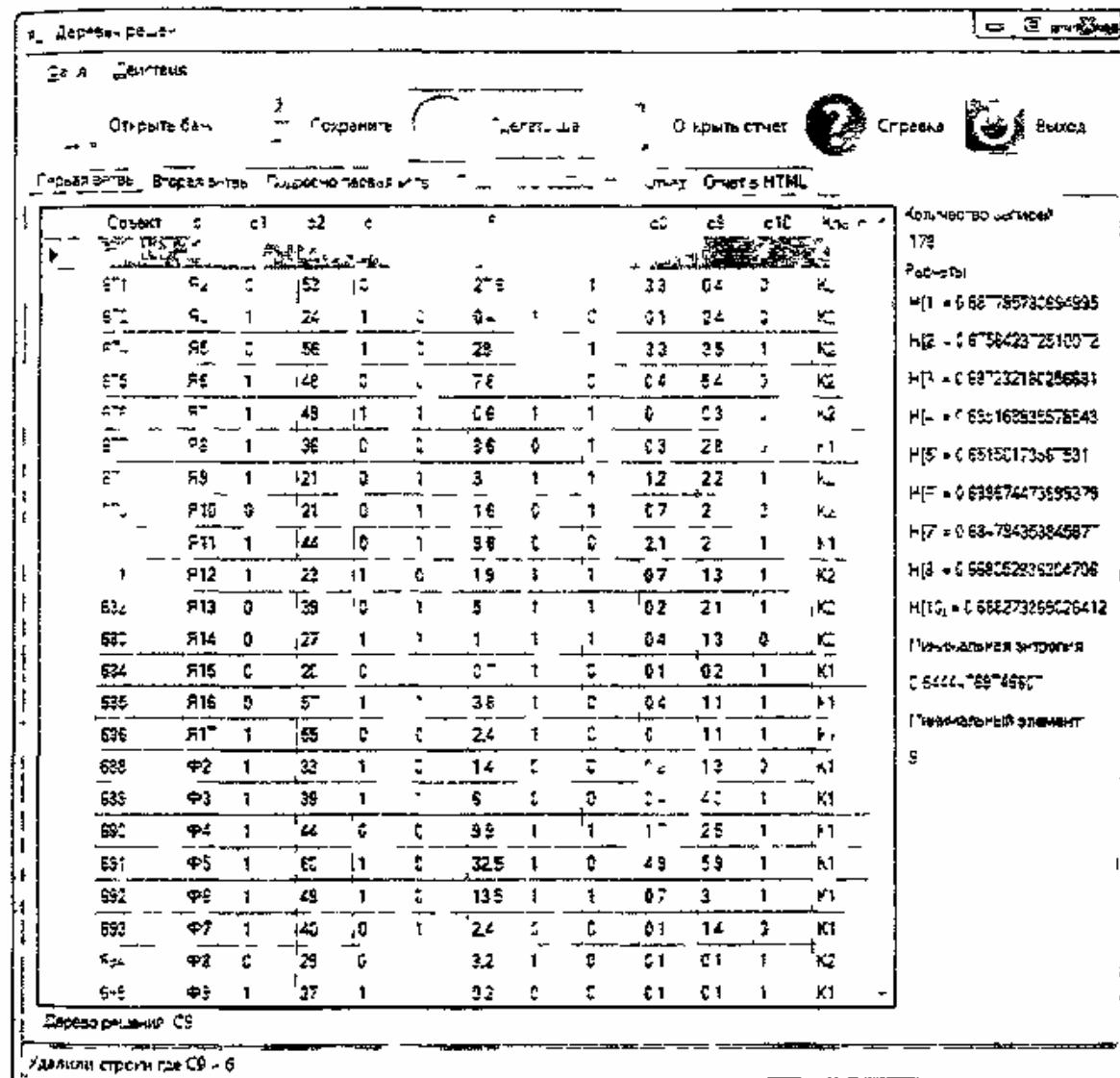


Рис. 3  
Первый шаг построения дерева решений

Далее по формуле Байеса находим апостериорные условные вероятности при условии, что атрибут C9 принял одно из двух значений в данной ветке C9<6. Выбираем минимальную энтропию и

## ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ НЕЙРОННЫЕ СЕТИ

находим номер атрибута, которому соответствует минимальная энтропия. Вторая ветвь ( $C9>=6$ ) приводит к готовому производственному правилу Построение дерева решений происходит с расчётом всех возможных комбинаций. Во вкладке «Вторая ветвь» выводятся удалённые строки из базы данных, и по ним производится параллельный расчёт. Аналогично первой ветке во второй происходит расчёт энтропии и нахождение минимального элемента. Количественные характеристики и энтропии по атрибутам можно просмотреть во вкладке «Первая ветвь подробно» (рис. 4). Подробные расчёты второй ветки дерева можно посмотреть, открыв вкладку «Вторая ветвь подробно».

