

# ИСПОЛЬЗОВАНИЕ НЕЙРОННЫХ СЕТЕЙ ДЛЯ РЕШЕНИЯ ЗАДАЧ КЛАССИФИКАЦИИ

**А.Ю. Лабинский, кандидат технических наук, доцент.  
Санкт-Петербургский университет ГПС МЧС России**

Рассмотрены особенности использования нейронных сетей для решения задач классификации. Представлены результаты формального описания системы классификации с использованием самоорганизующихся нейронных сетей Кохонена. Разработана компьютерная модель системы классификации, реализованная в виде программы для ЭВМ, позволяющей визуализировать результаты классификации.

*Ключевые слова:* кластерный анализ, система классификации, искусственная нейронная сеть, компьютерная программа, математическая модель

## THE SPECIAL FEATURE OF EMPLOYMENT THE NEURAL NETWORKS FOR SOLUTION THE CLASSIFICATION TASKS

A.Yu. Labinskiy. Saint-Petersburg university of State fire service of EMERCOM of Russia

This article presents the special feature of employment the neural network for solution the classification tasks. Present the result of formally description the system of classification with use the self-organizing neural networks.

*Keywords:* cluster analysis, system of classification, synthetic neural networks, computer program, mathematical model

Методом решения задачи классификации является группировка первичных данных (изучаемых объектов) [1]. Для решения задач классификации при наличии нескольких признаков, описывающих исследуемые объекты, традиционно используются методы кластерного анализа, которые отличаются от других методов многомерной классификации отсутствием обучающих выборок, то есть априорной информации о распределении значений свойств исследуемых объектов. Методы кластерного анализа (кластеризации) позволяют разделить изучаемую совокупность на группы «похожих» объектов, называемые кластерами [1]. Критерием качества классификации может служить возможность содержательной интерпретации найденных групп (кластеров). Однако методы кластерного анализа не дают какого-либо способа проверки статистической гипотезы об адекватности полученных классификаций. Иногда результаты кластеризации можно обосновать с помощью методов дискриминантного анализа [1].

В задачах кластерного анализа обычной формой представления исходных данных является прямоугольная таблица, содержащая  $N$  столбцов, элементы которой представляют собой значения, полученные в результате измерения  $N$  признаков исследуемых объектов. Другим способом представления исходных данных является матрица, содержащая элементы, числовые значения которых определяют меру близости сравниваемых объектов.

Методы кластеризации различаются способом определения меры близости между сравниваемыми объектами  $X_i$  и  $X_j$ . В кластерном анализе используются различные метрики или меры близости сравниваемых объектов [2].

*Расстояние Махаланобиса.* В случае зависимых компонент  $x_1, x_2, \dots, x_k$  вектора наблюдений  $X$  используется обобщенное расстояние Махаланобиса, вычисляемое по формуле:

$$R(X_i, X_j) = \sqrt{\{(X_i - X_j)^T * W^T / \Sigma[W * (X_i - X_j)]\}},$$

где  $W$  – некоторая симметрическая неотрицательно-определенная матрица весовых коэффициентов. Следующие меры близости сравниваемых объектов являются частными случаями расстояния Махаланобиса.

*Обычное Евклидово расстояние.* Используется для случая, когда компоненты вектора  $X$  однородны и одинаково важны для классификации и вычисляется по формуле:

$$R(X_i, X_j) = \sqrt{\Sigma(X_i - X_j)^2}.$$

*Взвешенное Евклидово расстояние.* Используется для случая, когда каждой компоненте вектора  $X$  можно приписать некоторый «вес»  $P_k$ , пропорциональный степени важности признака в задаче классификации, и вычисляется по формуле:

$$R(X_i, X_j) = \sqrt{\Sigma[P_k * (X_i - X_j)^2]}.$$

*Хеммингово расстояние.* Используется как мера различия объектов, задаваемых дихотомическими признаками, и вычисляется по формуле:

$$R(X_i, X_j) = \Sigma |X_i - X_j|.$$

Самоорганизующиеся нейронные сети Кохонена могут быть использованы для кластерного анализа только в том случае, если заранее известно число кластеров. Особенностью нейронных сетей Кохонена является устойчивость к зашумленным данным, а также быстрое обучение и возможность упрощения восприятия многомерных входных данных с помощью визуализации [3].

