

Харахинов В.А., аспирант,
Сосинская С.С., канд. техн. наук, доц.
Иркутский государственный технический университет

РАЗРАБОТКА ИНСТРУМЕНТА ДЛЯ АНАЛИЗА РЕЗУЛЬТАТОВ МЕТОДИЧЕСКОЙ И НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЫ ПРЕПОДАВАТЕЛЕЙ С ПОМОЩЬЮ НЕЙРОННЫХ СЕТЕЙ

kharakhinov@gmail.com

В наши дни практически во всех областях деятельности требуется накапливать информацию в базах данных. В сфере высшего образования необходимо учитывать все виды деятельности преподавателей: учебную, методическую, научно-исследовательскую. Была разработана единая база данных, содержащая 13 таблиц, которая хранит полную информацию о каждом преподавателе каждой кафедры (количество написанных им монографий, статей, учебных пособий, количество выступлений на научных конференциях, количество полученных патентов и т.д.). Однако организация хранения информации – это лишь один из этапов, вслед за которым должен следовать этап анализа. В данной статье описан один из способов проведения кластерного анализа данных. Рассматривается задача обучения самоорганизующейся нейронной сети (сети Кохонена) для проведения кластерного анализа результатов деятельности преподавателей. Реализовано приложение, позволяющее наглядно отобразить результаты кластерного анализа данных.

Ключевые слова: нейронная сеть Кохонена, кластерный анализ.

Введение. Кластерный анализ – многомерная статистическая процедура, выполняющая сбор данных, содержащих информацию о выборке объектов, и затем упорядочивающая объекты в сравнительно однородные группы [2]. Существует довольно большое количество методов проведения кластерного анализа, например, при помощи искусственных нейронных сетей.

Нейронные сети (Neural Networks) – это модели биологических нейронных сетей мозга, в которых нейроны имитируются относительно простыми, часто однотипными, элементами (искусственными нейронами). Искусственный нейрон – элемент искусственных нейронных сетей, моделирующий некоторые функции биологического нейрона. Главная функция искусственного нейрона – формировать выходной сигнал в зависимости от сигналов, поступающих на его входы [1,3].

В самой распространенной конфигурации входные сигналы обрабатываются адаптивным сумматором, затем выходной сигнал сумматора поступает в нелинейный преобразователь, где преобразуется функцией активации, и результат подается на выход.

Методология. Традиционно принято считать, что наиболее подходящим типом сети для выполнения кластерного анализа является сеть Кохонена [1,3]. На рис. 1 представлена архитектура слоя Кохонена.

Здесь использован блок ndist для вычисления отрицательного евклидова расстояния между вектором входа p и строками матрицы

сов IW^{11} . Вход функции активации n^1 – это результат суммирования вычисленного расстояния с вектором смещения b . Если все смещения нулевые, максимальное значение n^1 не может превышать 0. Нулевое значение n^1 возможно только тогда, когда вектор входа p оказывается равным вектору веса одного из нейронов. Если смещения отличны от 0, то возможны и положительные значения для элементов вектора n^1 [1].

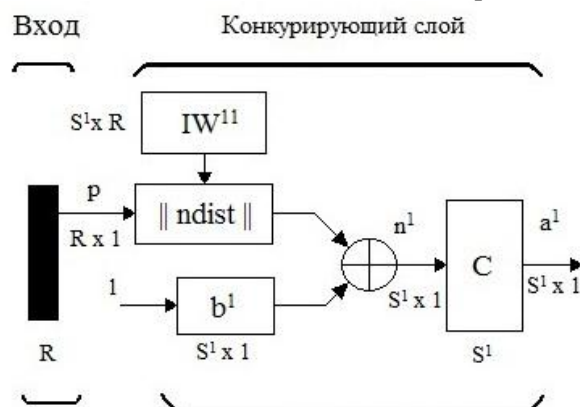


Рис. 1. Архитектура слоя Кохонена

Конкурирующая функция активации анализирует значения элементов вектора n^1 и формирует выходы нейронов, равные 0 для всех нейронов, кроме одного нейрона-победителя, имеющего на входе максимальное значение. Таким образом, вектор выхода слоя a^1 имеет единственный элемент, равный 1, который соответствует нейрону-победителю, а остальные равны 0. Такая активационная характеристика может быть описана следующим образом:

$$a_i^1 = \begin{cases} 1, i = i^*, i^* = \arg(\max n_i^1); \\ 0, i \neq i^*. \end{cases} \quad (1)$$

где i^* - номер активного нейрона, который определяет тот кластер, к которому наиболее близок входной вектор, а \arg - аргумент конкурирующей функции активации.

