

НЕЙРОННЫЕ СЕТИ КОХОНЕНА И НЕЧЕТКИЕ НЕЙРОННЫЕ СЕТИ В ИНТЕЛЛЕКТУАЛЬНОМ АНАЛИЗЕ ДАННЫХ

Манжула В.Г., Федяшов Д.С.

ГОУ ВПО «Южно-Российский государственный университет
экономики и сервиса», Шахты,
e-mail: mail@sssu.ru

Во введении рассмотрена актуальность использования нейронных сетей в интеллектуальном анализе данных. В первом разделе описан метод нейронных сетей в интеллектуальном анализе данных. Второй раздел представляет процесс анализа данных, основанный на искусственных нейронных сетях. В третьем разделе рассматриваются конкретные типы нейронных сетей, используемых в интеллектуальном анализе данных, а именно нейронные сети Кохонена и нечеткие нейронные сети. Выявлены достоинства и недостатки данных сетей. В заключении даны рекомендации по использованию рассмотренных нейронных сетей в технологиях интеллектуального анализа данных.

Ключевые слова: Data Mining, интеллектуальный анализ данных, нейронные сети, сети Кохонена, нечеткие нейронные сети

Введение. Развитие технологии баз данных и систем управления базами данных, способствует росту объема данных, хранящихся в базе. Эти данные содержат в себе много важной информации, которая имеет большой потенциал для прибыли. Ввиду этого многие компании используют технологию интеллектуального анализа данных (data mining), которая позволяет обрабатывать массивные базы данных и извлекать из них полезную информацию.

Задачей интеллектуального анализа данных является выявление латентных правил и закономерностей в наборах данных. Продолжительное время основным инструментом интеллектуального анализа данных была традиционная математическая статистика, но и она зачастую не в состоянии решить задачи из реальной жизни. Математическая статистика в основном полезна при проверке заранее сформулированных гипотез (verification-driven data mining) [3].

Первоначально применение нейронных сетей в интеллектуальном анализе данных вызывало скептическое отношение, ввиду недостатков, присущих нейронным сетям: сложная структура, плохая интерпретируемость и

долгое время обучения [8]. Однако их преимущества, такие как, высокая допустимость к зашумленным данным и низкий коэффициент ошибок, непрерывное усовершенствование и оптимизация различных алгоритмов обучения сетей, алгоритма извлечения правил, алгоритма упрощения сетей, делают нейронные сети все более и более перспективным направлением в data mining [8].

Области применения нейронных сетей обширны – автоматизация процессов распознавания образов, прогнозирование, адаптивное управление, создание экспертных систем, организация ассоциативной памяти, обработка аналоговых и цифровых сигналов, синтез и идентификация электронных цепей и систем [2].

Таким образом, можно сказать, что использование нейронных сетей в технологии интеллектуального анализа данных является актуальным направлением, которое непрерывно развивается, по пути устранения недостатков.

Метод нейронных сетей в Data Mining

Как правило, выделяют следующие существующие методы интеллектуального

анализа данных [6]: нейронные сети, деревья решений, генетические алгоритмы, нечеткая логика, алгоритмы ограниченного перебора, эволюционное программирование, системы рассуждения на основе аналогичных случаев, индукция правил, анализ с избирательным действием, логическая регрессия, алгоритмы определения ассоциаций и последовательностей, визуализация данных, комбинированные методы.

Большинство аналитических методов в технологии data mining – это известные математические алгоритмы и методы. Новым в их применении является адаптация их для решения тех или иных конкретных проблем, возможная благодаря появившимся техническим возможностям и программным средствам.

Основная часть методов data mining была разработана в рамках теории искусственного интеллекта.

