



Литература

1. Каширина, И.Л. Нейросетевые технологии: учебно-методическое пособие для вузов / И.Л. Каширина. – Воронеж: ИПЦВГУ, 2008. – 72 с.
2. Короткий, С. Нейронные сети Хопфилда и Хэмминга / С. Короткий. – 69 с.

U.S. Otajonov

DATA MINING METHODS

(Tashkent university of information technologies named after Muhammad al-Khwarizmi)

Data mining is information processing and identifying trends in it, which helps to make decisions. There are many different methods of data mining, modeling requests for processing and collecting information.

Intelligent analysis + data is one of the most relevant topics in the modern world. And this is not an exaggeration: business analysis + is only a small part of the scope of this powerful tool.

The principles of data mining have been known for a long time, but with the advent of big data, they have become even more widespread. When working with large data sets, relatively simple and straightforward statistics are no longer enough. Having about a million detailed records, you need to know not only the information about the location of several thousand records, but you also need to understand whether this category of records belongs to a specific group.

These requirements have created a complex data mining process. To solve problems, an analysis + data is required, in which a model is built to describe the information, and also a final report is created.

Big data combines three dimensions: volume, speed, and diversity.

Volume: when enterprises are overflowed with data of all types, the volume of which is growing and can easily reach terabytes and even petabytes.

Speed: sometimes 2 minutes is too long. In time-sensitive processes such as fraud detection, big data should be used as they enter the organization in order to maximize their value.

Diversity: big data can be of any structured or unstructured type, consisting of text data from sensors, sound, video recording, lists of visited web pages, protocol files and other data. When all these data of different types are analyzed, new ideas are born.

Intelligent analysis + data can be performed with relatively modest database systems and simple tools including creating your own or using ready-made software packages. Sophisticated data mining + is based on past experience and algorithms defined using existing software and packages, and various specialized tools are associated with various methods.



Working with big data sets, as well as their processing allows you to create complex generalizations of the results of data mining by groups and data comparisons. New systems and tools are now available, such as combined storage and processing systems. SQL databases strictly regulate the structure and strictly adhere to the schema, which simplifies queries to them and analysis + data with a known format and structure.

Consider some key data mining techniques.

Association (or relation) is the most famous method of data mining. This method consists in comparing two or more elements most often of the same type.

The classification method is used in describing several attributes to identify a particular class, as well as input to other methods. Decision trees are used to define the classification. Each class has certain properties that characterize its objects. A good example is the task of finding out the creditworthiness of a client. A bank employee thus operates with two classes known to him - “creditworthy” and “non-creditworthy”.

So, the task of classification is to determine to which class certain data belong; however, the set of classes to one of which can later be attributed to the object under study is known in advance.

Clustering allows the use of common attributes of various classifications in order to identify clusters. By examining one or more attributes, you can group the individual data elements together to get a structured conclusion. At a simple level, clustering uses one or more attributes as the basis for determining a cluster of similar results. Clustering is useful in defining various information because it is correlated with other examples, so you can see where the similarities and ranges agree with each other.

Clustering is a task at first glance very similar to the classification method. This method, also called cluster analysis, is a logical continuation or, if you wish, a generalized case of the classification problem when the set of classes is not known in advance.

Prediction methods use the values of some variables to predict unknown (target) values of variables. In combination with other data mining methods, forecasting involves analyzing + trends classification, comparison with the model and relationships.

The decision tree associated with most other methods is used within the framework of the selection criteria as well to support the selection of certain data within the overall structure. A decision tree begins with a simple question that has two answers (but perhaps more). Each answer leads to the next question helping to classify and identify data or make predictions.

Some of these processes can be automated. For example, building a predictive model to detect credit card fraud is reduced to determining the probabilities that can be used for the current transaction and then updating this model when adding new (confirmed) transactions. Then this information is recorded so that the next time the decision can be made faster.



Intelligent analysis + data is based on the construction of a suitable model and structure that can be used to process the identification and creation of the necessary information. Regardless of the form and structure of the data source, the information is structured and organized according to a format that allows you to perform intelligent analysis + data with the most efficient model.

Intelligent data analysis + is not only the execution of some complex data queries stored in the database. Regardless of whether you are using document-based SQL databases, simple, unstructured files need to work with the data to format or restructure them. It is required to determine the format of the information on which your method and analysis + will be based. Then, when the information is in the right format, you can apply various methods (individually or in combination) independent of the desired basic data structure or data set.

References

1. Parr-Rud O. Data Mining Cookbook Modeling Data for Marketing, Risk, and Customer Relationship Management, 2001
2. Gideon Dror, Marc Boule, Isabelle Guyon, Vincent Lemaire, David Vogel The 2009 Knowledge Discovery in Data Competition (KDD Cup 2009) Challenges in Machine Learning, Volume 3.

А.М. Ольшанский, И.А. Довгерд

О ПРИМЕНЕНИИ PID-РЕГУЛЯТОРА К УПРАВЛЕНИЮ ИНС С ПЕРЕМЕННОЙ ПРОВОДИМОСТЬЮ СИГНАЛА ПРИ ИСПОЛЬЗОВАНИИ СХЕМЫ ПОСТОБУЧЕНИЯ

(Самарский государственный университет путей сообщения, АО «НИИАС»)

Помимо существования широкого класса искусственных нейронных сетей с известной архитектурой, таких как сети Хопфилда, сети Кохонена, многослойные персептроны и др., существуют отдельные виды сетей, которые специализированы для решения отдельных задач, таких как, например, составления расписания процессов.

Такие сети отличаются тем, что, являясь системами нелинейными, не всегда могут быть описаны эволюционным уравнением для сети в целом или для каждого нейрона. С позиций теории автоматического регулирования такие сети не являются наблюдаемыми, а также являются частично управляемыми (по выходу) [1].

В этих условиях применить аналитическое решение по управлению сети невозможно, кроме того, значительные трудности представляет собой запись динамики состояния каждого веса. Для реализации управления ошибкой при наличии такой постановки задачи возможно применить идею реализации синтеза ПИД-управления [2] с последующей трансляцией сигнала управления на вход обучающих алгоритмов сети.