Отличительной чертой процесса обучения данной сети является то, что необходимо настроить веса синапсов нейронов, а не минимизировать ошибку обучения. Обучение происходит по правилу Кохонена:

$${}_iIW_{11}(q) = {}_iIW_{11}(q-1) + \alpha(p(q) - {}_iIW_{11}(q-1)) \quad (2)$$

Правило Кохонена представляет собой рекуррентное соотношение, которое обеспечивает коррекцию строки i матрицы весов добавлением взвешенной разности вектора входа и значения строки на предыдущем шаге. Таким образом, вектор веса, наиболее близкий к вектору входа, модифицируется так, чтобы расстояние между ними стало еще меньше. Результат такого обучения будет заключаться в том, что победивший нейрон, вероятно, выиграет конкуренцию и в том случае, когда будет представлен новый входной вектор, близкий к предыдущему, и его победа менее вероятна, когда будет представлен вектор, существенно отличающийся от предыдущего. Когда на вход сети поступает все большее и большее число векторов, нейрон, являющийся ближайшим, снова корректирует свой весовой вектор. В конечном счете, если в слое имеется достаточное количество нейронов, то каждая группа близких векторов окажется связанной с одним из нейронов слоем [1].

Основная часть. Искусственную нейронную сеть можно реализовать при помощи практически любого современного языка программирования, однако, пакет MATLAB позволяет заметно облегчить этот процесс.

Реализация любой нейронной сети состоит из следующих этапов [3]:

- Сбор данных для обучения;
- Подготовка и нормализация данных;
- Выбор топологии сети;
- Экспериментальный подбор характеристик сети;
- Экспериментальный подбор параметров обучения;
- Обучение сети;
- Проверка адекватности обучения;
- Корректировка параметров, окончательное обучение.

Выполнить сбор данных из реально существующей базы данных, фиксирующей деятельность преподавателей НИ ИрГТУ (количество изданных монографий, пособий и т.д.), не пред-

ставлялось возможным, так как приложение не внедрено в университете. В связи с этим было сгенерировано необходимое количество записей в системе MATLAB по каждой из таблиц базы данных. В ходе генерации были учтены все аспекты логической модели данных. На рис. 2 представлена часть таблицы, содержащая сгенерированные записи. Всего генерируется 200 записей о преподавателях, которые равномерно распределяются по кафедрам. Подобным образом генерируются записи и для других таблиц базы данных. Например, генерируется около 2000 записей по статьям, учитывается возможность написания статьи в соавторстве. Таким образом, создается результирующая матрица, которая содержит данные о кафедрах, а именно, суммарное количество изданных статей, монографий, учебных пособий, количество выступлений на конференциях, а также суммарное количество полученных патентов преподавателями кафедры за тот или иной рассматриваемый период. Общий период рассмотрения был взят с 1991 по 2013 годы, однако, это не так важно, поскольку данные были сгенерированы.

Данные из таблиц образуют обучающую выборку.

Приложение, выполняющее кластерный анализ, было реализовано при помощи GUI-интерфейса. На рис. 3 представлена готовая форма приложения.

Готовая форма включает в себя стандартные компоненты среды Guide, которая служит инструментом для создания графических приложений в среде MATLAB. В реализованном приложении пользователь может выбрать признаки, значения которых являются входом нейронной сети Кохонена. Также пользователь может выбрать временной интервал.

Задача кластеризации сходна с задачей классификации, является ее логическим продолжением, но ее отличие в том, что классы изучаемого набора данных заранее не предопределены. Таким образом, кластеризация предназначена для разбиения совокупности объектов на однородные группы. Если данные выборки представить как точки в признаковом пространстве (в реализованном приложении можно выбрать лишь 2 признака), то задача кластеризации сводится к определению "сгущений точек".