Дерево решений																																																																																																													
Файл		Действия																																																																																																											
Открыть	Сохранить	Сделать шаг		Открыть/закрыть		Сп. меню	Выход																																																																																																						
<b>Первая ветвь Вторая ветвь Годробно первая ветвь Годробно вторая ветвь</b>																																																																																																													
<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>Номер строки</th> <th>Число первых K1</th> <th>Число первых &lt;2</th> <th>Число строк K2</th> <th>Число строк &lt;2</th> <th>Число строк K3</th> <th>Число строк &lt;2</th> <th>Число строк K4</th> <th>Число строк &lt;2</th> <th>Число строк K5</th> </tr> </thead> <tbody> <tr><td>1</td><td>27</td><td>13</td><td>1</td><td>1</td><td>59</td><td>54</td><td>35</td><td>53</td><td></td></tr> <tr><td>2</td><td>55</td><td>42</td><td>53</td><td>42</td><td>52</td><td>53</td><td>25</td><td>53</td><td></td></tr> <tr><td>3</td><td>55</td><td>41</td><td>55</td><td>52</td><td>51</td><td>39</td><td>25</td><td>53</td><td></td></tr> <tr><td>4</td><td>115</td><td>42</td><td>57</td><td>57</td><td>27</td><td>26</td><td>24</td><td>53</td><td></td></tr> <tr><td>5</td><td>82</td><td>40</td><td>80</td><td>79</td><td>-</td><td>51</td><td>23</td><td>53</td><td></td></tr> <tr><td>6</td><td>21</td><td>33</td><td>45</td><td>57</td><td>61</td><td>45</td><td>23</td><td>53</td><td></td></tr> <tr><td>7</td><td>141</td><td>68</td><td>77</td><td>77</td><td>13</td><td>17</td><td>11</td><td>53</td><td></td></tr> <tr><td>8</td><td>17</td><td>24</td><td>-</td><td>1</td><td>1</td><td>6</td><td>-</td><td>53</td><td></td></tr> <tr><td>9</td><td>55</td><td>43</td><td>-</td><td>52</td><td>27</td><td>48</td><td>-</td><td>53</td><td></td></tr> </tbody> </table>										Номер строки	Число первых K1	Число первых <2	Число строк K2	Число строк <2	Число строк K3	Число строк <2	Число строк K4	Число строк <2	Число строк K5	1	27	13	1	1	59	54	35	53		2	55	42	53	42	52	53	25	53		3	55	41	55	52	51	39	25	53		4	115	42	57	57	27	26	24	53		5	82	40	80	79	-	51	23	53		6	21	33	45	57	61	45	23	53		7	141	68	77	77	13	17	11	53		8	17	24	-	1	1	6	-	53		9	55	43	-	52	27	48	-	53	
Номер строки	Число первых K1	Число первых <2	Число строк K2	Число строк <2	Число строк K3	Число строк <2	Число строк K4	Число строк <2	Число строк K5																																																																																																				
1	27	13	1	1	59	54	35	53																																																																																																					
2	55	42	53	42	52	53	25	53																																																																																																					
3	55	41	55	52	51	39	25	53																																																																																																					
4	115	42	57	57	27	26	24	53																																																																																																					
5	82	40	80	79	-	51	23	53																																																																																																					
6	21	33	45	57	61	45	23	53																																																																																																					
7	141	68	77	77	13	17	11	53																																																																																																					
8	17	24	-	1	1	6	-	53																																																																																																					
9	55	43	-	52	27	48	-	53																																																																																																					
$H_1 = 112/223 * (80/114 * \ln(80/114) + 56/14 * \ln(56/14) + 100/23 * \ln(100/23) + 42/103 * \ln(42/103) + 103/47 * \ln(103/47) + 103/1 - 0.68775772895 - 0.5$ $H_2 = 69/223 * (26/62 * \ln(26/62) + 33/66 * \ln(33/66) + 15/22 * \ln(15/22) + 53/155 * \ln(53/155) + 92/156 * \ln(92/156) + 15/1 - 0.6758427251472$ $H_3 = 126/223 * (61/106 * \ln(61/106) + 45/106 * \ln(45/106) + 117/223 * \ln(117/223) + 51/117 * \ln(51/117) + 55/117 * \ln(55/117) + 1.68723216220451$ $H_4 = 122/223 * (61/120 * \ln(61/120) + 98/120 * \ln(98/120) + 103/223 * \ln(103/223) + 61/103 * \ln(61/103) + 103/42 * \ln(103/42) + 1.625162235572543$ $H_5 = 117/223 * (45/117 * \ln(45/117) + 68/117 * \ln(68/117) + 108/223 * (73/106 * \ln(73/106) + 32/106 * \ln(32/106)) + 0.651501735673_1$ $H_6 = 105/223 * (57/105 * \ln(57/105) + 46/45 * \ln(46/45) + 112/223 * (65/118 * \ln(65/118) + 53/118 * \ln(53/118) + 52/118 * \ln(52/118) + 0.628577356927_6$ $H_7 = 134/223 * (62/104 * \ln(62/104) + 52/104 * \ln(52/104) + 113/223 * (77/119 * \ln(77/119) + 46/119 * \ln(46/119) + 0.684784353925_0$ $H_8 = 143/223 * (51/143 * \ln(51/143) + 73/143 * \ln(73/143) + 74/223 * \ln(74/223) + 51/74 * \ln(51/74) + 23/71 * \ln(23/71) + 0.66825230631_706$ $H_9 = 127/223 * (84/177 * \ln(84/177) + 53/177 * \ln(53/177) + 48/223 * \ln(48/223) + 38/48 * \ln(38/48) + 8/48 * \ln(8/48) + 0.64447857468_7$ $H_{10} = 114/223 * (64/114 * \ln(64/114) + 50/114 * \ln(50/114) + 103/223 * 59/103 * \ln(59/103) + 51/103 * \ln(51/103) + 0.6582373865025_12$																																																																																																													
Удалить строки где $C9 > 6$																																																																																																													

