

*ИСПОЛЬЗОВАНИЕ НЕЙРОСЕТЕВОГО МОДЕЛИРОВАНИЯ
В АНАЛИЗЕ ДЕЯТЕЛЬНОСТИ КРУПНЕЙШИХ КОМПАНИЙ
РОССИЙСКОЙ ФЕДЕРАЦИИ*

А.О. КУЗНЕЦОВА

*Кубанский государственный университет,
350040, Российская Федерация, г. Краснодар, ул. Ставропольская, 149;
электронная почта: florryboom@gmail.com*

В статье рассмотрена концепция метода нейросетевого моделирования на основе самоорганизующихся карт Кохонена. Анализ, проведенный по 400 крупнейшим компаниям РФ по объему выпускаемой продукции с помощью пакета программ Viscovery SOMine, основанных на самоорганизующихся картах, нацелен на отображение настоящего положения компаний в рейтинге, прогнозирования наиболее прибыльных отраслей экономики для дальнейшего грамотного управления ими.

Ключевые слова: нейрон, нейросетевое моделирование, сети Кохонена, самоорганизующаяся карта, кластеризация, обучение сети, объем реализации, темп прироста.

Актуальность и практическая значимость данной работы состоит в прогнозировании и дальнейшем выявлении наиболее прибыльных отраслей экономики Российской Федерации.

Нейросетевая методология для обработки сигналов использует явления, аналогичные тем, которые происходят в нейронах живых существ, являющихся одним из основных элементов нервной системы. Важнейшая особенность таких сетей состоит в параллельной обработке информации всеми звеньями, что позволяет значительно ускорить процесс обработки информации. Другим важным свойством нейронных сетей является способность к обучению и обобщению накопленных знаний. Самостоятельно концепция формирования искусственного происхождения нейронных сетей по типу аппарата нервной системы живых созданий с целью подборки новейших научно-технических заключений никак не нова. Ещё в 1943г. К. МакКаллох и А. Питсс формировали простую модификацию нейрона. Разум человека включает вплоть до 100 млрд (10^{11}) нейронов разных видов. Все без исключения они сложным способом взаимодействуют между собой и

подобраны в популяции – нейронные сети. Синтетический нейрон содержит подобное устройство с био нейроном, моделируя его [2].

Сети (слои) Кохонена (Kohonen T.) относятся к самоорганизующимся нейронным сетям. Самоорганизующаяся сеть позволяет выявлять кластеры (группы) входных векторов, обладающих некоторыми общими свойствами [6].

Кластеризация — это разделение исследуемого множества объектов на группы "похожих" объектов, называемых кластерами. Синонимами термина "кластер" являются термины класс, таксон, сгущение. Цель кластеризации сознательно различается от цели систематизации. Заключением проблемы систематизации считается причисление каждого из объектов к одному из предварительно определенных классов. В проблеме кластеризации совершается причисление объекта к одному из предварительно неясных классов. Разделение объектов по кластерам выполняется при осуществлении синхронного создания кластеров. Кластеризация позволяет сгруппировать сходные данные, что облегчает решение ряда задач:

1. Изучение данных, облегчение анализа. Содержательный анализ полученных кластеров позволяет обнаружить закономерности. Например, можно выявить группы клиентов сети сотовой связи, для которых можно предложить новый тарифный план. Другие примеры — выявление групп покупателей торговой сети, сегментация рынка. Анализ содержания кластера позволяет применить к объектам различных кластеров разные методы анализа.

2. Прогнозирование. Относя новый объект к одному из кластеров, можно прогнозировать поведение объекта, поскольку его поведение будет схожим с поведением объектов кластера.

3. Обнаружение аномалий. Содержательный анализ кластеров помогает выявить аномалии. Обычно, это кластеры, в которые попадает мало объектов [1].

В сетях Кохонена используется обучение без учителя. Для обучения сети применяются механизмы конкуренции. При подаче на вход сети вектора x побеждает тот нейрон, вектор весов которого в наименьшей степени

отличаются от входного вектора. Для нейрона-победителя выполняется соотношение [2]:

$$d(x, w_j) = \min d(x, w_i) \quad (1)$$

где:

n — количество нейронов,

j — номер нейрона-победителя,

$d(x, w)$ — расстояние (в смысле выбранной метрики) между векторами x и w .

