

БУТЫРНОВА Татьяна Валерьяновна – кандидат экономических наук, доцент кафедры мировой экономики и налоговых систем. Чебоксарский кооперативный институт (филиал) Российского университета кооперации. Россия. Чебоксары. E-mail: btv2303@mail.ru.

BUTYRNOVA, Tatyana Valeryanovna – Candidate of Economics Sciences, Associate Professor of the World Economy and Tax Systems. Cheboksary Cooperative Institute (branch) of the Russian University of Cooperation. Russia. Cheboksary. E-mail: btv2303@mail.ru.

УДК 65.011.56

О ПОДДЕРЖКЕ ПРИНЯТИЯ РЕШЕНИЯ В ТЕХНОЛОГИИ DATA MINING

В.И. Возяков, В.П. Филиппов

Рассмотрено применение технологии Data Mining для кластерного и дискриминантного анализа инвестиционной привлекательности предприятий потребительской кооперации.

Ключевые слова: Data Mining; классификация; кластеризация; дискриминантный анализ; кластерный анализ.

V.I. Vozyakov, V.P. Filippov. ABOUT DECISION-MAKING SUPPORT IN DATA MINING TECHNOLOGIES

The application of data mining technology for cluster and discriminant analysis of investment attractiveness of Consumer Co-operatives is considered in this article.

Keywords: Data Mining; classification; clustering; discriminant analysis; cluster analysis.

В настоящее время деятельность любого предприятия (коммерческого, производственного, научного и т.д.) сопровождается регистрацией и записью всех подробностей его функционирования. Эффективное использование этой информации связано с продуктивной переработкой данных. Специфика современных требований к такой переработке следующая:

- данные являются разнородными (количественными, качественными, текстовыми);
- результаты должны быть конкретны и понятны;
- инструменты для обработки данных должны быть просты в использовании.

Традиционная математическая статистика при обработке данных использует концепцию усреднения по выборке. Практика показывает, что эта процедура недостаточна для принятия решений в различных сферах человеческой деятельности, требующей обнаружения в данных ранее неизвестных, нетривиальных, практически полезных закономерностей. Исследовательский аспект изучения закономерностей в данной предметной области объектов и их отношений связан с использованием средств информационных и коммуникационных технологий.

В основу современной технологии Data Mining (discovery-driven data mining) положена концепция шаблонов, отражающих фрагменты

многоаспектных взаимоотношений в данных. Эти шаблоны представляют собой закономерности, свойственные подвыборкам данных, которые могут быть компактно выражены в понятной человеку форме. Поиск шаблонов производится методами, не ограниченными рамками априорных предположений о структуре выборки и виде распределений значений анализируемых показателей.

Data Mining является мультидисциплинарной областью, возникшей и развивающейся на базе достижений прикладной статистики, распознавания образов, методов искусственного интеллекта, теории баз данных и др. (рис. 1). Отсюда обилие методов и алгоритмов, реализованных в различных действующих системах Data Mining. Многие из таких систем интегрируют в себе сразу несколько подходов. Как правило, в каждой системе имеется какая-то ключевая компонента, на которую делается главная ставка.

Выделяют пять стандартных типов закономерностей, которые позволяют выявлять методы Data Mining: ассоциация, последовательность, классификация, кластеризация и прогнозирование.

Далее остановимся на классификационном анализе, включающем кластерный и дискриминантный анализ.

Кластерный анализ – это совокупность



Рис. 1. Data Mining — мультидисциплинарная область

многомерных статистических процедур, которая позволяет упорядочить объекты по однородным группам.

Решение задачи кластерного анализа выполняется в следующей последовательности:

1. Формируются выборки для анализа.
2. Выбирается совокупность признаков, характеризующих объект.
3. Выбираются меры сходства (расстояния) между объектами и производится их расчет.
4. Формируются кластеры.
5. Проводится анализ полученной информации.

Большинство алгоритмов кластерного анализа относят к так называемым агломеративным процедурам, которые сначала объединяют в группы самые близкие объекты, а затем к ним присоединяют более дальние.

Рассмотрим процедуру кластерного анализа, предлагаемую для обработки данных. Существуют две разновидности, которые могут давать различное разбиение на кластеры. Выбирать подходящую разновидность следует исходя из постановки задачи. Если это невозможно, необходимо провести разбиение двумя способами и попытаться определить, какой из них более соответствует фактически существующим структурам данных. При изотоническом разбиении группы объектов состоят из однородных по уровню значений, а при изоморфном в группы включаются объекты, близкие по структуре, т.е. те, в которых пропорции признаков мало отличаются. Это означает, что различные способы разбиения могут давать различное объединение по группам. Например, у нас есть данные, которые характеризуют распределение прибыли фирм на расширение производства, научные исследования, социальные выплаты и пр. Тогда при изотоническом разбиении группы будут состоять из фирм, в которых уровни прибыли близки, а

при изоморфном — в однородные группы будут включаться те компании, в которых структуры распределения прибыли сходны.

В обоих способах признаки сначала преобразуют таким образом, чтобы не было единиц измерения и размах шкалы был одинаковым.

Для нормирования шкал необходимо выполнить следующие преобразования. При изотоническом разбиении сначала каждое значение признака заменяется на вычисленное по формуле:

$$V_{ij} = x_{ij} / \sum_{i=1}^n x_{ij},$$

где x_{ij} — значение j -го признака для i -го объекта.

После этого каждому объекту ставится в соответствие одно число, вычисленное по формуле:

$$\omega_i = \sum_{j=1}^m V_{ij}.$$

Расстояния между двумя объектами определяют по формуле:

$$d_{ij} = |\omega_i - \omega_j|.$$

При изоморфном разбиении сначала выполняют нормирование шкал по формуле:

$$Z_{ij} = V_{ij} / \sum_{j=1}^m V_{ij}.$$

Расстояния между двумя объектами определяют по формуле:

$$d_{ik} = \sqrt{\sum_{j=1}^m (Z_{ij} - Z_{ik})^2}.$$

В изоморфном преобразовании расстояние будет минимальным в том случае, когда векторы коллинеарны, и максимальным, если они перпендикулярны.