Необходимо экспериментально подобрать параметр обучения - количество итераций. Было выбрано разбиение объектов на 3 кластера. В данном случае было использовано 50 итераций обучения, на рис. 4 показаны результаты разбиения кафедр, вошедших в обучающую выборку, на классы после обучения.

| | 1 | 2 | 3 | 4 | 5 | 6 |
|----|----|---------------|---------------|----------------|------------|------|
| 1 | 1 | 'Ульянов М... | 'Преподава... | 'Профессо... | 'Кандидат' | 8 |
| 2 | 2 | 'Сидоров В... | 'Профессор' | 'Доцент по ... | 'Доктор' | 66 |
| 3 | 3 | 'Ульянов Л... | 'Преподава... | 'Профессо... | 'Доктор' | 34 |
| 4 | 4 | 'Кузнецов ... | 'Доцент' | 'Профессо... | 'Доктор' | 62 |
| 5 | 5 | 'Кузнецов ... | 'Преподава... | 'Профессо... | 'Кандидат' | 4191 |
| 6 | 6 | 'Прохоров ... | 'Старший ... | 'Профессо... | 'Доктор' | 52 |
| 7 | 7 | 'Кузнецов ... | 'Преподава... | 'Доцент по ... | 'Кандидат' | 12 |
| 8 | 8 | 'Титов Ром... | 'Ведущий н... | 'Профессо... | 'Кандидат' | 4 |
| 9 | 9 | 'Петров Фе... | 'Старший ... | 'Профессо... | 'Кандидат' | 22 |
| 10 | 10 | 'Усов Рома... | 'Профессор' | 'Доцент по ... | 'Кандидат' | 27 |

Рис. 2. Часть таблицы «teacher»