Чаще всего в качестве меры расстояния используется евклидова мера [7]:

$$d(x, w_i) = \|x - w_i\| = \sqrt{\sum_{j=1}^m (x_j - w_{ji})^2} \quad (2)$$

Используются и другие меры расстояния (метрики). Конкурирующая функция активации анализирует значения сумматоров и формирует выходы нейронов, равные 0 для всех нейронов, кроме одного «нейрона-победителя», имеющего на выходе максимальное значение. Таким образом, вектор выхода имеет единственный элемент, равный 1, который соответствует нейрону-победителю, а остальные равны 0. Номер активного нейрона определяет ту группу (кластер), к которой наиболее близок входной вектор. В сети Кохонена входные значения желательно (хотя и не обязательно) нормировать. Для этого следует воспользоваться одной из следующих формул [5]:

$$x_{ni} = \frac{x_i}{\sqrt{\sum_{l=1}^m x_l^2}}, \quad x_{ni} = \frac{x_i}{|x_i|} \quad (3)$$

где:

x_{ni} — нормированный компонент входного вектора.

Нормирование входных данных положительным образом сказывается на скорости обучения сети.

Перед процессом обучения производится инициализация сети, то есть первоначальное задание векторов весов. В простейшем случае задаются

случайные значения весов. Процесс обучения сети Кохонена состоит из циклического повторения ряда шагов:

1. Подача исходных данных на входы. Обычно это случайная выборка одного из входных векторов.

2. Нахождение выхода каждого нейрона.

3. Определение «выигравшего» нейрона (веса которого в наименьшей степени отличаются от соответствующих компонентов входного вектора), или нейрона-победителя.

4. Корректировка весов «выигравшего» нейрона по правилу Кохонена:

$$w_i^{(k+1)} = w_i^{(k)} + \eta_i^{(k)} [x - w_i^{(k)}] \quad (4)$$

где:

x — входной вектор,

k — номер цикла обучения,

$\eta_i^{(k)}$ — коэффициент скорости обучения i -го нейрона в k -ом цикле обучения.

5. Переход на шаг 1, если обучение не завершено. Часто, например, в Neural Network Toolbox MATLAB, задается число циклов обучения. Можно проверять достижение малой величины функционала ошибки [9]:

$$E = \frac{1}{Q} \sum_{i=1}^Q \|x_i - w_{x_i}\|^2 \quad (5)$$

где:

w_{x_i} — вектор весов нейрона-победителя при предъявлении входного i вектора x_i ,

Q — размер обучающей выборки.

Подобным способом, нейрон, чей градиент весов был поближе к входному вектору, обновляется, для того чтобы быть ещё поближе. В следствии данный нейрон, скорее всего, победит конкурентную борьбу при подаче в доступ ближнего вектора и проиграет при подаче значительно отличающегося вектора. Уже после неоднократной подачи обучающих

векторов станет находиться нейрон, который дает 1, если градиент относится к кластеру, и 0, если градиент никак не относится к кластеру. Подобным способом, линия обучается систематизировать входные векторы. При обучении сети Кохонена появляется трудность так называемых «мертвых» нейронов. Одно из ограничений любого создающего конкуренцию покрова заключается в том, что определенные нейроны становятся резервными и не используются. Данный факт выражается в том, что нейроны, обладающие первоначальными весовыми векторами, существенно далеки от векторов входа, никогда не побеждают в конкурентной борьбе, вне зависимости от того, как долго длится подготовка. В следствии этого в действительности, то что подобные векторы никак не применяются при обучении и надлежащие нейроны ни в коем случае никак не становятся к фаворитам. Подобные «нейроны-аутсайдеры» именуют «мертвыми» нейронами, так как они не осуществляют практически никакой нужной функции. Подобным способом, входные сведения станут интерпретироваться наименьшим количеством нейронов. По этой причине необходимо предоставить возможность выиграть абсолютно всем нейронам. С этой целью метод преподавания видоизменяют подобным способом, для того чтобы «мертвые» нейроны принимали участие в обучении [4].

