

К вопросу использования самоорганизующейся карты Кохонена для обработки анализируемых данных

В.А. Жжонов, В.А. Евсина, С.Н. Широкова

*Южно-Российский государственный политехнический университет (НПИ)
имени М.И. Платова, Новочеркасск*

Аннотация: В процессе ведения информационной деятельности скапливается большой набор данных, который отражает специфические особенности производимой работы. Не всегда хранимая информация находится в упорядоченном и понятном виде, благодаря чему с ней очень тяжело работать. Это затрудняет выполнение анализа, увеличивает сроки обработки. Решить эту проблему в силах нейросети. Сегодня нейронные сети повсеместно используются во многих сферах деятельности, за счёт их применения, например, появляется возможность более тщательно анализировать ситуацию на рынке и принимать соответствующие решения, прямо влияющие на успех. Благодаря использованию нейросети можно привести набор информации в удобный для анализа вид. В статье будет приведен перечень информации о самоорганизующейся карте Кохонена, касающийся принципов работы нейронной сети. Рассмотрена обработка тестовых данных с визуализацией карт.

Ключевые слова: самоорганизующаяся карта Кохонена, сети Кохонена, нейронные сети, кластер, обработка, *Self Organizing Map, SOM*.

В современном мире повсеместно используются нейросети. Они нашли свое применение во многих областях, касающихся науки, производства, искусства и даже развлечений. Разумеется, наиболее ощутимую пользу от их применения можно ощутить в сферах, которые имеют непосредственную связь с обработкой данных. Накопленные объемы информации, которые продолжают увеличиваться, нужно оперативно обрабатывать, чтобы получить от них наибольшую пользу [1]. На данный момент существует множество различных нейронных сетей, позволяющих решать множество подобных проблем. Одним из вариантов сети, который отлично подходит под задачи кластеризации, анализа, а также прогнозирования, является нейронная сеть Кохонена, которую относят к самоорганизующимся сетям [2]. Её также разделяют на слои и карты Кохонена. Слои – это сети с неупорядоченными нейронами, а карты, напротив, с упорядоченными. Метод, включающий в

себя самоорганизующуюся карту (Self Organizing Map, SOM) Кохонена, был разработан в 1982 году финским ученым Тейво Кохоненом [3].

Основной особенностью нейросети является то, что она позволяет визуализировать многомерное пространство в виде двумерного [4-5]. Такая интерпретация обеспечивает более удобное восприятие обработанной информации. На рис. 1 показана схема самоорганизующейся карты, где видно, что данные из входного вектора распределяются по всей карте, занимая при этом свое место, соответствующее определенным признакам.

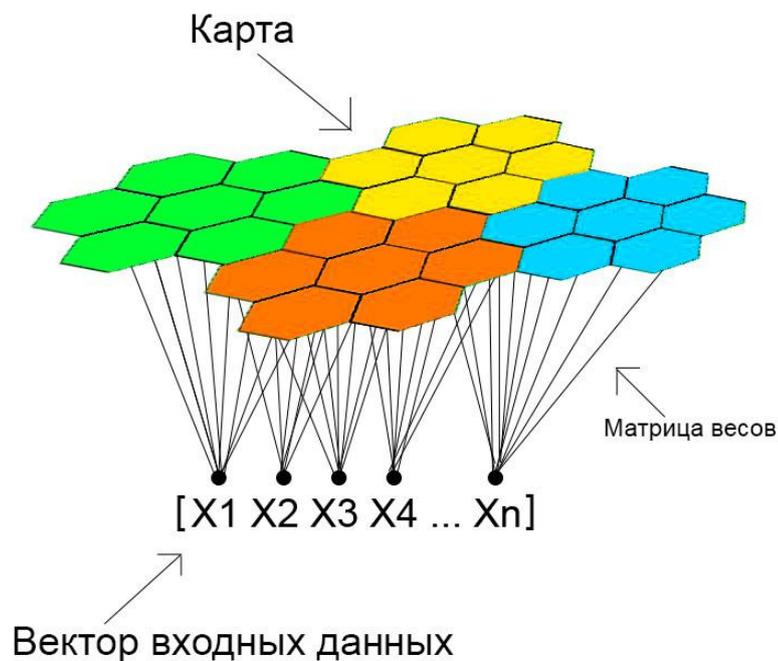


Рис. 1. – Самоорганизующаяся карта Кохонена

Сама карта состоит из ячеек, чаще всего, шестиугольной формы, что показано на рис. 2, однако, ячейки могут быть и прямоугольными. В ячейку попадают близкие по признакам объекты, после чего к ней применяется определённый цвет, который зависит от самих значений параметров объекта, присвоенного ей, а также помогает определить глубину параметра. Таким

образом для каждого отдельного параметра строится своя, соответствующая ему, двумерная карта.