Нейронная сеть Кохонена обучается без учителя за счет самоорганизации. Количество нейронов сети определяется той степенью подробности, с которой требуется выполнить кластеризацию. Нейронная сеть Кохонена для наглядности может быть представлена в виде двумерного изображения (карты), в связи с чем для сетей Кохонена часто используется другое название – самоорганизующиеся карты (Self-Organizing Maps – SOM) [3]. Схема нейронной сети Кохонена представлена на рис. 1.

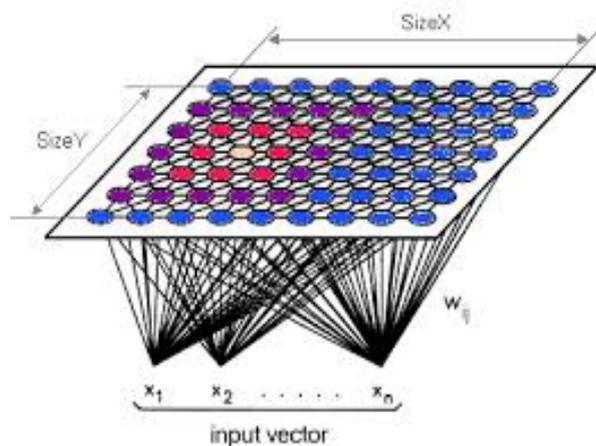


Рис. 1. Схема нейронной сети Кохонена

Нейронная сеть Кохонена должна иметь число выходов, равное числу кластеров. В качестве меры близости сравниваемых объектов используется Евклидово расстояние. В этом случае центр кластера, минимизирующий сумму мер близости для объектов этого кластера, совпадает с центром тяжести сравниваемых объектов.

Каждый нейрон в сети Кохонена запоминает один центр кластера и отвечает за определение объектов в своем кластере, то есть величина выхода нейрона тем больше, чем

ближе объект к данному центру кластера. Общее количество нейронов скрытого слоя в сети Кохонена совпадает с количеством кластеров. С помощью изменения количества нейронов скрытого слоя можно динамически менять количество кластеров.

Вектор входных данных в сети Кохонена обычно нормируется путем расположения компонентов вектора на отрезке  $[0, 1]$ . В этом случае каждый выход можно будет трактовать как вероятность того, что объект принадлежит данному кластеру.

Для запуска алгоритма классификации необходимо задать начальные значения весовых коэффициентов нейронной сети, что может быть выполнено несколькими способами, в том числе [4]:

- задание значений весовых коэффициентов путем присваивания случайных чисел;
- задание значений весовых коэффициентов путем присваивания случайных значений входных данных.

Обучение нейронной сети Кохонена происходит в процессе самоорганизации, состоящей в модификации весовых коэффициентов с использованием вектора обучающей выборки. Значения элементов вектора обучающей выборки последовательно поступают на вход нейронной сети, и каждый раз происходит корректировка весовых коэффициентов. Нейронная сеть Кохонена может обучаться и на искаженных данных входных векторов. Если искажения данных не носят систематический характер, то в процессе обучения искаженные данные сглаживаются.

Вариант алгоритма классификации с использованием сети Кохонена может быть записан следующим образом [4]:

1. Инициализация начальных значений весовых коэффициентов нейронной сети  $W_{ij}$ .
2. Цикл по количеству сравниваемых объектов.
3. Для каждого сравниваемого объекта вычисление меры близости:

$$L_j = D(W_{ij}, X_{ij}) = \|W_{ij} - X_{ij}\|,$$

где  $D$  – мера расстояния (близости).

4. Корректировка значений весовых коэффициентов нейронной сети:

$$W_{ij} = W_{ij} + \alpha * (W_{ij} - X_{ij}),$$

где коэффициент  $0 < \alpha < 1$ .

5. Минимизация целевой функции:

$$\min \sum D^2(X_{ij}, W(X_{ij})).$$

6. Конец цикла по количеству сравниваемых объектов.

