IOP Conf. Series: Materials Science and Engineering **392** (2018) 062202 doi:10.1088/1757-899X/392/6/062202

Research on Application of Machine Learning in Data Mining

Teng Xiuyi[1,2], Gong Yuxia[1,2]

[1]Economics and Management School, Tianjin University of Science and Technology, Tianjin China, 300222;

[2]Financial engineering and risk management research Center, Tianjin University of Science and Technology, Tianjin China, 300222.

Abstract. Data mining has been widely used in the business field, and machine learning can perform data analysis and pattern discovery, thus playing a key role in data mining application. This paper expounds the definition, model, development stage, classification and commercial application of machine learning, and emphasizes the role of machine learning in data mining. Understanding the various machine learning techniques helps to choose the right method for a specific application. Therefore, this paper summarizes and analyzes machine learning technology, and discusses their advantages and disadvantages in data mining.

1.Introduction

At the end of the last century, with the rapid development of computers and information technology, human society entered the era of information technology. And how to effectively obtain valuable information resources from vast amounts of data has become particularly important.

Data mining is a data mining method that explores and analyzes a large number of disorganized data to obtain potentially useful information and model it.

Most scholars believe that data mining was first proposed by Fayyad at the Knowledge Discovery Conference in 1995. He believed that data mining was a complex process that automatically or semi-automatically finds effective, meaningful, potentially useful, and easily understood data models from a large number of data. Data mining is a very complicated process and requires multi-step iteration. In the process of solving practical problems, scholars have gradually summed up the process of data mining: The first step is to select the data, usually selecting the appropriate historical data; then, the selected data is preprocessed to eliminate differences and inconsistencies between data. Finally, the data is analyzed, and the interpretable model is obtained and the generality is verified.

Data mining is a cross-cutting discipline that needs to combine knowledge from all walks of life. The main characteristic embodied in the combination of data mining and machine learning is the emphasis on the characteristics and distribution of data. Those feature is mainly reflected in the application of machine learning in big data.

2. Machine learning

Machine learning is a learning method that automates the acquisition of knowledge. Machine learning plays an important role in artificial intelligence research. An intelligent system without learning ability cannot be regarded as a real intelligent system, but the intelligent system in the past was generally lack of learning ability. For example, they cannot correct themselves in time when they encounter errors.It does not automatically acquire and discover the required knowledge. Its reasoning is restricted to deduction and lack of induction. Therefore, they can only prove existing facts and theorems, and can

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

not find new theorems, laws and rules. It does not improve its performance by accumulating experience.With the further development of artificial intelligence, these limitations are becoming more and more prominent. In this situation, machine learning has gradually become one of the core of artificial intelligence. Its application has spread to all over branches of artificial intelligence, such as expert systems, automatic reasoning, natural language understanding, pattern recognition, computer vision, intelligent robots and other fields.

2.1 The definition of machine learning

According to the viewpoint of H·Simon, an artificial intelligence master, learning is the enhancement or improvement of the system's ability in its repeated work. When the system performs the same or similar tasks at the next time, it will be better or more efficient. Machine learning is an important way for computers to acquire knowledge and an important indicator of artificial intelligence. It is a discipline that studies how to use computers to simulate or realize human learning activities. It is to study how to make machines obtain new knowledge and skills by identifying and using existing knowledge. It is generally believed that machine learning is a process of acquisition knowledge with a specific purpose. Its internal performance is a process of knowledge growth from unknown to known. Its external performance is the improvement of some performance and adaptability of the system, so that the system can finish the task that could not be completed or better completed.

2.2 The basic model of machine learning

Based on the learning definition from H. Simon, we can establish the basic model of Figure 1. The external environment is a collection of external information expressed in some form, which represents the source of external information. Learning is the process of processing external information into knowledge. First, external information is obtained from the environment; and then the information is processed into knowledge and the knowledge is put into the knowledge base. The general principles of execution are stored in the knowledge base. The execution link is the process of using knowledge in the knowledge base to complete a certain task, and feeds back some information that obtained in the process of completing the task to guide further study.

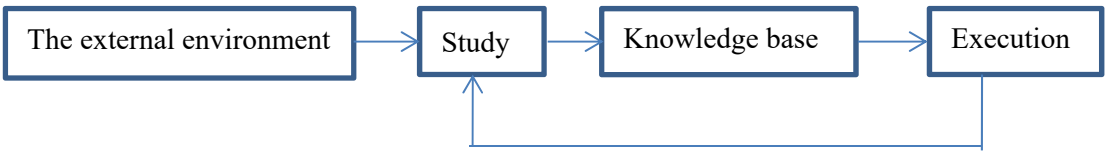

Figure 1. The basic model of machine learning

2.3 The development stage of machine learning

In the first stage, in the 1950s the main method was to construct the neural network and self-organizing learning system, and the learning performance was the feedback adjustment of the transmission signal of the threshold logic unit.In the second stage, in the early 1960s scholars began to study concept-oriented learning, which is symbolic learning. In concept acquisition, the learning system constructs the symbolic representation of concepts by analyzing a large number of positive and negative examples of relevant concepts. In the third stage, in the middle of 1970s, the first symposium on machine learning held in Carnegie Mellon University. It marks machine learning as an independent field of artificial intelligence.In the fourth stage, from the late 1980s to the present, machine learning research has entered into the fields of automation and pattern recognition. Various learning methods began to inherit and began to move from the laboratory to the field of application**.**