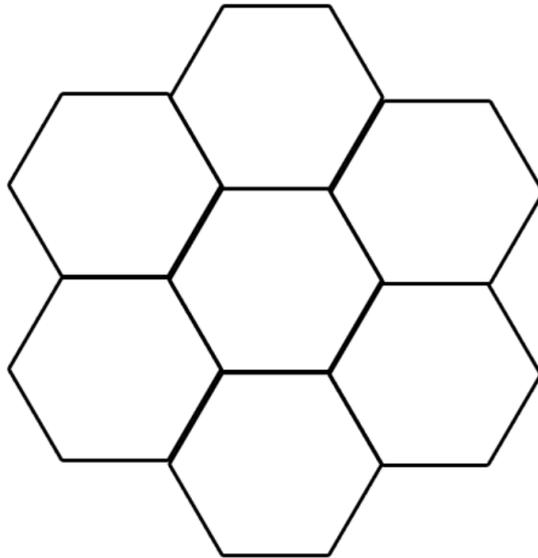


Рис. 2. – Пример ячейки карты Кохонена

В картах обучение производится без учителя и благодаря механизмам конкуренции. Это означает, что выигрывает нейрон, у которого направление весов менее всего отличается от направления входных данных. Подобный нейрон это – нейрон-победитель [6-7]. Иными словами, каждая итерация соответствует активации одного узла, где сети доступны лишь характеристики образца входных данных [8]. Вокруг него образуется радиус обучения, который определяет количество нейронов, помимо победителя, которые будут учувствовать в коррекции весов на текущей итерации. Туда войдут те нейроны, расстояние от вектора весов которых будет меньше радиуса обучения. В процессе обучения радиус уменьшается [9]. Веса нейронов, которые находятся в области радиуса, обучаются по правилу Кохонена:

$$w_i^{(k+1)} = w_i^{(k)} + n_i^{(k)} [x - w_i^{(k)}],$$

где x – входной вектор, k – номер цикла, n – коэффициент скорости обучения. Сам процесс обучения состоит из двух фаз:

- 1) Выбор наибольшего значения скорости и радиуса обучения
- 2) Точная настройка весов

Первая фаза применяется для начального распределения векторов нейронов, по основным направлениям в выборке данных, а заключительная фаза необходима для корректировки весов, при меньшей скорости обучения нейросети. Само обучение производится, пока погрешность сети не будет являться малой величиной:

$$E = \frac{1}{p} \sum_{i=1}^p \|x_i - w_j\|^2,$$

где p – количество элементов в наборе входных данных, w – вектор весов нейрона победителя.

Стоит также отметить, что входные значения следует нормировать для наиболее правильного обучения. Это можно сделать с помощью формулы:

$$x_{ij} = \frac{x_i}{|x_i|},$$

где x_{ij} - нормированное значение вектора.

Рассмотрим построение карты Кохонена на примере тестовых данных, представленных в таблице 1.

Таблица №1

Тестовый набор данных для анализа

Количество билетов	Длительность	Регион	Период
1	2	3	4
15	2	Европа	01.сен
16	4	Азия	02.сен
17	5	Европа	03.сен
18	5	Европа	04.сен
19	4	Европа	05.сен
20	6	Европа	06.сен
22	4	Азия	08.сен
23	1	Америка	09.сен
24	4	Азия	10.сен

Продолжение таблицы №1

1	2	3	4
25	8	Азия	11.сен
26	6	Америка	12.сен
27	5	Америка	13.сен
28	4	Америка	14.сен
29	8	Европа	15.сен
30	4	Америка	16.сен
31	2	Америка	17.сен
32	5	Европа	18.сен
33	1	Америка	19.сен

Для работы с данными и построения используем инструменты аналитического пакета *Deductor Academic* [10]. Определим количество билетов, длительность, регион в качестве входных параметров для обучения, представлено на рис. 3.



Рис. 3. – Входной набор данных

Далее произведем построение карт входов для данных. Итого сформировалось 4 карты, отражающие параметры, которые были загружены для обработки в виде кластеров, отмеченных на каждой карте черным контуром. Выделяя ячейку на одной из карт, можно увидеть, какие ячейки на других картах будут соответствовать ей (представлено на рис. 4).