В методе поочередного (Iterative) преподавания развитие весов выполняется уже после каждого обучающего примера. Подготовка нейронов выполняется подобно учебе нейронов сети Кохонена. Различие заключается в том, что помимо нейрона-фаворита учатся нейроны, вступающие в окрестность (neighborhood), либо радиус преподавания (radius of learning) нейрона-фаворита. Нейрон относится к кругу нейрона-фаворита, в случае если интервал между ним и нейроном-победителем на карте менее конкретной величины (в ходе преподавания меняются веса нейронов, однако их позиция на карте никак не меняется). Такого рода метод именуют методом вида WTM (Winner Takes Most — чемпион приобретает большее). В традиционном методе веса нейрона-фаворита и абсолютно всех нейронов, покоящихся в границах

его окружности, подвергаются учебе (приспособлению) согласно ряду модифицированного закона Кохонена (4). Веса нейронов, пребывающих за границами окрестности, не меняются. Объем окрестности и показатель быстроты преподавания считаются функциями, значимости которых сокращаются с ростом номером цикла преподавания. Перемена в законе Кохонена заключается в том, что показатель быстроты преподавания $\eta_i(k)$ разбивается на две части: функцию соседства $\eta_i(d,k)$ и функцию скорости обучения $a(k)$ [3]:

$$\eta_i^{(k)} = \eta_i(d,k) \cdot a(k) \tag{6}$$

В качестве функции соседства применяется или константа:

$$\eta_i(d,k) = \begin{cases} \text{const}, & d_i \leq \sigma(k) \\ 0, & d_i > \sigma(k) \end{cases} \tag{7}$$

или Гауссова функция:

$$\eta_i(d,k) = e^{-\frac{d_i^2}{2\sigma(k)^2}} \tag{8}$$

При этом лучший результат получается при использовании Гауссовой функции расстояния. В (6) и (7) $d_i = \|r_i - r_{c_j}\|$ — расстояние между i -м нейроном и нейроном-победителем c_j , r_i и r_{c_j} — координаты на сетке карты i -го и победившего c_j -го нейронов, $\|r_i - r_{c_j}\|$ — расстояние между ячейками i и c_j на сетке карты. Функция $\sigma(k)$ является убывающей функцией от номера цикла обучения. Наиболее часто используется функция, линейно убывающая от номера цикла обучения.

Рассмотрим теперь функцию скорости обучения $a(k)$. Эта функция также представляет собой функцию, убывающую от номера цикла обучения. Наиболее часто используются два варианта этой функции: линейная и обратно пропорциональная от номера цикла обучения вида [3]:

$$a(k) = \frac{A}{k+B} \quad (9)$$

где:

А и В — константы.

Использование данной функции приводит к тому, что все без исключения векторы из обучаемой подборки приносят приблизительно равнозначный вклад в итог обучения. Подготовка заключается в двух главных стадиях: в начальной стадии — стадии упорядочения векторов взвешенных коэффициентов в пространстве избирается довольно большая роль быстроты преподавания и радиуса подготовки, что дает возможность разместить векторы нейронов в согласовании с распределением образцов в выборке (меняется позиция векторов нейронов в месте свойств, однако никак не в карте). В завершающей стадии — стадии опции выполняется четкого подстраивания весов, когда значимости характеристик быстроты преподавания менее первоначальных. Подготовка длится вплоть до тех пор, пока ошибка связи (5) никак не будет довольно маленькой [8].

В данной работе проведено нейросетевое моделирование деятельности 400 крупнейших компаний Российской Федерации в 2014 – 2015гг. по объему реализации продукции. В качестве базы для исследований использован ежегодный рейтинг крупнейших компаний России по объему реализации продукции. В результате моделирования построены самоорганизующиеся карты, на которых представлены кластеры, включающие исследуемые компании. Анализ этих карт, согласно статистике средних значений экономических показателей в кластерах и отраслевой структуре кластеров, позволил определить динамику развития различных отраслей экономики, а также выявить отрасли, компании которых обладают перспективами роста.

В качестве исходных данных модели выбраны следующие показатели: объем реализации в текущем году (млн. руб.), темп прироста объема реализации продукции в текущем году по сравнению с предыдущим годом

(%), прибыль до налогообложения в текущем году (млн. руб.) и чистая прибыль в текущем году (млн. руб.).

Результатом моделирования за 2014 год стало разбиение данных на 5 кластеров, представленных на самоорганизующихся картах (рис.1).

Далее были рассчитаны средние значения анализируемых показателей по каждому кластеру и отраслевая структура кластеров в 2014 г. и тоже проделано с данными за 2015г.