Метод нейронных сетей [8] используется для классификации, кластеризации, прогнозирования и распознавания образов. Модель нейронной сети может быть разделена на три типа:

1) сети прямого распространения (backpropagation): одна из наиболее распространенных архитектур, в основном используется в таких областях, как прогнозирование и распознавание образов;

2) сети с обратной связью: такие, как дискретная модель Хопфилда, в основном используется для оптимизации вычислений и ассоциативной памяти;

3) самоорганизующиеся сети: включают модели адаптивной резонансной теории (ART) и модели Кохонена, в основном используется для кластерного анализа.

В настоящее время при анализе в data mining используются нейронные сети прямого распространения. Искусственные нейронные сети – активно развивающееся направление науки, но до сих пор некоторые теории окончательно не сформировались, такие как проблемы сходимости, устойчивости, локальный минимум и корректировка параметров. Для сети прямого

распространения частот встречающихся проблемы – обучение происходит медленно, она может попасть в локальный минимум и трудно определить параметры обучения.

Ввиду этих проблем многие перешли к методу объединения искусственных нейронных сетей с генетическими алгоритмами и достигли лучших результатов.

Одно из главных преимуществ нейронных сетей [3] состоит в том, что они, по крайней мере, теоретически могут аппроксимировать любую непрерывную функцию, что позволяет исследователю не принимать заранее какие-либо гипотезы относительно модели. К существенным недостаткам нейронных сетей можно отнести тот факт, что окончательное решение зависит от начальных установок сети и его практически невозможно интерпретировать в традиционных аналитических терминах.

Процесс анализа данных, основанный на нейронной сети

Процесс анализа данных (data mining) может быть представлен тремя основными фазами [8]: подготовка данных, анализ данных, выражение и интерпретация результатов. Подробнее показано на рис. 1.

Интеллектуальный анализ данных (data mining), основанный на нейронной сети, состоит из [8]: подготовки данных, извлечения правил и оценки правил, то есть трех этапов, как показано на рис. 2.

• Подготовка данных

Процесс подготовки данных должен определить и обработать добываемые данные, чтобы сделать их пригодными для конкретных методов интеллектуального анализа. Подготовка данных является первым важным шагом на пути интеллектуального анализа (data mining) и играет в нем решающую роль. Как правило, подготовка данных включает в себя четыре процесса:

1. Очистка данных

Очистка данных должна заполнить вакантные значения данных, устранить зашумленные данные и исправить несогласованность в данных.



Рис. 1. Общий процесс анализа данных

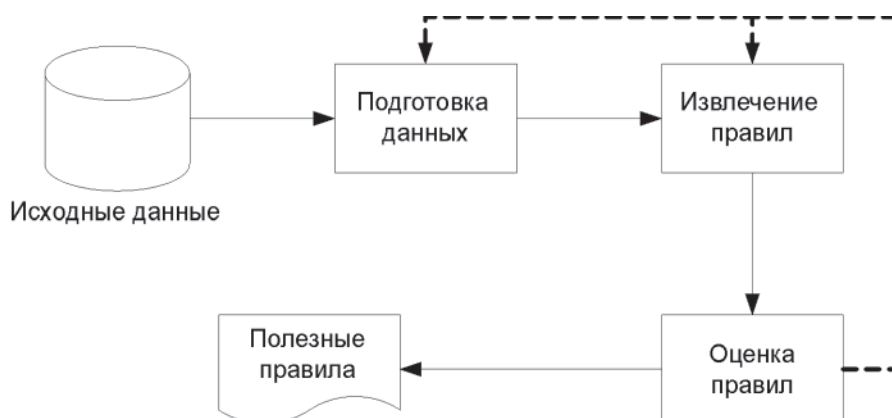


Рис. 2. Процесс анализа данных, основанный на нейронной сети

2. Выбор данных

Выбор данных должен определить расположение используемых в данном анализе данных.

3. Предварительная обработка данных

Предварительная обработка данных является расширением процесса очистки данных, которые были выбраны.

4. Выражение данных

Выражение данных должно преобразовать данные после предварительной обра-

ботки в форму, которая может быть принята по условию алгоритма анализа данных, основанного на нейронной сети.