После определения расстояний возможно разбиение на группы с помощью метода ближайшего соседа. Метод заключается в следующем: сначала для первого объекта находится ближайший и объединяется с ним в один кластер. Затем

для второго объекта ищется ближайший сосед и объединяется с ним в другой кластер и т.д. В случае, если ближайший сосед уже находится в одном из кластеров, происходит присоединение исследуемого объекта к нему. Таким образом образуются начальные кластеры.

На следующем этапе обычно выполняют построение дендритов и определение связности в системе кластеров. Это позволяет объединить первоначальные кластеры в более сложные структуры, которые в большей степени отвечают их реальной форме.

Дискриминантный анализ относится к методам классификаций многомерных наблюдений при наличии обучающих выборок (в отличие от кластерного анализа, осуществляющего классификацию автоматически – без обучения). Его цель состоит в идентификации новых объектов и отнесении их к уже имеющимся группам или совокупностям.

Пусть имеется n наблюдений, характеризующих набор из k признаков. Тогда каждое наблюдение представляет собой случайный вектор $x = (x_1, x_2, \dots, x_k)^T$. Задача дискриминации состоит в разбиении всего множества реализаций рассматриваемой многомерной величины на некоторое число групп (областей) R_i ($i=1, 2, \dots, l$) и последующем отнесении нового наблюдения к одной из них, используя некое решающее правило. При этом информация об истинной принадлежности объекта считается недоступной или требует чрезмерных материальных и временных затрат.

В ходе процедуры автоматически вычисляются функции классификации, предназначенные для определения той группы, к которой наиболее вероятно принадлежит новый объект. Количество функций классификации равно числу имеющихся групп. Наблюдение считается принадлежащим той группе (совокупности), для которой получено наивысшее значение функции классификации или наивысшее значение апостериорной вероятности – вероятности, с которой новое наблюдение принадлежит к этому классу, вычисленное на основе априорной вероятности и расстояния Махаланобиса. Расстоянием махаланобисского типа называют естественную меру отдаленности двух объектов друг от друга, если наблюдения извлекаются из нормальных генеральных совокупностей с одной и той же матрицей ковариации [4].

Один из методов автоматизированного классификационного анализа данных основан на использовании программной среды Excel [4]. Достоинством такого подхода является возможность получения искомого результата

без дорогостоящих специализированных программ для обработки данных.

Далее в качестве примера проведем классификационный анализ прогноза развития потребительской кооперации на 20хх год (таблица) с использованием программ для кластерного (Приложение 1) и дискриминантного анализа (Приложение 2).

Пример кластерного анализа. Первым этапом работы программы является вычисление расстояний между объектами (рис. 5).

На следующем этапе, с использованием найденных расстояний, происходит объединение объектов в кластеры (рис. 2):

	A	B	C	D	E	F	G	H
46	Цепочка	Архангельский	0,026227	Вологодский				
47								
48	Цепочка	Мурманский	0,094106	Коми				
49								
50	Цепочка	Карельский	0,077471	Ивановский				
51								
52	Цепочка	Ленинградский	0,048147	Псковский	0,0198	Тверской	0,1163	Тульский
53								
54	Цепочка	Калужский	0,004976	Ярославский				
55								
56	Цепочка	Кировский	0,063666	Удмуртский	0,0917	Пермский		
57								
58	Цепочка	Чувашский	0,035664	Татарский	0,5031	Башкирский		

Рис. 2. Начальные кластеры

При наличии большого числа показателей или более сложной структуры данных возможно образование большого числа кластеров. Путем расчета межкластерных расстояний возможно их дальнейшее объединение в более крупные структуры (рис. 3).

Таким образом, исходя из рис. 3. с учетом значений, приведенных на рис. 2, получаем следующие кластеры:

1. Архангельский, Вологодский, Мурманский, Карельский, Коми, Ярославский, Калужский, Ивановский.
2. Ленинградский, Псковский, Тверской, Тульский.
3. Кировский, Чувашский, Пермский, Башкирский, Удмуртский, Татарский.

Программа позволяет провести анализ влияния показателей деятельности потребительских союзов на состав получаемых кластеров (аналог факторного анализа).

Пример дискриминантного анализа. Рассмотрим пример классификации Ивановского и Татарского потребсоюзов при наличии обучающей выборки из предыдущего раздела. Заполним рабочий лист Microsoft Excel согласно рис. 6. Столбец *Группа принадлежности наблюдения* относит наблюдения к тому или иному кластеру (использовано деление на кластеры, полученное в кластерном анализе).