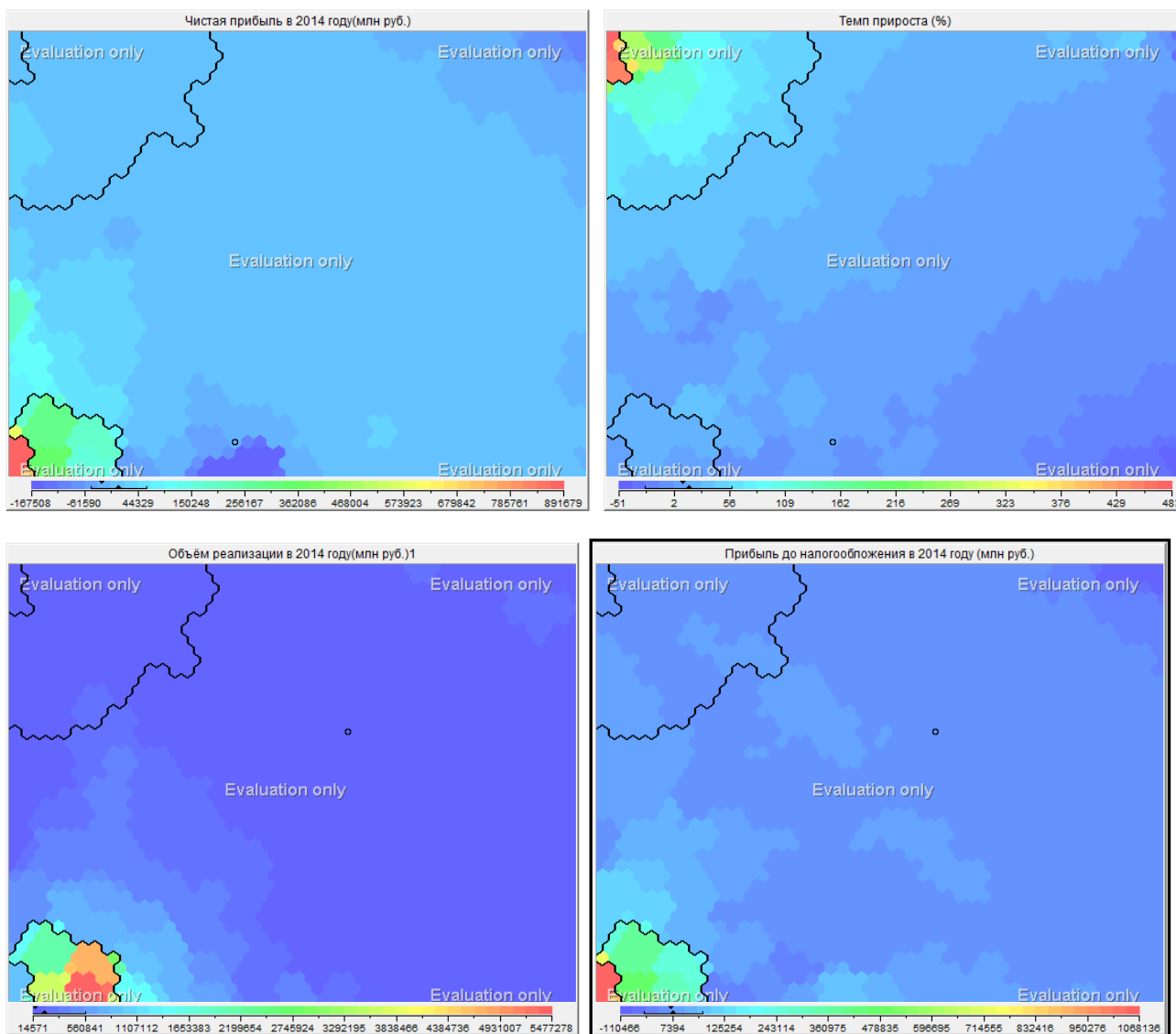


Рисунок – 1 Самоорганизующиеся карты 400 крупнейших компаний РФ за 2014 г.

Использование нейросетевого моделирования в анализе финансово-экономических показателей 400 крупнейших компаний РФ показало достаточно большое количество кластеров, а также присутствие заметных

различий в их размерах на СОК. Этот факт свидетельствует о неравномерности развития отраслей экономики Российской Федерации. В 2014 г. в бесспорные лидеры вышли компании нефтяной и нефтегазовой промышленности. На карте 2015 г. самые большие кластеры образовали компании машиностроения, химической и нефтехимической промышленности, транспорта, оптовой и розничной торговли. Они вышли на второе место по объему производства продукции и величине прибыли после явного лидера – компаний нефтяной и нефтехимической промышленности. В лидеры по темпам прироста вышли банки.

