

## КЛАСТЕРИЗАЦИЯ И РЕЙТИНГОВАНИЕ БАНКОВ С ПОМОЩЬЮ НЕЙРОННОЙ СЕТИ КОХОНЕНА В СРЕДЕ MS EXCEL

Аникин Валерий Иванович  
д.т.н., профессор кафедры «Информационный и электронный сервис», Поволжский  
Государственный Университет Сервиса, г. Тольятти, РФ  
Карманова Александра Александровна  
инженер-программист ООО «НетКрэкер», г. Тольятти, РФ

*Рассмотрено применение табличной реализации в Excel клеточной искусственной нейронной сетью Кохонена для кластеризации и составления рейтингов банков на основе их финансово-экономических показателей.*

Целью работы являлась разработка методики разделения банков на группы по ряду их финансово-экономических показателей для получения объективной картины состояния банковской сферы в разрезе ее участников. Полученные данные дают возможность построения рейтинга банков и прогнозирования их будущего в зависимости от положения в конкретной группе с учетом статистических характеристик кластера и актуальной рыночной информации о других банках, его сформировавших.

Для проведения экспериментальных исследований были выбраны данные об активах банков, характеризуемых четырьмя входными переменными: *Сумма активов, Собственные активы, Банковские активы, Средства в банке*, - и одной выходной переменной *Прибыль*. Выборка данных объемом 99 образцов взята из демонстрационного примера аналитической платформы Deductor Studio в связи с хорошей коррелированностью и сильной неоднородностью пространственного распределения нормированных образцов, что делает ее особенно привлекательной для кластерного анализа с помощью нейронной сети Кохонена.

В работе использован оригинальный алгоритм обучения нейронной сети Кохонена клеточным автоматом [1] и произведено сравнение наших результатов кластеризации с аналогичными результатами, получаемыми в Deductor Studio (<http://www.basegroup.ru>) посредством классического алгоритма обучения сети Кохонена [2].

Важной особенностью данной работы является то, что нейронная сеть Кохонена была реализована чисто табличными средствами в виде итерационной табличной модели MS Excel без программирования на языке VBA, что само по себе является нетривиальной задачей [3, 4].

Преимуществом алгоритма обучения нейронной сети Кохонена клеточным автоматом в сравнении с классическим алгоритмом обучения является небольшое число настраиваемых параметров, к которым относятся следующие (рис.1):

- *число прогонов* – число эпох обучения нейронной сети; в каждой эпохе учебные образцы подаются на вход нейронной сети последовательно в порядке возрастания их номеров;
- *скорость обучения* – параметр, регулирующий скорость обучения сети Кохонена клеточным автоматом; обычно этот параметр равен 1, что обеспечивает максимальную временную эффективность алгоритма обучения;
- *число эпох взаимодействия* – доля эпох обучения, в течение которых связи между соседними нейронами сети включены;
- *эпох разворачивания сети* – число начальных эпох взаимодействия, в течение которых нейроны сети саморазворачиваются в пространстве учебных образцов в соответствии со структурой самоорганизующейся карты Кохонена и пространственным распределением многомерных данных.

Флаг включения итераций	0
Число итераций	68
Счетчик итераций	0

Число нейронов	30
Ширина карты	6

Фиксировать веса

Выполнить

Число эпох	1,0000
Счетчик эпох	0
Счетчик итераций эпохи	0

Число прогонов	50
----------------	----

Шаг

Сброс

Число образов	68
Номер текущего образа	1

Скорость обучения	1,00	1,00
Число эпох взаимодействия	25	
Эпох разворачивания сети	2	2

Рис.1. Пользовательский интерфейс управления процессом обучения нейронной сети Кохонена.

Флажок «Фиксировать веса» пользовательского интерфейса предотвращает возможность непреднамеренного изменения весов нейронов после обучения сети в результате ошибочных действий пользователя, таких как случайное нажатие кнопки «Сброс». Кнопка «Выполнить» запускает цикл обучения нейронной сети в течение заданного числа прогонов, кнопка «Шаг» позволяет выполнять обучение сети в шаговом режиме и используется в режиме отладки, кнопка «Сброс» приводит нейронную сеть Кохонена в исходное необученное состояние.

