Research on Machine Learning Methods for Intelligent Decision Problems

Dong Xiao[1], liang Erwei[2], Zi Li[1,*]
[1]Daqing Normal University, Heilongjiang,China,163000
[2]PetroChina Kunlun Gas Co., Ltd.. Jiangxi Branch, China

Abstract—**In order to improve the accuracy, objectivity and scientificity of decision-making, people propose to use machine learning to automatically learn data and extract complex patterns from it, so as to make objective and fair decisions. Machine learning provides a new way to effectively solve decision support problems in many fields by virtue of its powerful ability to automatically learn data and extract complex patterns from them to make intelligent decisions. This paper mainly studies the construction of a supervisory classification model with strong generalization ability to solve complex decision-making problems such as disease diagnosis and financial risk prediction. Supervised classification learning methods such as SVM and FKNN are mainly studied. Diagnostic decision support methods based on combination RS and SVM, combined LFDA and SVM are proposed.**

Keywords- Reservoir; Dynamic analysis system; Web

I. Introduction

Machine learning is a subject about how to use computer programs to simulate human learning activities. In the real world, there are many problems such as handwritten numeral recognition, automatic driving, which cannot be solved by direct programming. Machine learning refers to the enhancement or improvement of the system's own ability in repeated work, so that the next time the system performs the same task or similar task, it will be better or more efficient than it is now. Generally speaking, if human behaviour changes under the influence of experience, we call this phenomenon learning. Experience is an important factor in learning, and the results of learning lead to changes in behaviour. Because computer programs are the main body of machine learning, their definitions are slightly different. Up to now, there is no unified, accepted and accurate definition of machine learning, but generally speaking, machine learning refers to computer programs using experience to improve the performance of the system itself.

II. Classification of machine learning

According to the learning strategies, machine learning methods can be divided into mechanical learning, instructional learning, analogical learning, and inductive learning and so on. At present, inductive learning (i.e. learning outcomes from training samples, which is a broad definition of learning from examples) is the most studied. If inductive learning is subdivided, there are also many classification systems. For example, in terms of learning initiative, it can be divided into Active Learning and Passive Learning; in terms of the starting time of training process, it can be divided into Eager Learning and Lazy Learning; in terms of learning experience, it can be divided into Supervised Learning, Unsupervised Learning and Reinforcement Learning.

Supervised learning refers to the training experience clearly telling the correct results. Just as people learn diagnostic technology through known cases, computers need to learn to have the ability to recognize all kinds of things and phenomena. If the output of supervised learning algorithm is continuous, it is called Regression, and if it is discrete, it is called Classification. Most of the machine learning tasks belongs to supervised learning. The main goal of unsupervised learning is to model the input set without class information. Common methods of unsupervised learning are Clustering and data dimension reduction Enhanced learning refers to the system's mapping from environment to behavior. The goal is to maximize the value of the reward signal function. Intensified signal is an evaluation of the quality of the action produced. Each action will have an impact on the environment, and then the environment will provide feedback in the form of rewards to guide the learning algorithm. To sum up, if we only evaluate the behavior of the program, the program will make the behavior that is more likely to be evaluated positively. Enhanced learning is widely used in intelligent control robots and analysis and prediction.

III. Supervised classification technique for machine learning

Classification is very common in real life, such as medical diagnosis, mechanical fault diagnosis, speech recognition, credit evaluation, enterprise bankruptcy prediction, spam filtering, network intrusion detection, gene selection, face recognition, web page classification, all of them belong to supervised classification learning. The supervised learning process consists of the following steps, as showed in figure 1.