Анализ данных, основанный на нейронной сети, может работать только с числовыми данными, из чего следует, что необходимо преобразовывать символьные данные в числовые. Простейший способ заключается в создании таблицы соответствий между символьными данными и числовыми. Другой, более сложный подход

заключается в принятии хэш-функций для создания уникальных числовых данных, соответствующих данной строке. Несмотря на то, что в реляционной базе данных есть множество типов данных, все они в основном могут быть приведены к символьным, дискретным числовым и непрерывным числовым данным, то есть три логических

типа данных. Например, слово «Яблоко» на рис. 3 может быть преобразовано в соответствующие дискретные числовые данные при использовании таблицы символов или хэш-функции. Затем дискретные числовые данные могут быть определены количественно в непрерывные числовые данные, а также могут быть зашифрованы.

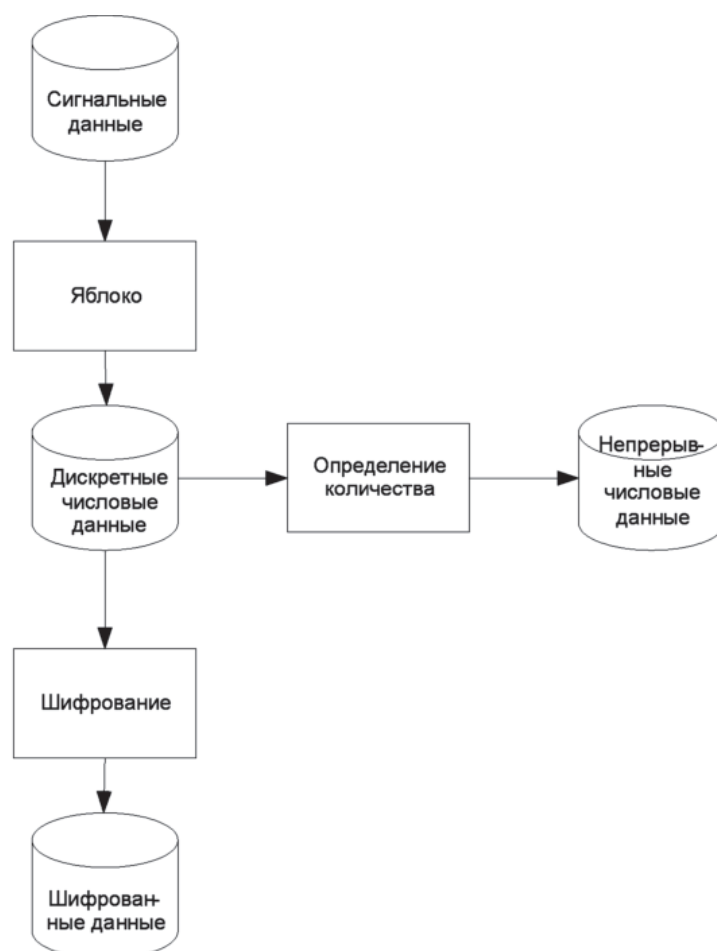


Рис. 3. Выражение данных и преобразования в анализе данных на основе нейронной сети

• *Извлечение правил*

Существует множество методов извлечения правил, среди которых наиболее часто используются LRE (Limited Relative Error) метод, метод черного ящика, метод извлечения нечетких правил, метод извлечения правил из рекурсивной сети, алгоритм извлечения правил двоичного входа и выхода (BIO-RE), алгоритм частичного извлечения правил (Partial-RE) и алгоритм полного извлечения правил (Full-RE).

• *Правила оценки*

Несмотря на то что цель правил оценки зависит от конкретного применения, в общем они могут быть оценены в соответствии со следующими задачами:

- 1) найти оптимальную последовательность извлечения правил. Сделав это, получим лучшие результаты в ряде определенных данных;
- 2) проверить точность извлеченных правил;

3) определить количество знаний в нейронной сети, которые не были извлечены;

4) определить противоречия между извлеченными правилами и обученной нейронной сетью.