Рис. 4  
Подробные расчёты энтропии первой ветки дерева

## ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ НЕЙРОННЫЕ СЕТИ

Ветвь с C9<6 разбиваем на подмножества по следующему выбранному атрибуту. Критерием выбора атрибута, по которому должно пойти разбиение соответствующего подмножества, является минимальная энтропия. Далее шаги повторяются до тех пор, пока не получим вершину, для которой апостериорная вероятность принадлежности объекта к определенному классу равна единице. На последнем шаге можно увидеть атрибуты, влияющие на построение дерева решений, их порядок и значения энтропии. После выполнения последнего шага можно посмотреть готовое дерево решений в приложении на вкладке «Отчёт» или во внешнем браузере, выбрав пункт меню «Открыть отчёт». Результаты построения дерева показаны на рис. 5. Для проверки работоспособности и эффективности разработанного алгоритма было проведено сравнение результатов классификации с использованием разных алгоритмов.

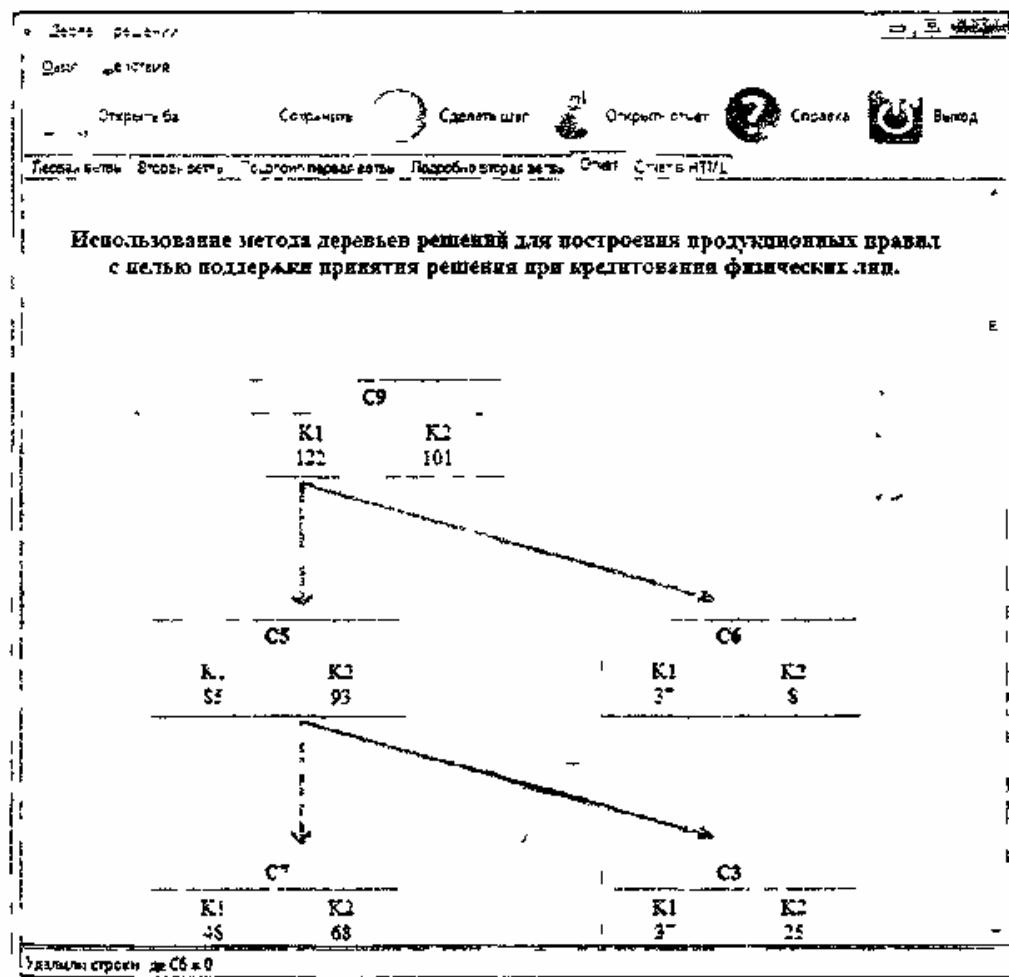


Рис. 5  
Дерево решений

## ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ НЕЙРОННЫЕ СЕТИ

После построения дерева решений можно формировать производственные правила. Выбирая пункт меню «Сформировать правила» (рис. 6), формируем производственные правила вида IF () AND () AND () ... AND () THEN () .

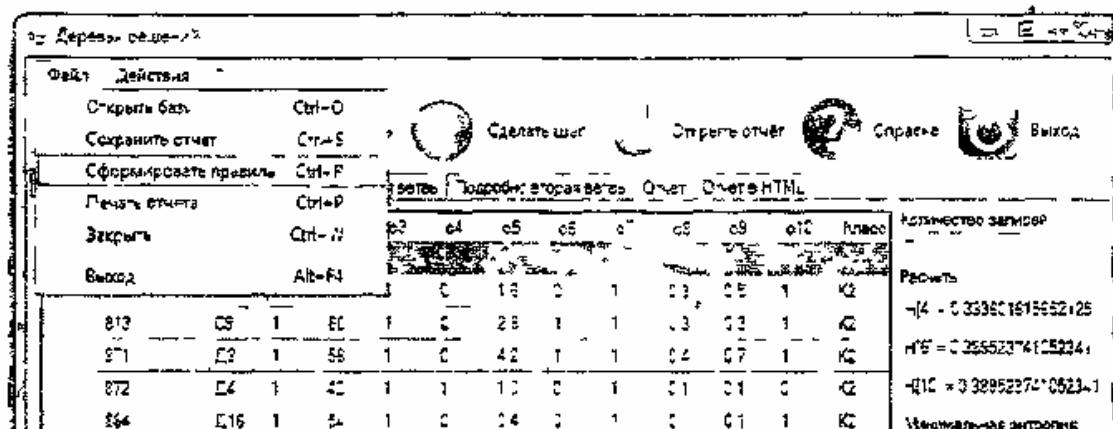


Рис.6  
Формирование производственных правил

Сформированные правила автоматически сохраняются в текстовом формате (рис. 7).

```
IF C9<6 AND C5<5 AND C7=1 AND C8<1 AND C1=1 AND C2>30 AND C3=1 AND C4=1 THEN Выдать кредит FS87
IF C9<6 AND C5<5 AND C7=1 AND C8<1 AND C1=1 AND C2>30 AND C3=1 AND C4=0 THEN Выдать кредит FS63
IF C9<6 AND C5<5 AND C7=1 AND C8<1 AND C1=1 AND C2>30 AND C3=0 THEN Выдать кредит FS57
IF C9<6 AND C5<5 AND C7=1 AND C8<1 AND C1=1 AND C2<30 THEN Не выдать кредит FS43
IF C9<6 AND C5<5 AND C7=1 AND C8<1 AND C1=0 THEN Не выдать кредит FS63
IF C9<6 AND C5<5 AND C7=1 AND C8>1 THEN Не выдать кредит FS73
IF C9<6 AND C5<5 AND C7=0 THEN Не выдать кредит FS78
IF C9<6 AND C5>5 THEN Не выдать кредит FS84
IF C9>6 THEN Не выдать кредит FS100
```