Основные показатели прогноза развития потребительской кооперации на 20xx год

№ п/п	Наименование потребительских союзов	Весь розничный товароборот		Объем бытовых услуг, млн руб.	Производство основных видов промышленной продукции						Закупки основных видов сельскохозяйственной продукции					
		млн руб.	в т. ч. оборот общественного питания, млн руб.		хлеб, тыс. т	колбасные изделия, т	кондитерские изделия, т	консервы, туб	безалкогольные напитки, тыс. дкл	пиво, тыс. дкл	Мясо, т	Молоко т	Яйца, млн шт.	Картофель, т	Овощи, т	Фрукты, т
1	Архангельский	1900	130	7	18,7	610	900	130	20	12	1800	1000	11,7	5000	1830	650
2	Вологодский	2550	140	5	22,8	600	1100	15	110		2400	4000		8000	1000	80
3	Мурманский	145	12	1	1,3		15				110	40	0,9	700	540	100
4	Карельский	630	30	2	3,5	250	200		8		440	500	40	1000	400	380
5	Коми	530	28	2	2,4		75		7		500	1500	2,5	5800	620	100
6	Ленинградский	1900	150	1,5	4	1100	300	9500	10		5000	2000	12	5000	2000	1100
7	Псковский	3150	225	10	32	160	1600		240		2300	7000	23	8400	2300	1200
8	Тверской	1950	150	6	21	240	1200	1000	300	45	2000	8000	18	8000	4000	800
9	Калужский	860	68	7	12	410	600	170	40		2200	4000	5	7500	2400	520
10	Тульский	850	45	3	12,5	1000	400	7750	60		3750	3000	8	8000	4500	950
11	Ярославский	1600	95	6	8		800		8		2200	2000	11	3000	3000	1300
12	Кировский	3800	450	8	44	730	1730	700	400	33	3000	5000	22	8000	4000	1100
13	Чувашский	2550	310	10	23,5	2000	2700	1900	920	12	6300	20000	12,1	11000	5000	1200
14	Пермский	1850	115	5	25	140	900	3900	100	200	3000	5000	5,9	17000	8000	450
15	Башкирский	3000	300	20	33	1000	1600	1500	600	60	11000	25000	15	20000	5000	8000
16	Удмуртский	2150	240	15	22	1500	3000	3000	560		6200	13000	10,5	13500	3500	250
17	Татарский	3100	300	29	40	1500	1900	1900	510		6500	20000	15	19000	5000	1000
18	Ивановский	560	30	1,5	6,7		200	2000	30		900	1000	4,5	1000	1000	700

	A	B	C	D	E	F	G	H
60	Матрица расстояний между начальными кластерами							
61		Цепочка1	Цепочка2	Цепочка3	Цепочка4	Цепочка5	Цепочка6	Цепочка7
62	Цепочка1	0	0,378731	0,201162	0,2206	0,0815	0,6632	0,9785
63		(0; 0)	(2; 5)	(2; 4)	(1; 10)	(2; 11)	(1; 12)	(1; 13)
64	Цепочка2	0,378731	0	0,100098	0,6256	0,2922	1	1
65		(2; 5)	(0; 0)	(5; 18)	(5; 10)	(5; 9)	(0; 0)	(0; 0)
66	Цепочка3	0,201162	0,100098	0	0,448	0,1146	0,8906	1
67		(2; 4)	(5; 18)	(0; 0)	(4; 10)	(4; 9)	(4; 12)	(0; 0)
68	Цепочка4	0,220622	0,62558	0,448011	0	0,3284	0,3262	0,6415
69		(1; 10)	(5; 10)	(4; 10)	(0; 0)	(10; 11)	(8; 12)	(8; 13)
70	Цепочка5	0,081548	0,292208	0,114638	0,3284	0	0,771	1
71		(2; 11)	(5; 9)	(4; 9)	(10; 11)	(0; 0)	(11; 12)	(0; 0)
72	Цепочка6	0,663198	1	0,890588	0,3262	0,771	0	0,16
73		(1; 12)	(0; 0)	(4; 12)	(8; 12)	(11; 12)	(0; 0)	(14; 13)
74	Цепочка7	0,978507	1	1	0,6415	1	0,16	0
75		(1; 13)	(0; 0)	(0; 0)	(8; 13)	(0; 0)	(14; 13)	(0; 0)
76								
77	Мин. расс.	0,081548	0,100098	0,100098	0,2206	0,0815	0,16	0,16
78	Между це	(1; 5)	(2; 3)	(3; 2)	(4; 1)	(5; 1)	(6; 7)	(7; 6)

Рис. 3. Матрица расстояний между начальными кластерами

В результате вычислений с использованием программы (рис. 4) получаем, что Ивановский потребительский союз относится к первой группе,

а Татарский – к третьей. Данный результат соответствует полученному при кластерном анализе, а также согласуется с выводами работы, выполненной в программной среде Statistica.

	B	C	D	E	F
25	Результаты дискриминантного анализа				
26	Имя	объем	Макс. з	Класс	
27	Татарский	420,3		3	
28	Ивановский	3,921		1	

Рис. 4. Результат работы программы дискриминантного анализа

В заключение отметим, что настоящая схема классификации может быть использована для анализа инвестиционной привлекательности предприятий потребительской кооперации.