3. The classification of machine learning methods

3.1 Rule induction
Rule induction is to produce a decision tree or a set of decision rules from the training set to classify. The main advantage of rule induction is that it has strong ability to process large data sets and is suitable for classification and predictive tasks. The results are easy to interpret and technically easy to implement.

3.2 Neural networks
The neural network consists of processing nodes similar to human brain neurons. The input node is connected to the output node through a hidden node to form a multi-layer network structure. The neural network learns through repeated network training on historical sample data. The greatest advantage of neural network is that it can accurately predict complex problems.

3.3 Case reasoning
Each case consists of two parts: problem description and solution to the problem. After asking questions, the system will look for matching cases and solutions. Its advantage is that it can better deal with pollution data and missing data, which is very suitable for a large number of cases.

3.4 Genetic algorithms
Genetic algorithm is a combinatorial optimization method based on biological evolution process. The basic idea is the survival of the fittest and the best or better individual. The operation process includes reproduction, hybridization and mutation. The advantage of genetic algorithm is that it is easy to integrate with other systems.

3.5 Inductive logic programming
Inductive logic programming uses first-level attribute logic to define and describe concepts. First,it define positive and negative examples and then rank the new examples. This method has a strong conceptual description mechanism and can express complex relations well.

4.Classification of machine learning tasks in data mining

4.1 Classification
The training data set is used to obtain a classification model. Then, the classification model can automatically divide the data into multiple categories. The existing classification algorithms of machine learning include KNN classification algorithm, naive bayes classification algorithm, decision tree, artificial neural network and support vector Machine, etc.

4.2 Regression analysis
By analyzing the data and applying statistical methods, the relational expression between variables and variables is obtained. These inherent laws are used to estimate and predict future trends. Regression models are constructed by regression tree, artificial neural network, linear regression and logic regression.

4.3 Association rules
There are association rules among transactional data. By mining the relationship between transactional data, frequent itemsets can be obtained. Based on this, the probability of certain transactions occurring simultaneously is predicted. Apriori is a classical algorithm for mining association rules.

4.4 Clustering
By using the mining algorithm, multiple data without category labels are aggregated in a number of different clusters, so that the data objects in the cluster are similar to each other, and the data objects

between clusters are different from each other. K-means is a classical clustering algorithm.

5.Conclusion

The machine learning technology in data mining has been applied in many industries, including financial industry, retail industry, insurance industry, telecommunication industry and so on. For example, in the financial industry, financial analysts use data mining to build prediction models to identify the patterns that have caused market volatility in history, thereby improving the ability of predicting market volatility. In the retail industry, salespeople can build predictive models through data mining to understand who are most likely to respond to correspondence, thereby increasing sales. When enterprises apply data mining technologies, they should fully understand the advantages and disadvantages of various technologies and methods, and select appropriate technologies for specific environments and tasks.

References:

[1] R·Groth·HouDi.Data Mining - Building Competitive Advantages of Enterprises[M]. Xi'an:Xi'an Jiaotong University press,2001.
[2] ZhaoYijun,ShangMengjiao.The characteristics of data mining as a cross discipline[J].Times Finance,2017(03):263-264.
[3] LiYun.Application of machine learning algorithms in data mining[D].Beijing:Beijing University of Posts and Telecommunications,2014.
[4] HeQing.A Survey of Machine Learning Algorithms for Big Data[J].Pattern Recognition and Artificial Intelligence,2014(4):327 -336.
[5] WangXiao.Research on Trends of Machine Learning Algorithms in Big Data Environments[J].Natural sciences journal of harbin normal university,2013(4):48-50.
[6] AnZengbo,ZhangYan.The Application Study of Machine Learning[J].Journal of Changzhi University,2007,24(2):21-24.
[7] ZouYi.Overview of Datamining technology[J].Information & Communications,2016(12):164-165.
[8] YangJingfang.The application of machine learning algorithm in data mining[J].Electronic Technology & Software Engineering,2018(04):191.
[9] ZhangShaocheng.Research and Application of Machine Learning in Data Mining Based on Big Data[J].Journal of Liaoning university Natural Sciences Edition,2017,44(1):15-17.
[10] ChenXiao.Application of machine learning algorithm in data mining[J].Modern Electronics Technique,2015,38(20):11-14.