Пользовательский интерфейс табличной модели клеточной нейронной сети Кохонена позволяет легко изменять параметры обучения и визуально наблюдать за ходом обучения сети.

На рис.2 показан интерфейс управления конфигурацией обучающего клеточного автомата, используемого в фазе взаимодействия нейронов сети Кохонена. Как видно из рисунка, клеточная нейронная сеть Кохонена может содержать до 30 связанных нейронов на плоской решетке размером 5\*6, где активными являются нейроны с номерами 0, 1, 6 и 7, остальные нейроны не используются.

Связи по выбору  
 Все связи

Связи по выбору

0	+	1	-	2	-	3	-	4	-	5
+	+	-	-	-	-	-	-	-	-	-
6	+	7	-	8	-	9	-	10	-	11
-	-	-	-	-	-	-	-	-	-	-
12	-	13	-	14	-	15	-	16	-	17
-	-	-	-	-	-	-	-	-	-	-
18	-	19	-	20	-	21	-	22	-	23
-	-	-	-	-	-	-	-	-	-	-
24	-	25	-	26	-	27	-	28	-	29

Рис.2. Интерфейс управления связями между соседними нейронами обучающего клеточного автомата

Процесс обучения ИНС Кохонена клеточным автоматом в Excel включает следующие последовательные этапы:

- настройка параметров обучения через пользовательский интерфейс табличной модели сети, таких как структура сети, скорость обучения, доля эпох взаимодействия, число прогонов, конфигурация обучающего клеточного автомата, начальное состояние нейронов сети;
- запуск процесса обучения сети в течение заданного числа прогонов или ее пошаговое обучение;
- фиксация весов нейронов обученной сети или приведение нейронной сети в исходное необученное состояние с целью повторения эксперимента при других входных параметрах.

Предварительные эксперименты показали, что обучение искусственной нейронной сети (ИНС) Кохонена классическим алгоритмом в Deductor Studio и клеточным автоматом в Excel дает одинаковые результаты, надежно разделяя пространство учебных образцов на четыре кластера, однако проблема заключалась в том, что 75% наиболее интересных и разнообразных образцов попали в один кластер, что лишает результаты кластеризации практического смысла. Простое увеличение размерности сети Кохонена приводило к дроблению уже обнаруженных кластеров на более мелкие и дальнейшему усложнению ситуации.

Для решения возникшей проблемы было решено использовать многослойную кластеризацию, когда в каждом новом слое для обучения сети используется усеченная выборка данных, не содержащая образцов, попавших в выделенные предыдущими слоями кластеры.

Методом многослойной кластеризации в первом слое из 99 образцов клеточным автоматом было надежно выделено 2 кластера (по 5 и 6 банков соответственно); во втором слое из оставшихся 88 образцов – еще 2 кластера (по 11 и 9 банков). Обучение 3-го слоя выборкой из 68 оставшихся образцов дало неожиданный результат: банки не удалось однозначно разделить на кластеры ни классическим алгоритмом обучения сети в Deductor, ни с помощью клеточного автомата, что объясняется невысокой коррелированностью обучающей выборки 3-го слоя. Тем не менее, определенная организованность образцов в данных третьего слоя еще оставалась, так как все коэффициенты корреляции между координатами образцов были положительными, а распределение учебных образцов по перцентильям заметно неравномерным.

Для увеличения коррелированности учебных образцов обучающей выборки данных 3-го слоя был использован метод линейного преобразования их координат, что позволило надежно выделить еще 4 кластера (по 9, 13, 19 и 27 банков). В итоге во всей выборке было выделено 8 кластеров со стабильными участниками. О правильности выделения всех кластеров говорит совпадение центра тяжести кластеров со средними значениями характеристик банков, их сформировавших.