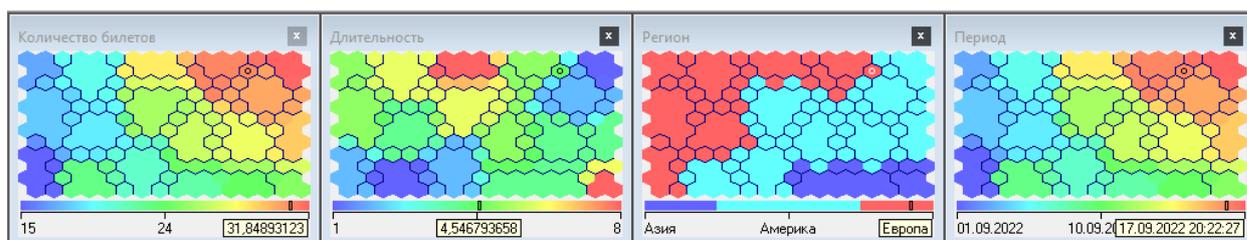


Рис. 4. – Сформированные карты входов

Итак, возможности, предоставляемые картами Кохонена, позволяют значительно упростить анализ информации, путем вывода их в удобной для работы с данными форме.

Литература

1. Питкевич П.И. Методы объединения, сокращения размеров и обработка больших данных // Инженерный вестник Дона, 2022, №7. URL: ivdon.ru/ru/magazine/archive/n12y2021/7338.
2. Кохонен, Т. Самоорганизующиеся карты. М.: Лаборатория знаний, 2021. 660 с.
3. Хайкин С. Нейронные сети: полный курс. М.: Издательский дом «Вильямс», 2006. 1104 с.
4. Алёшин С.П. Бородин Е.А. Нейросетевое распознавание классов в режиме реального времени // Инженерный вестник Дона, 2022, №7. URL: ivdon.ru/ru/magazine/archive/n1y2013/1494.
5. Achraf Khazri. Self Organizing Maps (Kohonen's maps) // Towards Data Science, 2019 URL: towardsdatascience.com/self-organizing-maps-1b7d2a84e065
6. Рутковская Д., Пилиньский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы. М.: Горячая линия – Телеком, 2006. 383 с.
7. Тархов, Д. А. Нейронные сети. Модели и алгоритмы. М. : Радиотехника, 2005. 256 с.
8. Eklavya. Kohonen Self-Organizing Maps A special type of Artificial Neural Network // Towards Data Science, 2019. URL: towardsdatascience.com/kohonen-self-organizing-maps-a29040d688da.
9. Осовский, С. Нейронные сети для обработки информации. М.: Финансы и статистика, 2002. 344 с.
10. Карта Кохонена. URL: basegroup.ru/deductor/function/algorithm/kohonen.

References

1. Pitkevich P.I. Inzhenernyj vestnik Dona, 2022, №7. URL: ivdon.ru/ru/magazine/archive/n12y2021/7338.
2. Koxonen, T. Samoorganizuyushhiesya karty. [Self Organizing Maps]. M.: Knowledge Lab, 2021. 660 p.
3. Hajkin S. Nejronnye seti: Polnyj Kurs. [Neural networks: Full Course]. M.: Williams Publishing House, 2006. 1104 p.
4. Alyoshin S.P. Borodina E.A. Inzhenernyj vestnik Dona, 2022, №7. URL: ivdon.ru/ru/magazine/archive/n1y2013/1494.
5. Achraf Khazri. Towards Data Science, 2019. URL: towardsdatascience.com/self-organizing-maps-1b7d2a84e065
6. Rutkovskaya D., Pilinskij M., Rutkovskij L. Nejronnye seti, geneticheskie algoritmy i nechetkie sistemy. [Neural networks, genetic algorithms and fuzzy systems]. M.: Hotline - Telecom, 2006. 383 p.
7. Tarxov, D. A. Nejronnye seti. Modeli i algoritmy. [Neural networks. Models and algorithms]. M.: Radio Engineering, 2005. 256 p.
8. Eklavya. Towards Data Science, 2019. URL: towardsdatascience.com/kohonen-self-organizing-maps-a29040d688da.
9. Osovskij, S. Nejronnye seti dlya obrabotki informacii. [Neural networks for information processing]. M.: Finance and statistics, 2002. 344 p.
10. Karta Koxonena. [Kohonen Map]. URL: basegroup.ru/deductor/function/algorithm/kohonen.