	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK
25	Матрица расстояний между объектами																		
26	Арх Волс Мур Каре Коми Ленг Псков Тверс Калу Туль Ярос Курс Чувс Перм Баш Удм Татм Иван																		
27	Архангельск	0	0,03	0,50	0,23	0,40	0,27	0,32	0,34	0,11	0,22	0,11	0,66	0,98	0,82	1,52	0,73	1,01	0,30
28	Вологодский	0,03	0	0,47	0,20	0,38	0,30	0,34	0,36	0,09	0,25	0,08	0,69	1,00	0,84	1,54	0,75	1,04	0,28
29	Мурманский	0,50	0,47	0	0,27	0,09	0,77	0,82	0,84	0,39	0,72	0,39	1,16	1,48	1,32	2,02	1,23	1,51	0,19
30	Карельский	0,23	0,20	0,27	0	0,18	0,50	0,54	0,56	0,11	0,45	0,12	0,89	1,21	1,05	1,74	0,95	1,24	0,08
31	Коми	0,40	0,38	0,09	0,18	0	0,67	0,72	0,74	0,29	0,63	0,30	1,07	1,38	1,22	1,92	1,13	1,42	0,10
32	Ленинградский	0,27	0,30	0,77	0,50	0,67	0	0,05	0,07	0,38	0,05	0,38	0,39	0,71	0,55	1,25	0,46	0,75	0,57
33	Псковский	0,32	0,34	0,82	0,54	0,72	0,05	0	0,02	0,43	0,10	0,42	0,35	0,66	0,50	1,20	0,41	0,70	0,62
34	Тверской	0,34	0,36	0,84	0,56	0,74	0,07	0,02	0	0,45	0,12	0,44	0,33	0,64	0,48	1,18	0,39	0,68	0,64
35	Калужский	0,11	0,09	0,39	0,11	0,29	0,38	0,43	0,45	0	0,33	0,00	0,78	1,09	0,93	1,63	0,84	1,13	0,19
36	Тульский	0,22	0,25	0,72	0,45	0,63	0,05	0,10	0,12	0,33	0	0,33	0,44	0,76	0,60	1,30	0,51	0,79	0,53
37	Ярославский	0,11	0,08	0,39	0,12	0,30	0,38	0,42	0,44	0,00	0,33	0	0,77	1,09	0,93	1,63	0,83	1,12	0,20
38	Кировский	0,66	0,69	1,16	0,89	1,07	0,39	0,35	0,33	0,78	0,44	0,77	0	0,32	0,16	0,85	0,06	0,35	0,97
39	Чувашский	0,98	1,00	1,48	1,21	1,38	0,71	0,66	0,64	1,09	0,76	1,09	0,32	0	0,16	0,54	0,25	0,04	1,28
40	Пермский	0,82	0,84	1,32	1,05	1,22	0,55	0,50	0,48	0,93	0,60	0,93	0,16	0,16	0	0,70	0,09	0,20	1,12
41	Башкирский	1,52	1,54	2,02	1,74	1,92	1,25	1,20	1,18	1,63	1,30	1,63	0,85	0,54	0,70	0	0,79	0,50	1,82
42	Удмуртский	0,73	0,75	1,23	0,95	1,13	0,46	0,41	0,39	0,84	0,51	0,83	0,06	0,25	0,09	0,79	0	0,29	1,03
43	Татарский	1,01	1,04	1,51	1,24	1,42	0,75	0,70	0,68	1,13	0,79	1,12	0,35	0,04	0,20	0,50	0,29	0	1,32
44	Ивановский	0,30	0,28	0,19	0,08	0,10	0,57	0,62	0,64	0,19	0,53	0,20	0,97	1,28	1,12	1,82	1,03	1,32	0
45																			

Рис. 5. Матрица расстояний между объектами

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Основные показатели прогноза развития потребительской кооперации на 20xx год																
2																	
3	Группа принадлежностей наблюдения	Наименование потребностей	Весь розничный млн. руб	в т. ч. оборот общественного питания, млн. руб	Объем бытовых услуг, млн. руб	Производство основных видов промышленной						Закупки основных видов					
4						хлеб, тыс., тонн	колбасные изделия, тонн	кондитерские изделия, тонн	консервы, туб	безалкогольные напитки, тыс. дкл	пиво, тыс. дкл	Мясо, тонн	Молоко тонн	Яйца, млн. штук	Картофель, тонн	Овощи тонн	Фрукты, тонн
5	1	Архангельск	1900	130	7	18,7	610	900	130	20	12	1800	1000	11,7	5000	1830	650
6	1	Вологодский	2550	140	5	22,8	600	1100	15	110		2400	4000		8000	1000	80
7	1	Мурманский	145	12	1	1,3		15				110	40	0,9	700	540	100
8	1	Карельский	630	30	2	3,5	250	200		8		440	500	40	1000	400	380
9	1	Коми	530	28	2	2,4		75		7		500	1500	2,5	5800	620	100
10	1	Ярославский	1600	95	6	8		800		8		2200	2000	11	3000	3000	1300
11	1	Калужский	860	68	7	12	410	600	170	40		2200	4000	5	7500	2400	520
12	2	Ленинградский	1900	150	1,5	4	1100	300	9500	10		5000	2000	12	5000	2000	1100
13	2	Псковский	3150	225	10	32	160	1600		240		2300	7000	23	8400	2300	1200
14	2	Тверской	1950	150	6	21	240	1200	1000	300	45	2000	8000	18	8000	4000	800
15	2	Тульский	850	45	3	12,5	1000	400	7750	60		3750	3000	8	8000	4500	950
16	3	Кировский	3800	450	8	44	730	1730	700	400	33	3000	5000	22	8000	4000	1100
17	3	Чувашский	2550	310	10	23,5	2000	2700	1900	920	12	6300	20000	12,1	11000	5000	1200
18	3	Пермский	1850	115	5	25	140	900	3900	100	200	3000	5000	5,9	17000	8000	450
19	3	Башкирский	3000	300	20	33	1000	1600	1500	600	60	11000	25000	15	20000	5000	8000
20	3	Удмуртский	2150	240	15	22	1500	3000	3000	560		6200	13000	10,5	13500	3500	2500
21		Татарский	3100	300	29	40	1500	1900	1900	510		6500	20000	15	19000	5000	10000
22		Ивановский	560	30	1,5	6,7		200	2000	30		900	1000	4,5	1000	1000	700

Рис. 6. Начальные данные для дискриминантного анализа

Список литературы

1. Возяков В.И., Филиппов В.П. Инструментальная поддержка управленческих решений в условиях экономического соперничества // Обзор прикладной и промышленной математики. 2008. Т. 15. № 3. С. 456-457.
2. Возяков В.И., Филиппов В.П. Модель рынка с прогнозируемыми ценами // Обзор прикладной и промышленной математики. 2007. Т. 14. № 4. С. 694-695.
3. Пичужкин А.Б., Романов Ю.А., Филиппов В.П. Об управлении бизнес-процессами торгового

предприятия // Вестник Российского университета кооперации. 2013. № 2 (12). С. 144-148.

4. Филиппов В.П., Возяков В.И. О поддержке принятия решения в технологии Data Mining // Обзор прикладной и промышленной математики. 2006. Т. 13, вып. 6. С. 1063-1064.

5. Филиппов В.П., Шульдяшева Е.О., Ярченков Н.А. Классификационный анализ деятельности потребительских организаций за 2008-2011 гг. // Актуальные вопросы науки и образования: сб. науч. ст. Чебоксары: ЧКИ РУК, 2013. С. 325-330.