Оценка рейтинга  $R_t$  банков по всей выборке производилась на основе среднего значения нормированных входных координат учебных образцов благодаря высокой коррелированности  $R_t$  с этими координатами по формуле:

$$R_t = \frac{1}{d} \sum_{k=1}^d x_k, \quad (1)$$

где  $x_k$  - нормированное значение  $k$ -координаты,  $d$  – размерность пространства образцов.

Для визуализации результатов обработки данных был использован такой уникальный инструмент Excel, как условное форматирование, которое с помощью выбранной цветовой шкалы автоматически разбивает выделенный диапазон на нужное количество равных интервалов и окрашивает ячейки для каждого значения цветом интервала, в которое попало данное число.

На рис.3 приведен фрагмент визуального представления результатов кластеризации и рейтингования банков. К каждому столбцу входных и выходного параметров было применено условное форматирование с градиентной заливкой ячеек, начиная от красного

цвета для самого большого значения до зеленого цвета с минимальным значением. Такое же форматирование было применено и к правому столбцу Rt, хранящему значения с коэффициентом рейтинга банков.

АКТИВЫ И РЕЙТИНГ БАНКОВ								
№	Банк	Сумма актива	Собственные активы	Банковская сумма активы	Средства в бан	Прибыль	Класс	Rt
1	Внешторгбанк	101660298	23236327	84543558	30863714	245820	1	91,79
3	ООО "Международный Промышленн	77888642	26409116	58647197	6712070	306196	1	65,75
2	Газпромбанк	79012789	9255041	74409960	18833125	355197	1	62,25
4	Международный Московский Банк	63910966	1176462	62436148	38973392	0	1	60,20
5	ОАО "АЛЬФА-БАНК"	57510886	12446938	52348562	28996990	1218928	1	59,88
8	АКБ "РОСБАНК" (ОАО)	28105202	4691449	25807591	8848426	257710	4	24,40
7	Банк Москвы	31352124	3335734	30287158	3781859	53771	4	22,01
6	ОАО "ПСБ"	49406525	1275859	17091603	2626908	271635	4	19,90
9	АКБ "ДИБ"	27350369	2616993	26986210	3418864	363622	4	19,12
12	ОАО РИКБ "Башкредитбанк"	23304860	2833897	21190313	4294120	330620	4	17,15
10	КБ "Ситибанк Т/О"	26240408	2063168	23291673	281327	839596	4	15,21
17	АКБ "НРБАНК"	15791359	5075844	15773889	1714380	153763	2	14,14
14	БНП-Дрезднер Банк	17689833	84312	13082069	9186855	-140918	2	13,88
21	ГЛОБЭКСБАНК	12769906	5016721	12576820	1711287	4756	2	12,38
18	АКБ "Еврофинанс"	15191022	2835944	13398039	2157847	290095	2	11,45
15	АВТОБАНК	17229288	1530584	15502811	225050	76747	2	11,41
13	ЗАО Банк "МЕНАТЕП СПб"	19541965	2268343	9479194	2025822	128038	2	10,75
11	ОАО МАКБ "Возрождение"	25446423	1520076	9879932	565976	5367	2	10,69
20	КБ "ГУТА-БАНК"	14799899	2008471	12398573	2696602	70863	2	10,62
19	ООО Райффайзенбанк Австрия	14892329	593460	11685822	4164173	0	2	10,03
24	АБМ АМРО Банк АО	10566709	1079184	10007563	4715039	160417	2	9,27
16	ОАО Банк "Петрокоммерц"	16690301	727690	10372701	2593355	251850	2	9,21
25	Банк ЗЕНИТ	10407230	1000661	8303786	2587086	171525	2	7,28
28	АКБ "Ак-Барс"	8401583	2515240	7218500	1228651	63534	2	7,02