Типы интеллектуального анализа данных, основанного на нейронных сетях

Существует множество типов анализа данных, основанных на нейронных сетях, но можно выделить два из них, наиболее популярных. Они основаны на самоорганизующихся нейронных сетях и на нечетких сетях.

- *Анализ данных (data mining), основанный на самоорганизующейся нейронной сети*

Самоорганизационный процесс – процесс обучения без учителя. При таком обучении [6] обучающее множество состоит из значений входных переменных, а в процессе обучения нет сравнения выходов нейронов с желаемыми значениями. Можно сказать, что такая сеть учится понимать структуру данных.

Идея сети Кохонена принадлежит финскому ученому Тойво Кохонену. Принцип работы этих сетей заключается во введении

в правило обучения нейрона информации о его расположении, то есть составляются карты размещения нейронов.

Самоорганизующиеся карты Кохонена используются для моделирования, прогнозирования, поиска закономерностей в больших массивах данных, выявления наборов независимых признаков и сжатия информации.

- *Анализ данных (data mining), основанный на нечеткой нейронной сети*

В основе нечетких нейронных сетей лежит идея использования существующей выборки данных для определения параметров функций принадлежности, выводы делаются на основе аппарата нечеткой логики, а для нахождения параметров функций принадлежности используются алгоритмы обучения нейронных сетей. Такие системы могут использовать заранее известную информацию, обучаться, приобретать новые знания, прогнозировать временные ряды, выполнять классификацию образов. Но одним из главных достоинств является наглядность работы такой сети для пользователя [2].

Каждый из рассмотренных типов нейросетей обладает своими преимуществами и недостатками применительно к интеллектуальному анализу данных (таблица).

Преимущества и недостатки популярных нейронных сетей в data mining

Тип нейронной сети	Область применения	Преимущества	Недостатки
Сеть Кохонена	Классификация, кластерный анализ, прогнозирование, сжатие данных	Устойчивость к зашумленным данным, неуправляемое обучение, быстрое обучение, возможность визуализации, возможность упрощения многомерной структуры	Эвристичность алгоритма обучения, предопределенность числа кластеров
Нечеткая нейронная сеть	Классификация, прогнозирование	Хорошая сходимость, быстрое обучение, интерпретируемость накопленных знаний, наглядность работы, легко определить размер сети, допустимость к зашумленным и неточным данным, способны аппроксимировать функции любой степени нелинейности, параллельные вычисления	Априорное определение компонентов

Из таблицы видно, что и сети Кохонена, и нечеткие нейронные сети обладают преимуществами и недостатками.

Основное отличие сетей Кохонена от других типов нейронных сетей состоит в наглядности и удобстве использования. Эти сети позволяют упростить многомерную структуру, их можно считать одним из методов проецирования многомерного пространства в пространство с более низкой размерностью. Другое принципиальное отличие сетей Кохонена от других моделей нейронных сетей – неуправляемое или неконтролируемое обучение, что позволяет задавать лишь значения входных переменных.

Важнейшим преимуществом нейронной сети является возможность построения одной сети для вычисления нескольких выходных значений по нескольким входным [5], а также способность к логическому описанию процессов и ручной корректировке функций принадлежности [7].

Однако нечеткие нейронные сети выгодно отличаются от других типов тем, что вобрали в себя все плюсы нечетких множеств. Таким образом, объединив нечеткие множества и нейронные сети, получили универсальные системы, компенсирующие недостатки нейронных сетей.

Заключение. Основным достоинством применения нейронных сетей является возможность решать различные неформализованные задачи. При этом можно очень просто моделировать различные ситуации, подавая на вход сети различные данные и оценивая выдаваемый сетью результат.