Приложение 1

Программа для кластерного анализа

```

'Разбиение на кластеры методом изотонических расстояний
Option Base 1
Dim s_1, s_2(), s_3(), RowCount, ColCount
Dim mat_res()
Dim arr_Max_Min() As Double
Dim arr_min(), Max_Min As Double
Dim arr_clast_dist() As Double 'матрица расстояний между кластерами
Dim arr_num() As Integer
Dim Ind_arr() As Integer
Dim myCell3, myCELL2

'Разбиение на кластеры методом изотонических расстояний
Sub clast_izoton()
Set myCELL = Application.InputBox(Prompt:="", Title:="Выберите исходную матрицу данных", Default:=Cells(1, 255).Value, Type:=8)
Set myCell3 = Application.InputBox(Prompt:="", Title:="Выберите ячейки, содержащие имена объектов", Default:=Cells(2, 255).Value, Type:=8)
Set myCELL2 = Application.InputBox(Prompt:="", Title:="Выберите ячейку, с которой будут выводиться результаты", Default:=Cells(3, 255).Value, Type:=8)
Cells(1, 255) = myCELL.Address
Cells(2, 255) = myCell3.Address
Cells(3, 255) = myCELL2.Address

Num_row = myCELL.Rows.count 'Вычисление количества строк
Num_col = myCELL.Columns.count 'Вычисление количества столбцов
' Выполним нормировку исходных данных
ReDim s_1(Num_col)
ReDim s_2(1 To Num_col, 1 To Num_row)
'Вычисление суммы по столбцам и помещение ее в массив s_1
For i = 1 To Num_col
s_1(i) = Application.Sum(myCELL.Columns(i))
For j = 1 To Num_row
s_2(i, j) = myCELL.Columns(i).Cells(j) / s_1(i)
Next j
Next i
ReDim s_3(Num_row)
' Расчет длин векторов
ReDim mas_min(Num_row)
For i = 1 To Num_row
s_3(i) = 0
mas_min(i) = 1
For j = 1 To Num_col
s_3(i) = s_3(i) + s_2(j, i)
Next j
Next i
ReDim mat_res(1 To Num_row, 1 To Num_row)
' Расчет матрицы расстояний
For i = 1 To Num_row
For j = 1 To Num_row
mat_res(i, j) = Abs(s_3(i) - s_3(j))
If mat_res(i, j) < 1.1E-15 Then mat_res(i, j) = 0
Next j
Next i
'Формирование и вывод изотонической матрицы расстояний
myCELL2.Offset(0, 1).Value = "Матрица расстояний между объектами"
For i = 1 To Num_row
myCELL2.Offset(1, i).Value = myCell3.Cells(i)
myCELL2.Offset(1, i).Font.Italic = True
myCELL2.Offset(1, i).Borders.Item(xlEdgeBottom).LineStyle = xlDouble
myCELL2.Offset(i + 1, 0).Value = myCell3.Cells(i)
myCELL2.Offset(i + 1, 0).Font.Italic = True
myCELL2.Offset(i + 1, 0).Borders(xlEdgeRight).LineStyle = xlDouble
Next i
myCELL2.Offset(1, 0).Borders.Item(xlEdgeBottom).LineStyle = xlDouble
myCELL2.Offset(1, 0).Borders(xlEdgeRight).LineStyle = xlDouble
For i = 1 To Num_row
For j = 1 To Num_row
myCELL2.Offset(i + 1, j).Value = mat_res(i, j)
Next j
Next i
'Разбиение на кластеры и поиск следов
Dim Sum_dist As Double ' Суммарное расстояние
Dim Num_Links As Integer ' Количество связей
ReDim arr_num(1 To RowCount) As Integer
ReDim Ind_arr(1 To RowCount) As Integer
Num_of_trek = 0 ' Количество кластеров
Num_in_treks = 0 ' Счетчик уже классифицированных объектов
shift1 = 0
'Создание индексного массива для запоминания, включен ли
'рассматриваемый объект в какой - либо кластер или нет: 0-еще не включен;
'н-включен в n-й кластер.
Dim k_A As Integer 'Номер объекта A
Dim k_B As Integer 'Номер объекта B
Dim Ind_arr_order() As Integer 'Массив порядка включения объектов в кластер
ReDim Ind_arr_order(RowCount)
Dim Num_obj_in_trek() As Integer
For i = 1 To RowCount
Ind_arr(i) = 0
Ind_arr_order(i) = 0
arr_min(i) = 1
Next i
'Max_Min = 0.088977
Max_Min = InputBox("", "Введите значение критиче-