26	НОМОС-БАНК	8916553	1416318	8799375	1798646	16548	2	6,95
31	ЗАО "Конверсбанк"	7618503	1716464	6341373	932007	89416	3	5,62
23	ОАО "Банк "Петровский"	12589085	439907	5943321	100843	12461	3	5,00
27	АКБ "Московский Индустриальный	8467224	938868	6145063	219797	55390	3	4,58
22	ОАО "Ханты-Мансийский банк "	12722589	323652	2987333	705074	30679	3	4,43
29	ИМПЭКСБАНК (ООО)	7963291	1015977	3631085	816171	21202	3	4,16
36	АКБ "БИН"	6160046	739890	5824773	876788	81122	3	4,14
37	ОАО "Меткомбанк"	4879981	329236	4248641	2012002	48249	3	3,70
35	Транскредитбанк	6252523	721208	2733597	1491818	262046	3	3,63
66	КБ "ВИЗАВИ"	2493992	1535115	2493167	1580337	670	3	3,44
30	АКБ "ЧЕЛИНДБАНК"	7767092	463445	3934709	304154	63956	3	3,35
50	КБ "Олимпийский"	3170956	1216732	3163933	1133884	6613	3	3,23
33	Запсибкомбанк ОАО	6538257	335942	4257359	338994	45827	3	3,05
46	АКБ "Пробизнесбанк"	3481038	784577	2402092	1292998	6308	3	2,77
49	АКБ "МФК"	3346237	1204941	3184388	235949	32600	3	2,69
53	КБ "Легпромбанк	3021143	1450421	2929932	52346	21083	3	2,65
32	ОАО "Банк "Санкт-Петербург"	6730121	255345	2866942	283465	3057	3	2,57
47	ЗАО "Международный акционерный	3381733	357083	3365156	967892	25178	3	2,42
48	КБ "Русский генеральный банк"	3351413	502505	3333821	753790	22459	3	2,40
38	ЗАО АБ "Капитал"	4568332	230101	3579674	400875	61528	3	2,29
42	ЗАО "Нижегородпромстройбанк"	3735969	789338	2798278	145675	164187	3	2,22
39	Балтийский Банк	4134372	268751	3359926	470731	65889	3	2,20
63	ОАО КБ "МОСКОВСКИЙ КРЕДИТН	2571278	617162	2567898	857835	4518	3	2,15
46	Вост ЛБ Восток	3417928	609375	3398459	0	119340	3	2,05
41	АБ "ИБО"НИКОЙЛ"	3838123	666703	2804599	6362	47143	3	2,04
68	ООО КБ "ЕВРОТРАСТ"	2413941	562653	2401145	813065	25082	3	1,98
40	АКБ "Югбанк"	4120333	140846	2598064	677731	26991	3	1,98
34	КБ "СахаДаймондБанк"	6338764	70658	2112569	90028	3441	3	1,95
74	АКБ "ВЕКСИМ БАНК"	2118559	123497	2111697	1634809	4411	3	1,94
67	КУБ ОАО	2485820	414721	2389431	837810	276897	3	1,88
52	ОАО "Инкасбанк"	3024579	331051	2885567	495960	7584	3	1,86
61	АКБ "Межрегиональный инвестицио	2589355	814636	2328399	144115	13417	3	1,82
64	Банк "Авангард"	2560051	836620	2318671	88391	13202	3	1,79
70	МНБ Лтд	2326185	745298	2126985	350330	0	3	1,76
51	ОАО "Международный банк Санкт -	3139712	249190	3013127	353112	13912	3	1,76
54	АКБ "ГАЗБАНК"	3008408	302715	2646762	237551	14081	3	1,59
58	АКБ "Металлинвестбанк"	2649453	408386	2007521	494121	13807	3	1,58
65	Промторгбанк	2555697	354812	2364194	386095	7592	3	1,54
56	ОАО КБ "Солидарность"	2927714	209409	2569185	319268	32603	3	1,51
57	АКБ "ИНТЕРПРОМБАНК"	2691874	288063	2596142	158982	9637	3	1,43
77	АКБ "МЕРР"	1919367	766181	1741201	34710	0	3	1,36
76	КБ "Гаранти Банк-Москва"	1975412	466340	1951263	355105	41700	3	1,36
60	КБИнтурбанк"	2590835	685006	1369933	47248	6092	3	1,35
59	ОАО КБ "Далькомбанк"	2625776	169803	2196948	312573	77817	3	1,28
62	Уоэленвнештоогбанк	2573907	117258	2073443	417488	75944	3	1,25