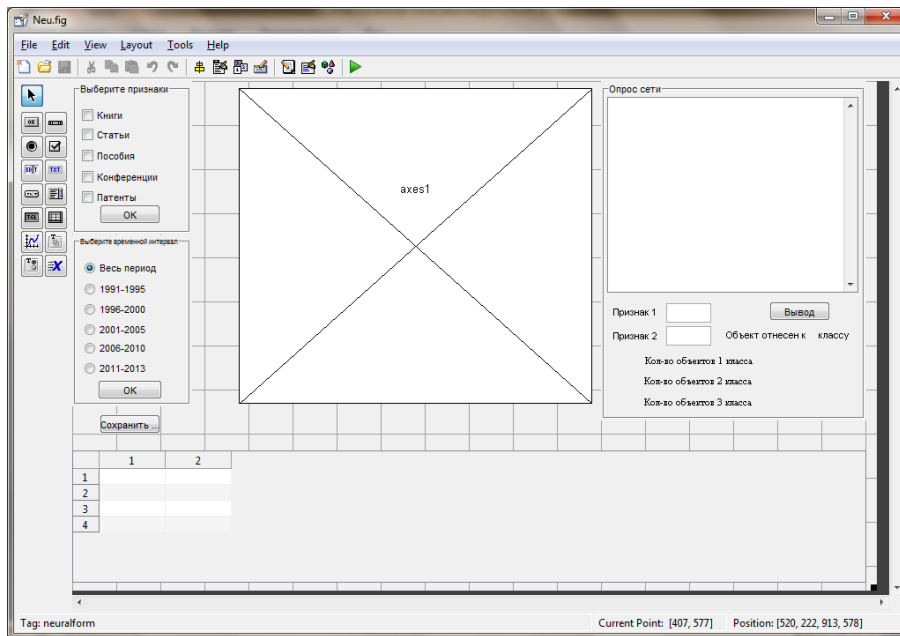


Рис. 3. Форма GUI-приложения

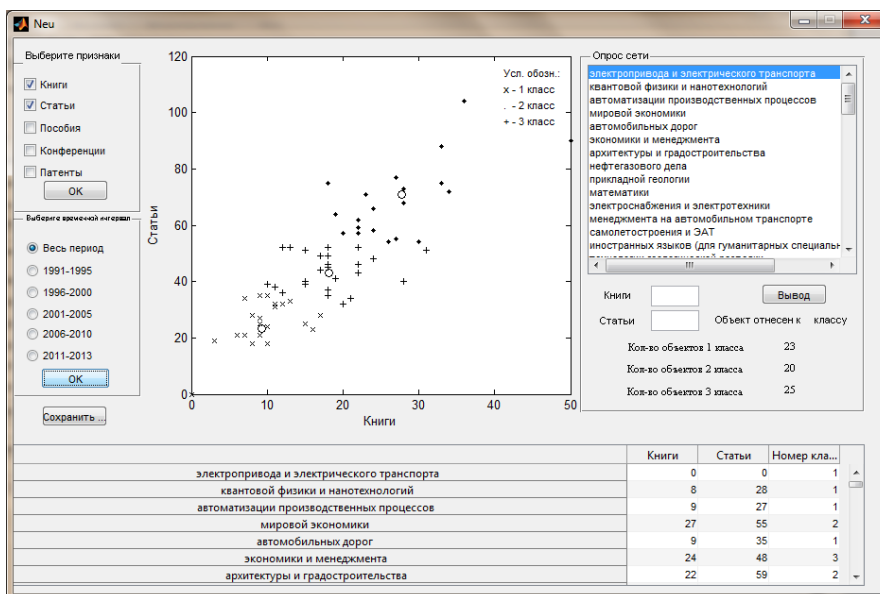


Рис. 4. Результат проведения кластерного анализа

Анализ проводился по изданным монографиям и статьям за весь рассматриваемый период (1991-2013). На графике присутствуют центры кластеров (окружности). Также отображены кафедры (представлены точками с разными пиктограммами). Это позволяет наглядно показать, к какому классу принадлежит та или иная кафедра. Кроме того, результаты проведения кластерного анализа отображаются в виде таблицы. Есть возможность выполнить классификацию кафедр, то есть провести опрос сети. Например, были выбраны кафедра вычислительной техники, кафедра автоматизированных систем и кафедра начертательной геометрии и технического черчения. Информация об этих кафедрах автоматически поступает на вход нейронной сети при выборе их из списка, после чего система

решает к какому классу отнести кафедру. На рис. 5 видно, что кафедра вычислительной техники имеет суммарно 18 изданных книг и 45 опубликованных статей за взятый период. В данном случае кафедра была отнесена к 3 классу, который можно охарактеризовать как «средняя кафедра». На рис 6 приведен результат опроса сети с данными о кафедре начертательной геометрии и технического черчения, система отнесла объект к 1 классу, посчитав его менее «успешным» по выбранным признакам, то есть класс 1 – это «неуспешные» кафедры. Наконец, на рис. 7 отображен результат опроса по кафедре автоматизированных систем, объект отнесен ко 2 классу (класс лучших кафедр по выбранным признакам).