```

```

ского расстояния", Max_Min)
'Формирование вектора минимальных значений
For i = 1 To RowCount
If Ind_arr(i) = 0 Then
k_A = i
For j = 1 To RowCount
If i <> j Then
'If arr_min(i) > mat_res(i, j) Then
If Max_Min > mat_res(i, j) Then
'arr_min(i) = mat_res(i, j)
k_B = j
Else: GoTo 111000
End If
Else: GoTo 111000
End If
'Next j
'End If
If ((Ind_arr(k_B) <> 0) And (Ind_arr(k_A) <> 0)) Then
GoTo 111000
If (Ind_arr(k_B) <> 0) Or ((Ind_arr(k_B) = 0) And (Ind_
arr(k_A) <> 0)) Then 'Запись объекта в уже существующий кластер
If ((Ind_arr(k_B) = 0) And (Ind_arr(k_A) <> 0)) Then
Ind_arr(k_B) = Ind_arr(k_A)
Else
Ind_arr(k_A) = Ind_arr(k_B)
End If
Num_obj_in_trek(Ind_arr(k_B)) = Num_obj_in_
trek(Ind_arr(k_B)) + 1
Ind_arr_order(k_A) = Num_obj_in_trek(Ind_arr(k_B))
Sum_dist = Sum_dist + arr_min(i)
Else 'Создание нового кластера
Num_of_trek = Num_of_trek + 1
ReDim Preserve Num_obj_in_trek(Num_of_trek)
Ind_arr(k_B) = Num_of_trek
Ind_arr(k_A) = Num_of_trek
Num_obj_in_trek(Num_of_trek) = 2
Ind_arr_order(k_A) = 1
Ind_arr_order(k_B) = 2
Sum_dist = Sum_dist + arr_min(i)
End If
111000
Next j
End If
Next i
Dim Tmp_arr() As Integer
Dim ind As Integer
Dim tp_ind As Integer
For i = 1 To Num_of_trek
Num_in_treks = 0
myCELL2.Offset(RowCount + i * 2 + 1, 0).Value = "Кла-
стер" & i & " (" & Num_obj_in_trek(i) & ")" & ":"
ReDim Tmp_arr(2, Num_obj_in_trek(i))
For j = 1 To RowCount
If Ind_arr(j) = i Then
Num_in_treks = Num_in_treks + 1
Tmp_arr(1, Num_in_treks) = j
Tmp_arr(2, Num_in_treks) = Ind_arr_order(j) '(1,7)(1,6)
End If

```

```

Next j
'Сортировка кластеров для вывода
ind = 1
While ind = 1
ind = 0
For i1 = 1 To Num_obj_in_trek(i) - 1
If Tmp_arr(2, i1) > Tmp_arr(2, i1 + 1) Then
ind = 1
tp_ind = Tmp_arr(2, i1)
Tmp_arr(2, i1) = Tmp_arr(2, i1 + 1)
Tmp_arr(2, i1 + 1) = tp_ind
tp_ind = Tmp_arr(1, i1)
Tmp_arr(1, i1) = Tmp_arr(1, i1 + 1)
Tmp_arr(1, i1 + 1) = tp_ind
End If
Next i1
Wend
'Выведение кластеров
For j = 1 To Num_obj_in_trek(i)
myCELL2.Offset(RowCount + i * 2 + 1, j * 2 - 1).Value
= myCell3(Tmp_arr(1, j))
myCELL2.Offset(RowCount + i * 2 + 1, j * 2 - 1).Hori-
zontalAlignment = xlCenter
myCELL2.Offset(RowCount + i * 2 + 1, j * 2 - 1).Bor-
ders.LineStyle = xlDouble
If j < Num_obj_in_trek(i) Then
myCELL2.Offset(RowCount + i * 2 + 1, j * 2).Value =
mat_res(Tmp_arr(1, j), Tmp_arr(1, j + 1))
myCELL2.Offset(RowCount + i * 2 + 1, j *
2).Borders(xlEdgeBottom).LineStyle = xlContinuous
End If
Next j
Next i
'Формирование матрицы расстояний между кластера-
ми
Dim arr1() As Integer
Dim arr2() As Integer
Dim Arr_Links() As Integer
Dim arr_trek_dist() As Double 'матрица расстояний
между кластерами
ReDim arr_trek_dist(Num_of_trek, Num_of_trek)
ReDim arr1(RowCount)
ReDim arr2(RowCount)
ReDim Arr_Links(2, Num_of_trek, Num_of_trek)
For i = 1 To Num_of_trek - 1
' Формирование массива номеров объектов i-го кла-
стера
k_A = 0
For j = 1 To RowCount
If Ind_arr(j) = i Then
k_A = k_A + 1
arr1(k_A) = j
End If
Next j
For j = i + 1 To Num_of_trek
' Формирование массива номеров объектов j-го кла-
стера
k_B = 0
For k = 1 To RowCount

```

```

If Ind_arr(k) = j Then
k_B = k_B + 1
arr2(k_B) = k
End If
Next k
' Поиск наименьшего расстояния между объектами,
находящимися в массивах arr1 и arr2
arr_trek_dist(i, j) = 1
For i1 = 1 To k_A
For j1 = 1 To k_B
If arr_trek_dist(i, j) > mat_res(arr1(i1), arr2(j1)) Then
arr_trek_dist(i, j) = mat_res(arr1(i1), arr2(j1))
Arr_Links(1, i, j) = arr1(i1)
Arr_Links(2, i, j) = arr2(j1)
End If
Next j1
Next i1
arr_trek_dist(j, i) = arr_trek_dist(i, j)
Arr_Links(1, j, i) = Arr_Links(1, i, j)
Arr_Links(2, j, i) = Arr_Links(2, i, j)
Next j
Next i
' Выведение матрицы расстояний между кластерами
shift1 = Num_of_trek * 2 + 2 + 1
myCELL2.Offset(RowCount + shift1, 0).Value = " Ма-
трица межкластерных расстояний "
For i = 1 To Num_of_trek
myCELL2.Offset(RowCount + shift1 + 1, i).Value =
"Кластер" & i
myCELL2.Offset(RowCount + shift1 + 1, i).Font.Italic
= True
myCELL2.Offset(RowCount + shift1 + 1, i).Borders.
Item(xlEdgeBottom).LineStyle = xlDouble
myCELL2.Offset(RowCount + shift1 + 2 * i, 0).Value =
"Кластер" & i
myCELL2.Offset(RowCount + shift1 + 2 * i, 0).Font.
Italic = True
myCELL2.Offset(RowCount + shift1 + 2 * i,
0).Borders(xlEdgeRight).LineStyle = xlDouble
myCELL2.Offset(RowCount + shift1 + 2 * i + 1,
0).Borders(xlEdgeRight).LineStyle = xlDouble
Next i
myCELL2.Offset(RowCount + shift1 + 1, 0).Borders.
Item(xlEdgeBottom).LineStyle = xlDouble
myCELL2.Offset(RowCount + shift1 + 1,
0).Borders(xlEdgeRight).LineStyle = xlDouble