Алгоритм выполняется до тех пор, пока весовые коэффициенты  $W_{ij}$  не перестанут изменяться.

### **Компьютерная модель системы классификации**

В целях решения задачи классификации была создана компьютерная модель искусственной самообучающейся нейронной сети Кохонена. С использованием разработанной модели производилась классификация путем разбиения на три кластера 144-х графических объектов размером  $10*10$  пикселей трех цветов (цветовая модель RGB), размещенных случайным образом в пределах прямоугольной области размером  $250*250$  пикселей. В качестве центров кластеров были выбраны центры абсцисс трех прямоугольных областей размером  $(250/3)*250$  пикселей.

Целью обучения искусственной нейронной сети являлось получение таких значений коэффициентов связи нейронов сети (синаптических весов), которые обеспечивают для множества значений входных данных требуемую степень минимизации целевой функции. Корректировка значений весовых коэффициентов нейронной сети производится по приведенной выше формуле.

Рассмотренная компьютерная модель искусственной нейронной сети была реализована в виде программы для ЭВМ, интерфейс которой представлен на рис. 2.

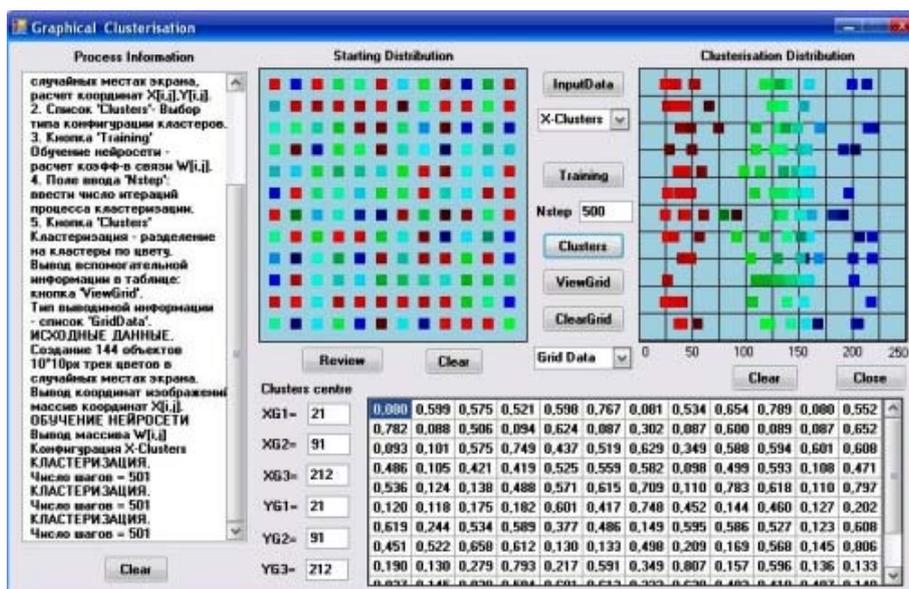


Рис. 2. Интерфейс программы классификации с помощью нейронной сети

Процесс классификации объектов трех типов (кластеризации) наглядно представлен в двух окнах графического вывода (рис. 2). В левом окне представлены исходные данные – 144 графических объекта размером 10\*10 пикселей трех цветов в цветовой модели RGB. В правом окне представлены результаты классификации путем разбиения совокупности объектов на три кластера. Программа классификации позволяет задавать несколько конфигураций областей кластеров. Примеры классификации для двух конфигураций областей кластеров представлены на рис. 3.

На левом фрагменте в качестве центров кластеров выбраны центры абсцисс трех прямоугольных областей размером (250/3)\*250 пикселей.