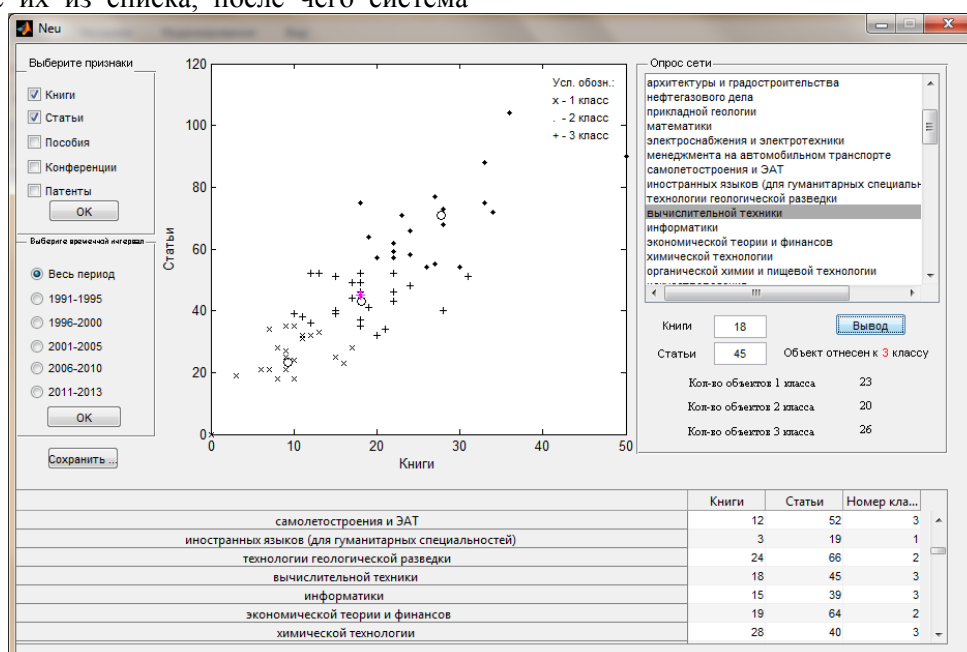


Рис. 5. Классификация кафедры вычислительной техники

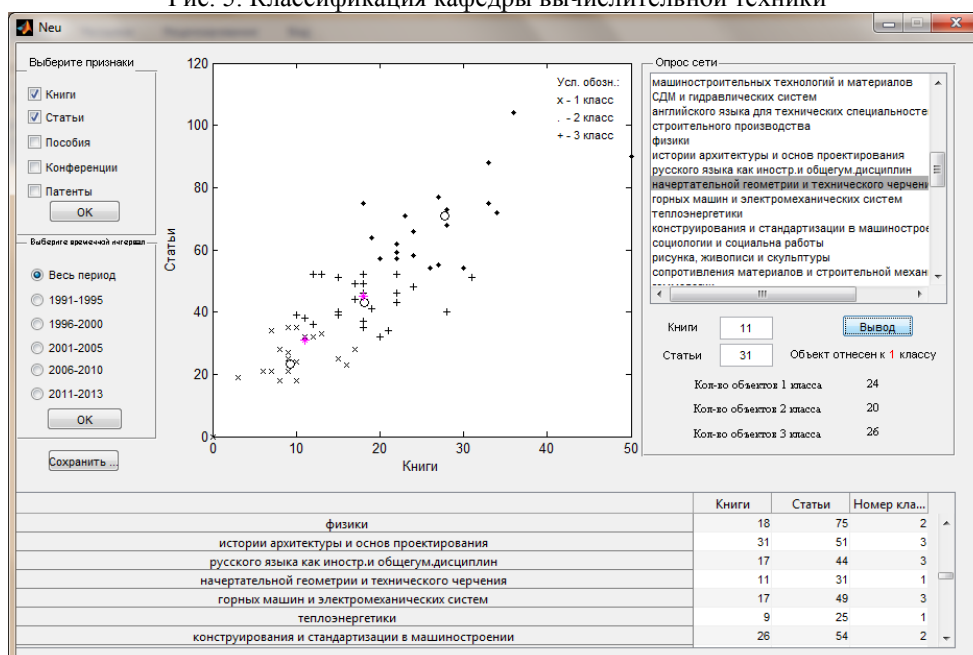


Рис. 6. Классификация кафедры начертательной геометрии и технического черчения

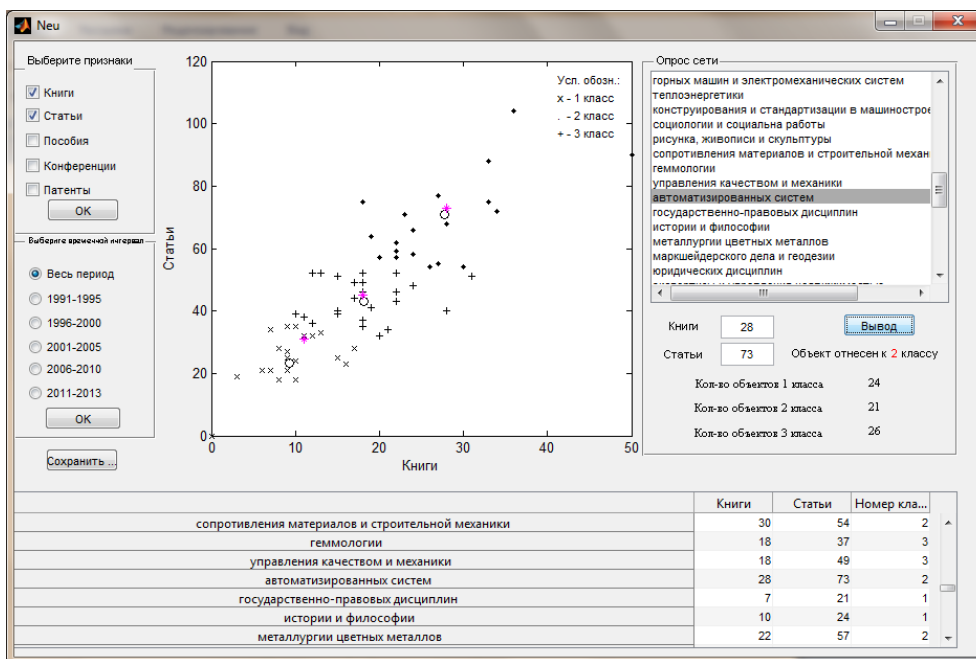


Рис. 7. Классификация кафедры автоматизированных систем

Выводы. Исходя из полученных результа-

тов, можно сделать выводы:

1. Аппарат искусственных нейронных сетей позволяет классифицировать знания о научно-исследовательской и методической деятельности кафедр;
2. Реализованное приложение позволяет автоматизировать процесс анализа результатов методической и научно-исследовательской деятельности преподавателей, сделать его более объективным.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Медведев В.С., Потемкин В.Г. Нейронные сети. MATLAB 6. М.: ДИАЛОГ-МИФИ, 2002. 496 с. ISBN 5-86404-135-1.
2. Дюран Б., Оделл П. Кластерный анализ / пер. с англ. Е.З. Демиденко. М.: Статистика, 1977. 128 С.
3. Каллан Р. Основные концепции нейронных сетей / пер. с англ. А.Г. Сивака. М.: ИД «Вильямс», 2001. 288 С.