ЛИТЕРАТУРА

1. Алеников А.С., Булин Г.В. Системный инструментарий моделирования и управления транспортным кластером мегаполиса// Экономика устойчивого развития: региональный научный журнал № 3 (23). 2015 г. С. 14-27.
2. Алеников А.С., Макаров М.В. Инструментарий системного анализа и его применение при моделировании нестационарной экономики // Национальные интересы: приоритеты и безопасность №4 (241). 2014 г. С. 36-47
3. Дебок Г., Кохонен Т. Анализ финансовых данных с помощью самоорганизующихся карт. М.: АЛЬПИНА. - 2001. С. 101-104
4. Каллан Р. Основные концепции нейронных сетей. М: Вильямс. 2003. С. 137
5. Кофман А. Введение в теорию нечетких множеств. М.: Радио и связь. 1982. С. 5-25
6. Кохонен Т. самоорганизующиеся карты // труды Института инженеров по электротехнике и электронике. 1990. вып. 78. С. 19-32
7. Кузнецов Ю.А., Перова В.И. Использование нейросетевого моделирования в анализе деятельности крупнейших компаний РФ. – 2010. С. 64-67
8. Мамонова И.В. Оценка качества системного подхода к формированию налогового законодательства РФ// Налоги и финансовое

право. 2012. № 5. С. 240-244.

9. Хайкин С. Нейронные сети: полный курс. М.: Вильямс. 2006. С. 348-411.

REFERENCES

1. Alenikov A.S., Bulin G.V. Sistemnyy instrumentariy modelirovaniya i upravleniya transportnym klasterom megapolisa// *Ekonomika ustoychivogo razvitiya: regionalnyy nauchnyy zhurnal* № 3 (23). 2015 g. С. 14-27.

2. Alenikov A.S., Makarov M.V. Instrumentariy sistemnogo analiza i ego primeneniye pri modelirovanii nestatsionarnoy ekonomiki // *Natsionalnye interesy: priority i bezopasnost* №4 (241). 2014 g. С. 36-47

3. Debok G., Kokhonen T. Analiz finansovykh dannykh s pomoshchyu samoorganizuyushchikhsya kart. М.: ALPINA. - 2001. S. 101-104

4. Kallan R. Osnovnye kontseptsii neyronnykh setey. М: Vilyams. 2003. S. 137

5. Kofman A. Vvedeniye v teoriyu nechetkikh mnozhestv. М.: Radio i svyaz.1982. S. 5-25

6. Kokhonen T. samoorganizuyushchiesya karty // *trudy Instituta inzhenerov po elektrotekhnike i elektronike*. 1990. vyp. 78. S. 19-32

7. Kuznetsov Yu.A., Perova V.I. Ispolzovaniye neyrosetevogo modelirovaniya v analize deyatel'nosti krupneyshikh kompaniy RF. – 2010. S. 64-67

8. Mamonova I.V. Otsenka kachestva sistemnogo podkhoda k formirovaniyu nalogovogo zakonodatel'stva RF// *Nalogi i finansovoe pravo*. 2012. № 5. S. 240-244.

9. Khaykin S. Neyronnye seti: polnyy kurs. М.: Vilyams. 2006. S. 348-411.

*THE USING OF NEURAL NETWORK MODELING IN THE ANALYSIS OF THE
LARGEST COMPANIES OF THE RUSSIAN FEDERATION*

A.O. KUZNETSOVA

*Kuban State University,
149, Stavropolskaya, st., Krasnodar, Russian Federation, 350040;
e-mail: floppyboom@gmail.com*

Concept of the method of neural network modeling based on the Kohonen's self-organizing maps considered in this article. The analysis conducted on the 400 largest companies of Russia by volume of production with the help of the software package Viscovery SOMine, based on self-organizing maps, aimed at mapping the present situation of companies in the ranking, predicting the most profitable sectors of the economy for further competent management.

Key words: neuron, neural network modeling, network of Kohonen, self-organizing map, clustering, network training, sales volume, growth rate.