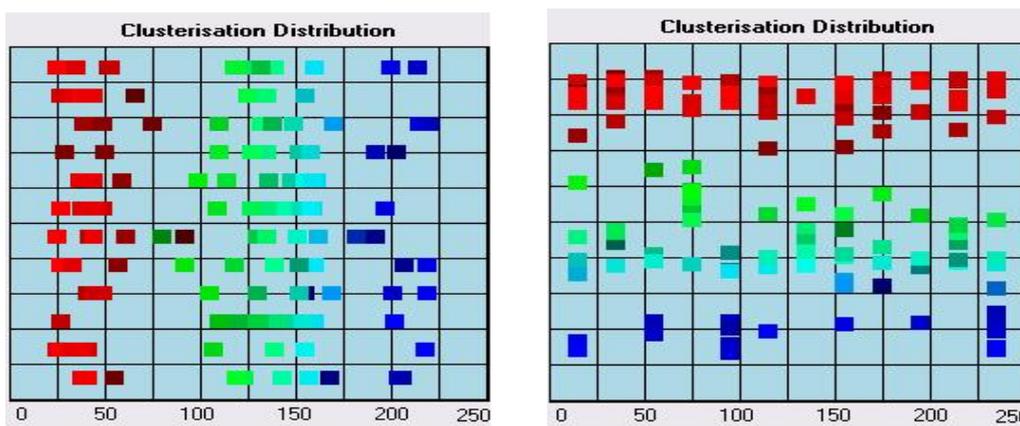


Рис. 3. Примеры кластеризации 144 графических объектов по трем признакам

На правом фрагменте в качестве центров кластеров выбраны центры ординат трех прямоугольных областей размером 250\*(250/3) пикселей.

Программа позволяет задавать квадратную конфигурацию областей кластеров. Пример кластеризации 144 объектов трех типов представлен на рис. 4.

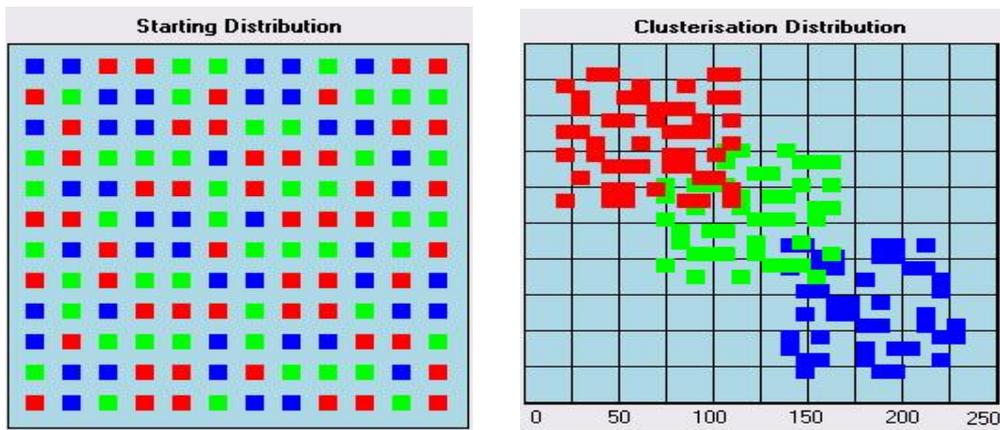


Рис. 4. Слева исходные данные – 144 объекта, справа результат классификации по трем признакам

Возможна классификация объектов шести типов. Примеры кластеризации 144 графических объектов по шести признакам представлены на рис. 5.