```

```

For i = 1 To Num_of_trek
For j = 1 To Num_of_trek
If (i = j) Then myCELL2.Offset(RowCount + shift1 + 2
* i + 1, i).Font.Bold = True
myCELL2.Offset(RowCount + shift1 + 2 * i, j).Value =
arr_trek_dist(i, j)
myCELL2.Offset(RowCount + shift1 + 2 * i + 1,
j).Value = "(" & Arr_Links(1, i, j) & "; " & Arr_Links(2,
i, j) & ")"
myCELL2.Offset(RowCount + shift1 + 2 * i + 1,
j).HorizontalAlignment = xlCenter
Next j
Next i
' Поиск минимальных расстояний между кластерами
и их связей
ReDim arr_min(Num_of_trek)
Dim arr_trek_num() As Integer
ReDim arr_trek_num(Num_of_trek)
For i = 1 To Num_of_trek
arr_min(i) = 1
For j = 1 To Num_of_trek
If i <> j Then
If arr_min(i) > arr_trek_dist(i, j) Then
arr_min(i) = arr_trek_dist(i, j)
arr_trek_num(i) = j
End If
End If
Next j
Next i
shift1 = shift1 + 2 * Num_of_trek + 3
myCELL2.Offset(RowCount + shift1, 0).Value = "Мин.
расст."
myCELL2.Offset(RowCount + shift1 + 1, 0).Value =
"Между класт."
For i = 1 To Num_of_trek
myCELL2.Offset(RowCount + shift1,
i).Borders(xlEdgeTop).LineStyle = xlDouble
myCELL2.Offset(RowCount + shift1, i).Value = arr_
min(i)
myCELL2.Offset(RowCount + shift1 + 1, i).Value = "("
& i & "; " & arr_trek_num(i) & ")"
myCELL2.Offset(RowCount + shift1 + 1,
i).HorizontalAlignment = xlCenter
Next i
End Sub

```

Приложение 2

Программа для дискриминантного анализа

```

Sub diskrim_n_m()
'Подпрограмма дискриминантного анализа для n со-
вокупностей объектов, позволяющая сразу классифи-
цировать целую матрицу новых объектов
Dim Mean_n() As Double
Dim Mean_n_tmp() As Double
Dim Cov_mat_n() As Double
Dim Cov_all() As Double
Dim Cof_disk() As Double
Dim mat_res()

```

```

Dim mas_min() As Double
Dim myfactor_n() As Double
Dim Obj_count As Integer 'Счетчик совокупностей
(классов) объектов
Dim Obj_mas() As Object 'Массив ссылок на совокуп-
ности (классы) объектов
Dim Row_n() As Integer
Dim Col_n() As Integer
Dim All_Row As Double
Dim Count_New_Obj As Integer 'Количество новых

```



```

объектов, предназначенных для классификации
'Ввод ссылки на ячейку, в которой хранится количест-
во 'совокупностей объектов
Set CELL_count = Application.InputBox(Prompt:="",
Title:="Выберите ячейку, в которой содержится ко-
личество совокупностей объектов", Default:=Cells(1,
255).Value, Type:=8)
Cells(1, 255) = CELL_count.Address
Obj_count = CELL_count.Cells.Value
ReDim Obj_mas(Obj_count)
ReDim Row_n(Obj_count) As Integer
ReDim Col_n(Obj_count) As Integer
'Ввод выборки совокупностей объектов
For i = 1 To Obj_count
Set Obj_mas(i) = Application.InputBox(Prompt:="",
Title:="Выберите " & i & "-ю совокупность объектов",
Default:=Cells(i + 1, 255).Value, Type:=8)
Cells(i + 1, 255) = Obj_mas(i).Address
Next i
'Ввод ссылки на вторую совокупность объектов
Set myObj = Application.InputBox(Prompt:="",
Title:="Выберите группу объектов, которые предназ-
начены для классификации", Default:=Cells(10, 255).
Value, Type:=8)
Cells(10, 255) = myObj.Address
Count_New_Obj = myObj.Rows.count
Set Name_myObj = Application.InputBox(Prompt:="",
Title:="Выберите названия объектов, которые пред-
назначены для классификации", Default:=Cells(11,
255).Value, Type:=8)
Cells(11, 255) = Name_myObj.Address
Set myCell3 = Application.InputBox(Prompt:="",
Title:="Выберите ячейку, с которой будут выводиться
результаты", Default:=Cells(12, 255).Value, Type:=8)
Cells(12, 255) = myCell3.Address
All_Row = 0
For i = 1 To Obj_count
Row_n(i) = Obj_mas(i).Rows.count 'Вычисление коли-
чества строк
All_Row = All_Row + Row_n(i)
Col_n(i) = Obj_mas(i).Columns.count 'Вычисление ко-
личества столбцов
Next i
'1. Вычисление векторов средних значений для каждо-
го класса (совокупности) объектов
ReDim Mean_n(Col_n(1), Obj_count)
ReDim Mean_n_tmp(Col_n(1), 1)
For j = 1 To Obj_count
For i = 1 To Col_n(1)
Mean_n(i, j) = Application.Average(Obj_mas(j).
Columns(i))
Next i
Next j
'2. Вычисление оценок ковариационных матриц для
каждого класса
ReDim Cov_mat_n(1 To Obj_count, 1 To Col_n(1), 1 To
Col_n(1))
For k = 1 To Obj_count
For i = 1 To Col_n(1)