В ходе применения нейронных сетей отмечен существенный недостаток: сложность понимания процесса получения сетью результата. Первым шагом к устранению данной проблемы является разработка новой технологии, которая позволяет генерировать описание процесса решения задачи нейронной сетью. Используя таблицу экспериментальных данных, описываю-

щих предметную область, можно будет получить явный алгоритм решения поставленной задачи.

Из рассмотренных типов анализа данных, основанных на нейронных сетях, можно сказать, что нейронные сети, системы нечеткой логики являются обязательным инструментом интеллектуального поиска и извлечения знаний, т. к. обладают способностью выявления значимых признаков и скрытых закономерностей в анализируемых экономических показателях.

Список литературы

1. Барсегян А.А. Методы и модели анализа данных: OLAP И Data Mining / А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод. – СПб.: BHV, 2004. – 331 с.
2. Бирюков Е.В., Корнев М.С. Практическая реализация нечеткой нейронной сети при краткосрочном прогнозировании электрической нагрузки [Электронный ресурс] // Портал магистров ДонНТУ : сайт. – URL: http://www.masters.donntu.edu.ua/2006/kita/chuykov/library/library/article_5.htm (дата обращения 02.11.2010).
3. Дюк В.А. Data Mining – интеллектуальный анализ данных // Информационные технологии: сайт. – URL: <http://www.inftech.webservis.ru/it/database/datamining/ar2.html> (дата обращения 01.11.2010)
4. Манжула, В.Г. Методы «мягких» вычислений для аналитической обработки информации в условиях неопределенности / В.Г. Манжула, С.А. Морозов, С.В. Федосеев // Фундаментальные исследования. – 2009. – № 4. – С. 75–76.
5. Назаров А.В., Лоскутов А.И. Нейросетевые алгоритмы прогнозирования и оптимизации систем/ А.В. Назаров, А.И. Лоскутов – СПб.: Наука и Техника, 2003. – 384 с.
6. Чубукова И.А. Data Mining. – М.: Изд-во «Интернет-университет информационных технологий – ИНТУИТ.ру», 2006. – 384 с.
7. Ярушкина Н.Г. Основы теории нечетких и гибридных систем: учеб. пособие. – М.: Финансы и статистика, 2004. – 320 с.
8. Xianjun Ni Research of Data Mining Based on Neural Networks // World Academy of Science, Engineering and Technology. – 2008. – № 39. – P. 381–384.

Рецензенты:

Савельев М.В., д.т.н., профессор, профессор кафедры «Электронных вычислительных

машин» ГОУ ВПО «Южно-Российский государственный технический университет (НПИ)», Новочеркасск;

Галушкин Н.Е., д.т.н., профессор, зав. кафедрой «Прикладная информатика и математика» филиал федерального государственного автономного образовательного учреждения ВПО «Южный федеральный университет», Новошахтинск;

Кириянов Б.Ф., д.т.н., профессор, профессор кафедры прикладной математики и информатики ГОУ ВПО «Новгородский государственный университет имени Ярослава Мудрого», Великий Новгород;

Пен Р.З., д.т.н., профессор, профессор кафедры целлюлозно-бумажного производства ГОУ «Сибирский технологический университет», Красноярск.

KOHONEN NEURAL NETWORKS AND FUZZY NEURAL NETWORKS IN DATA MINING

Manzhula V.G., Fedyashov D.S.

*The South Russian state university of economy and service, Shahty,
e-mail: mail@sssu.ru*

In the introduction, we consider the relevance of using neural networks in data mining. The first section describes a method of neural networks in data mining. The second section presents the data analysis process, based on artificial neural networks. The third section looks at specific types of neural networks used in data mining, namely neural networks and fuzzy Kohonen neural network. Identified strengths and weaknesses of these networks. In conclusion, recommendations on the use of neural networks considered in the technologies of data mining.

Keywords: data mining, neural networks, Kohonen network, fuzzy neural networks