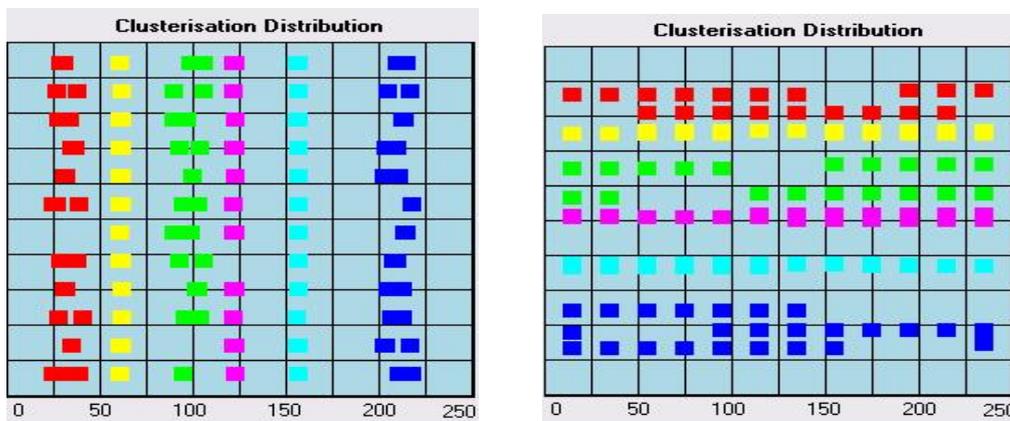


Рис. 5. Примеры кластеризации 144 объектов по шести признакам: слева вертикальная, справа горизонтальная конфигурация кластеров

Пример классификации 144 графических объектов по восьми признакам для случая квадратной конфигурации областей кластеров представлен на рис. 6.

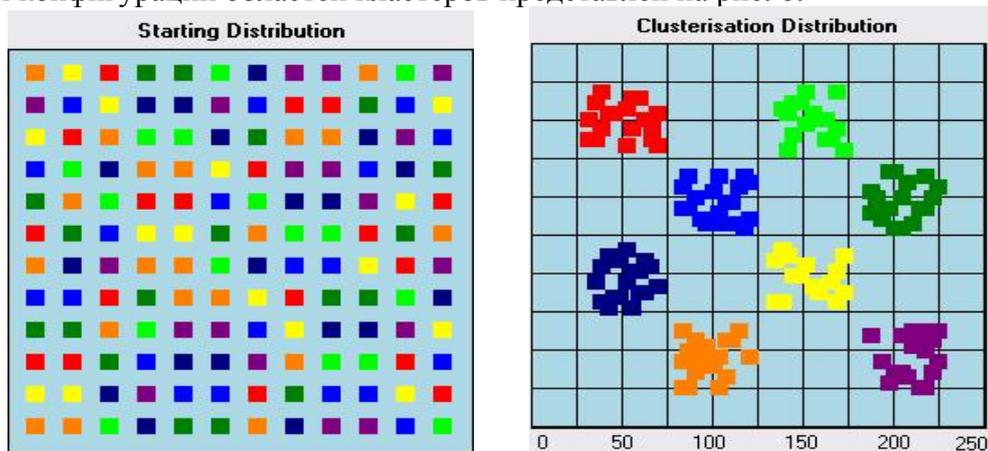


Рис. 6. Слева исходные данные – 144 объекта, справа результат классификации по восьми признакам

Разработанная программа классификации позволяет проводить классификацию до 600 объектов с максимальным количеством признаков 25.

Пример классификации 576 объектов по 24 признакам представлен на рис. 7.

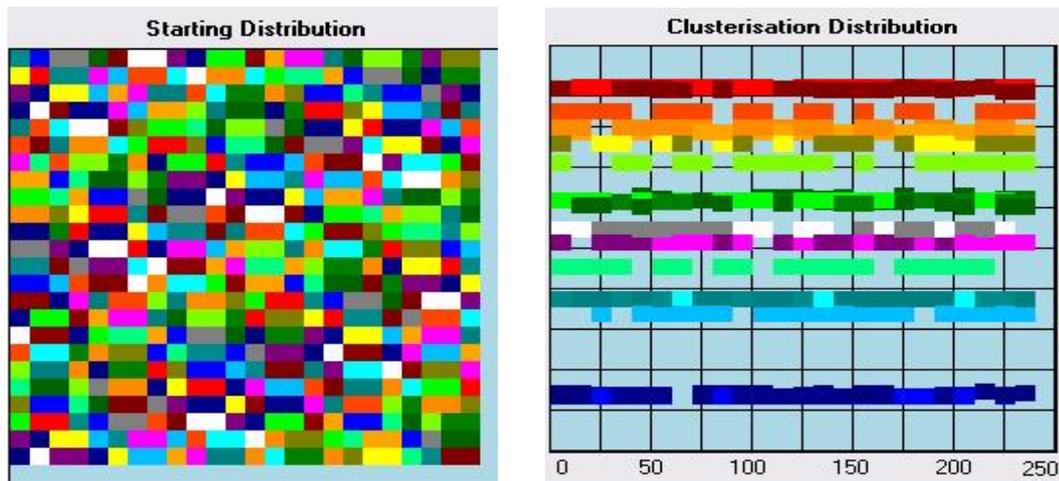


Рис. 7. Слева исходные данные – 576 объектов, справа результат классификации по 24 признакам