```

```

For j = i To Col_n(1)
Cov_mat_n(k, i, j) = Application.Covar(Obj_mas(k).
Columns(i), Obj_mas(k).Columns(j))
Next j
Next i
Next k
'3. Вычисление несмещенной оценки объединенной
ковариационной матрицы
ReDim Cov_all(1 To Col_n(1), 1 To Col_n(1))
For k = 1 To Obj_count
For i = 1 To Col_n(1)
For j = i To Col_n(1)
Cov_all(i, j) = Cov_all(i, j) + (1 / (All_Row - Obj_count
- 1))
* (Row_n(k) * Cov_mat_n(k, i, j))
If (i <> j) Then Cov_all(j, i) = Cov_all(i, j)
Next j
Next i
Next k
'4. Вычисление матрицы, обратной объединенной ко-
вариационной матрице
ReDim Cov_inv(1 To Col_n(1), 1 To Col_n(1))
Cov_inv() = Application.MInverse(Cov_all())
Dim Cof_tmp() As Double
ReDim Cof_tmp(1 To Col_n(1), 1 To 1)
'Вычисление векторов оценок коэффициентов дискри-
минации A(i)
ReDim Cof_disk(1 To Col_n(1), 1 To Obj_count)
For i = 1 To Obj_count
For j = 1 To Col_n(1)
Cof_disk(j, i) = 0
For l = 1 To Col_n(1)
Cof_disk(j, i) = Cof_disk(j, i) + Cov_inv(j, l) * Mean_n(l,
i)
Next l
Next j
Next i
'Вычисление дискриминантных констант лямда
Dim lamda_n() As Double
ReDim lamda_n(1 To Obj_count)
Dim Xl_tmp() As Double
ReDim Xl_tmp(1, Col_n(1))
Dim Res_tmp() As Double
ReDim Res_tmp(1, Col_n(1))
For i = 1 To Obj_count
For j = 1 To Col_n(1)
Xl_tmp(1, j) = Mean_n(j, i)
Mean_n_tmp(j, 1) = Mean_n(j, i)
Next j
For j = 1 To Col_n(1)
Res_tmp(1, j) = 0
For l = 1 To Col_n(1)
Res_tmp(1, j) = Res_tmp(1, j) + Xl_tmp(1, l) * Cov_inv(l,
j)
Next l
lamda_n(i) = lamda_n(i) + Res_tmp(1, j) * Mean_n_
tmp(j, 1)
Next j
lamda_n(i) = lamda_n(i) / 2

```

```

Next i
'Определение, к какому классу относится каждый из
новых объектов
Dim cof_prinadl() As Double
ReDim cof_prinadl(Obj_count) As Double
myCell3.Offset(0, 0).Value = "Результаты дискри-
нантного анализа"
'формирование шапки таблицы результатов
myCell3.Offset(1, 0).Value = "Имя объекта"
myCell3.Offset(1, 1).Value = "Макс. знач."
myCell3.Offset(1, 2).Value = "Класс"
For m = 1 To Count_New_Obj
For k = 1 To Obj_count
cof_prinadl(k) = 0
For i = 1 To Col_n(1)
cof_prinadl(k) = cof_prinadl(k) + myObj.Cells(m, i) *
Cof_disk(i, k)
Next i
cof_prinadl(k) = cof_prinadl(k) - lamda_n(k)
Next k
k = 1
Max_val = cof_prinadl(1)
For i = 2 To Obj_count
If Max_val < cof_prinadl(i) Then Max_val = cof_
prinadl(i): k = i
Next i
'Вывод результатов дискриминантного анализа для
одного объекта
myCell3.Offset(1 + m, 0).Value = Name_myObj.Cells(m)
myCell3.Offset(1 + m, 1).Value = Max_val
myCell3.Offset(1 + m, 2).Value = k
Next m
End Sub

```

ВОЗЯКОВ Владимир Иванович – доктор физико-математических наук, профессор, зав. кафедрой математических и инструментальных методов экономики. Чебоксарский кооперативный институт (филиал) Российского университета кооперации. Россия. Чебоксары. E-mail: vvozyakov@rucoop.ru

ФИЛИПPOB Владимир Петрович – магистрант. Чебоксарский кооперативный институт (филиал) Российского университета кооперации. Россия. Чебоксары. E-mail: filippov_v_p@rambler.ru

VOZYAKOV, Vladimir Ivanovich – Doctor of Physical and Mathematical Sciences, Professor, Department Chair of Mathematical and Tool Methods of Economics. Cheboksary Cooperative Institute (branch) of Russian University of Cooperation. Russia. Cheboksary. E-mail: vvozyakov@rucoop.ru

FILIPPOV, Vladimir Petrovich – Undergraduate. Cheboksary Cooperative Institute (branch) of Russian University of Cooperation. Russia. Cheboksary. E-mail: filippov_v_p@rambler.ru

УДК 334.7

ПОСТРОЕНИЕ ТРЕХУРОВНЕВОЙ СИСТЕМЫ КРЕДИТНОЙ КООПЕРАЦИИ В РОССИЙСКОЙ ФЕДЕРАЦИИ ПУТЕМ УЧРЕЖДЕНИЯ КРЕДИТНОЙ ОРГАНИЗАЦИИ

А.И. Воробьев

Посвящено исследованию трехуровневой системы кредитной кооперации в Российской Федерации. Проанализировано текущее состояние развития и предложены направления совершенствования. Разработана модель системы кредитной кооперации с небанковской кредитной организацией на верхнем уровне.

Ключевые слова: кредитный потребительский кооператив; система кредитной кооперации; кооперативный банк; кредитный кооператив второго уровня.

A.I. Vorobyev. THREE-LEVEL SYSTEM CONSTRUCTION OF CREDIT COOPERATIVES IN THE RUSSIAN FEDERATION THROUGH THE ESTABLISHMENT OF A CREDIT ORGANIZATION

The article investigates the three-level system of credit cooperatives in the Russian Federation. The analyze the current state of development is made and suggested ways of its improving. The model system of credit cooperatives with non-bank credit organization at the top level has been worked out.

Keywords: credit consumer cooperative; system of credit cooperatives; Cooperative Bank; credit cooperative second level.