Рис.3. Фрагмент визуализации результатов кластеризации и рейтингования банков

Отсортировав кластеризованную исходную выборку по номеру кластера и применив условное форматирование к каждому из столбцов измерений, мы получаем информацию о взаимном относительном расположении образцов исходной выборки

данных. Таким образом, условное форматирование в Excel позволяет получить так называемый атлас кластера, визуально описывающий, что представляет каждый кластер на основе информации о параметрах каждого образца в кластере.

С помощью механизма сводных таблиц для каждого из кластеров были также определены описательные статистики — средние, минимальные, максимальные значения, несмещенная дисперсия и несмещенное отклонение входных параметров в целом по выборке и в разрезе кластеров.

Результаты рейтингования подтвердили правильность результатов кластеризации с точки зрения экономических характеристик банков, попавших в один кластер: сильные банки с высоким рейтингом оказались в одних кластерах, мелкие банки с более низким рейтингом в других. При сортировке коэффициента  $R_t$  по убыванию рейтинг банков закономерно уменьшался, сохраняя правильный порядок убывания средних финансовых показателей кластеров.

Ранжирование банков в целом по выборке основывается на составлении их рейтинга, как на среднем уровне кластеров, так и в пределах кластера, что позволяет выделить банки, находящиеся в пограничном состоянии между двумя кластерами. Если банк по рейтингу находится на границе с более сильным кластером, то можно прогнозировать его развитие в будущем в сторону усиления финансового положения на рынке и возможный переход в более сильный кластер. И наоборот, если банк по рейтингу находится на границе с кластером, представленным банками – банкротами, то этот банк также близок к банкротству.

Обнаруженные закономерности позволяют обобщить предложенную методику многослойной кластеризации и рейтингования учебных образцов с помощью ИНС Кохонена, обучаемой клеточным автоматом, на большой класс многомерных выборок данных небольшого объема. Показано, что линейное преобразование координат обучающей выборки является важным инструментом для подготовки данных к кластеризации с помощью ИНС Кохонена и их надлежащего статистического анализа. Описанная авторская методика разделения банков на группы позволила учесть пространственные и корреляционные особенности обучающей выборки и получить обоснованные и стабильные результаты кластеризации как классическим, так и авторским алгоритмами обучения нейронной сети Кохонена. В отличие от классического алгоритма, предложенный алгоритм обучения сети Кохонена клеточным автоматом работает в десятки раз быстрее и имеет существенно меньшее число настраиваемых параметров.

Табличная реализация сети Кохонена, помимо удобства и гибкости реализации без необходимости программирования, предоставила возможность удобного анализа результатов кластеризации и рейтингования банков за счет встроенных визуальных возможностей электронных таблиц. Обученная сеть Кохонена позволяет дать оценку новым банкам, появляющимся на рынке, относя их к одному из выявленных кластеров.

## Литература

1. Аникин В.И. О возможности обучения искусственной нейронной сети Кохонена с помощью клеточного автомата / В.И. Аникин, А.А. Тураева. - Вектор науки ТГУ, 2011. - № 3(17). – с. 22-24
2. Кохонен Т. Самоорганизующиеся карты / Т. Кохонен. - Пер. 3-го англ. изд. – М.: БИНОМ. Лаборатория знаний, 2008. – 655 с.: ил.
3. Аникин В.И. Визуальное табличное моделирование клеточных автоматов в Microsoft Excel: монография / В.И. Аникин, О.В. Аникина. – Тольятти: Изд-во ПВГУС, 2013. – 324 с.
4. Аникин В.И. Алгоритмическое табличное моделирование в Microsoft Excel: итерационные модели / В.И.Аникин, О.В. Аникина. - «Информатика и образование», 2009. - № 9, с. 88-95.