В задачах кластерного анализа классификационные признаки могут принимать как непрерывные, так и дискретные значения, включая бинарное представление. В случае непрерывных значений признаков необходима нормализация входных значений путем приведения к заданному диапазону (a, b), где для цветовой модели RGB: a=0, b=255. Нормализация может быть выполнена по формуле:

$$X^* = a + (b - a) * (X - X_{\min}) / (X_{\max} - X_{\min}),$$

где X – исходное входное значение; X\* – нормализованное входное значение в диапазоне (a, b). Далее необходимо определить количество предполагаемых кластеров и выполнить цветовое кодирование признаков.

В случае бинарного представления значений классификационных признаков цветовое кодирование может быть выполнено следующим образом. Пусть число признаков, условно разделенных на три группы, равно 24. Тогда RGB-компонента одной группы из восьми признаков может принимать значения в диапазоне от 0 до 255, вычисленные по формуле:

$$Z_{\text{RGB}} = \sum_{i=0}^7 a_i * 2^i,$$

где  $a_i$  – значение признака (0 или 1),  $i$  меняется от 0 до 7.

На рис. 8 представлено цветовое кодирование 24 бинарных признаков в цветовой модели RGB, выполненное в разработанной программе.

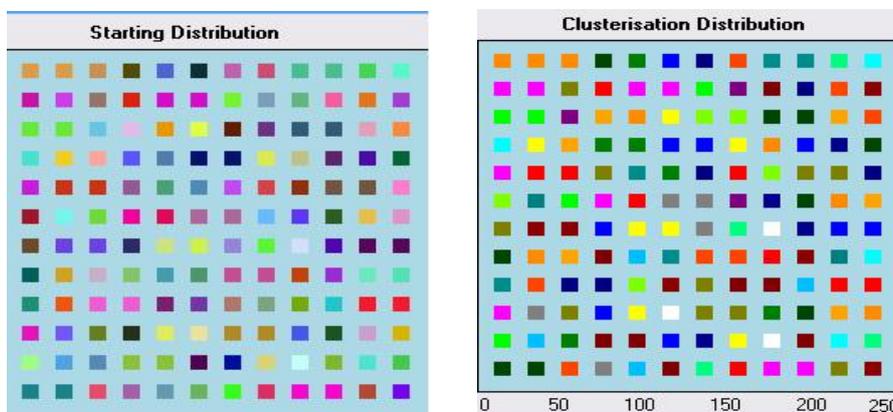


Рис. 8. Цветовое кодирование 144 объектов по 24 признакам

На рис. 8 слева представлены исходные данные в цветовой модели RGB, справа представлен результат цветового кодирования по 24 базовым цветам.

Таким образом, искусственные самообучающиеся нейронные сети Кохонена способны не только производить классификацию, но и путем преобразования n-мерного

пространства в двухмерное успешно решать проблему отображения пространственных структур большой размерности.

Результаты компьютерных экспериментов показали, что созданная компьютерная модель искусственной самоорганизующейся нейронной сети способна обеспечить классификацию до 600 графических объектов, имеющих от 3 до 24 классификационных признаков, путем разбиения на кластеры нескольких конфигураций и представлять результаты классификации в наглядном графическом виде. Компьютерная модель может производить цветное кодирование 24 бинарных признаков в цветовой модели RGB.

### **Литература**

1. Енюков И.С. Методы многомерного статистического анализа. М.: Финансы и статистика, 1986.

2. Кендэлл М., Стюарт А. Факторный, дискриминантный и кластерный анализ. М.: Финансы и статистика, 1989.

3. Хайкин С. Нейронные сети: полный курс. М.: Вильямс, 2006.

4. Ханжулин В.Г., Кудряшов Д.С. Нейронные сети Кохонена в кластерном анализе данных. М., 2011.