Рис. 7  
Текстовый файл с сформированными правилами

Сравнение решений, принятых сотрудниками банка, с результатом, полученным с применением предлагаемого алгоритма, проводилось на примере базы данных одного из региональных банков за 2002-2006 гг. (бралась выборка в размере 140 записей в каж-

## ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ НЕЙРОННЫЕ СЕТИ

дом году). Все записи в базе данных (отдельно по годам и в целом за 5 лет) были разбиты на две группы:

обучающая и тестовая.

Следует отметить, что анализ данных показал влияние человеческого фактора при принятии решения о выдаче кредита. В базе данных из 700 записей было отмечено 96, когда было принято неправильное решение:

- 83 записи о выданных кредитах, которые в последствии не были возвращены;
- 13 записей о не выданных кредитах заемщикам, которые удовлетворяют всем критериям для выдачи.

То есть работники в среднем ошибаются в 14 случаях из 100 при принятии решений.

Предлагаемая модель, в среднем, выдает 7-8 неправильных рекомендаций из 140 за год по вопросу выдачи кредитов. При этом, как показал анализ, 4-6 рекомендаций относятся к упущеной выгоде банка (программа рекомендует не выдавать кредит благонадежному заемщику) и только 2-3 рекомендации можно отнести к потери банком собственных денежных средств (программа рекомендует выдавать кредит, при этом он не будет возвращен). При рассмотрении дерева решений с количеством ветвей больше двух процент ошибок возрастает незначительно, максимальная ошибка составила 7,2%.

В статье рассматривается один из методов Data Mining – метод деревьев решений, обеспечивающий автоматизацию построения правил на основе статистических данных. Приводится пример, иллюстрирующий алгоритм построения продукционных правил на основе статистических данных с целью поддержки принятия решений о выдаче кредита

В заключение следует отметить, что использование вариационных алгоритмов в задачах классификации является весьма актуальным в связи с постоянным ростом вычислительной мощности компьютеров. Такого рода алгоритмы позволяют добиться хороших (адекватных) результатов, но в связи с большой долей эвристики исследование их свойств сильно затрудняется. Таким образом, имеет смысл продолжать исследования в этом направлении и создавать новые алгоритмы, использующие вариационный принцип, которые будут более универсальными и адекватными.

### Литература

1. Зайцева Т.В., Игрунова С.В., Путивцева Н.П., Пусная О.П., Манзуланич М.Ю. Компьютерная технология генерации правил для гибридных производственно-фреймовых экспертных систем. – "Вопросы радиоэлектроники", сер. ЭВТ, 2011, вып. I, с. 105-114.
2. Матвеев М.Г., Свиридов А.С., Алейникова Н.А. Модели и методы искусственного интеллекта. Применение в экономике. Учеб. пособие. М., Финансы и статистика, 2008. 448 с.

*Статья поступила 19.05.2011*

**К.т.н., доц. М.Ф. Тубольцев, д.т.н., проф. С.И. Маторин,  
О.М. Тубольцева (НИУ «БелГУ»)**

**M.F. Tuboltsev, S.I. Matorin, O.M. Tuboltseva**

### **ПРОГРАММНЫЕ СРЕДСТВА СИСТЕМНО-ОБЪЕКТНОГО АНАЛИЗА ФИНАНСОВЫХ МОДЕЛЕЙ БИЗНЕС-ПРОЦЕССОВ**

### **SOFTWARE OF SYSTEM-OBJECT ANALYSIS OF BUSINESS-PROCESS FINANCIAL MODELS**

*В статье рассмотрены вопросы разработки финансовых моделей бизнес-процессов показаны недостатки традиционного системного подхода и преимущества системно-объектного анализа, а также необходимость применения системного метода агрегирования показателей доходности. Показаны возможности информационной поддержки методик финансового моделирования*

*The article deals with issues of development of business processes financial models identifying shortcomings of traditional system approach and advantages of the system-object analysis, as well as the need for a systematic method of aggregation rates of return. The possibilities of informational support of financial modeling techniques are shown*

*Ключевые слова* бизнес процесс, финансовая модель, системно-объектный анализ

*Key words* business process, financial model system-object analysis, UFO-toolkit, FinStream