978-1-7281-3681-3/19/$31.00 ©2019 IEEE
DOI 10.1109/CISCE.2019.00139

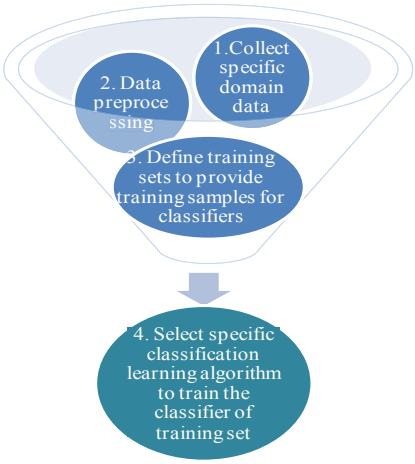

Fig1. The supervised learning process

Data preprocessing and classified learning are two important parts of the whole supervised learning process. Existing research shows that the performance of classification algorithm can be greatly improved by pre-processing data before selecting specific classification learning algorithm to learn data.

A. Support Vector Machine

SVM is a learning method developed on the basis of Statistical Learning Theory (SLT) to solve the learning rules of small samples. Compared with other traditional learning methods, it has many obvious advantages. For example, it is specifically for the case of limited samples (not only when the number of samples tends to be infinite), and its goal is to obtain the optimal solution under the existing information, which can effectively avoid the occurrence of over-learning phenomenon; its algorithm solution is ultimately transformed into the solution of quadratic optimization problem, therefore, in theory, the global optimal solution can be obtained. By introducing the nuclear technology, the dimension disaster problem is solved skillfully, so that the complexity of the algorithm is independent of the dimension of the sample.

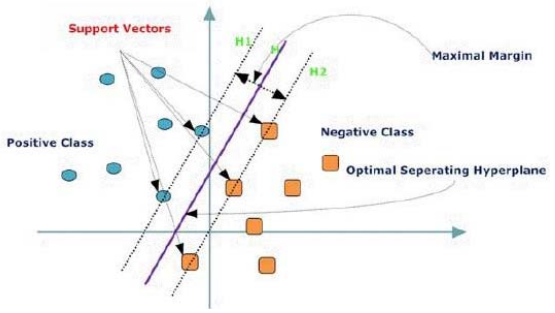

Fig2. Optimal classification surface in linear separable case

The SVM algorithm in the case of linear separability is described as shown in Figure 2, there are two types of samples in the figure. The circle represents positive class C_1 and the square represents negative class C_2. Because the training samples of two classes are linearly separable, there is an isolation zone between the two classes. Countless planes can be found between the isolation zones to separate the two types of data points, so how to choose the classification plane? The criterion of SVM is to find a classification surface H with maximum margin to separate the two types of samples as far as possible. In the graph, H1, H2 and H are parallel, H is its bisector, the sample on H1 is the nearest point from the first sample to H, and the sample on H2 is the nearest point from the second sample to H. These two sample points are very special and are on the edge of the isolation zone. These points are called Support Vectors, which determine the isolation zone. Between H1 and H2 is the margin of classification planes. The larger the margin is, the stronger the generalization ability, which means that the classification effect will be better.

B. Fuzzy k-Nearest Neighbor

For a test sample, the FKNN (Fuzzy k-Nearest Neighbor) classification algorithm first finds the K neighbors closest to the test sample from the training data set, and then calculates the class label with the largest membership degree by using the fuzzy decision-making method and regards it as the final category of the test sample. Since the proposed FKNN method, it has been widely used in many fields. For example, protein structure and function predict medical diagnosis.

IV. Medical model application with integrated feature dimensionality reduction technology and support vector machine

When there are irrelevant and redundant features in the data, the training complexity of SVM classifier will increase, the training speed will slow down, and the phenomenon of over-fitting will easily occur, which will eventually have a negative impact on the classification results. In order to reduce the complexity of SVM training and improve its classification speed and generalization ability, the combination dimensionality reduction technology and SVM classifier are applied to diagnose diseases. Using RS-based feature selection method, redundant features in data are reduced and core features are obtained. Furthermore, this method is combined with SVM to construct a high precision prediction model RS-SVM for medical diagnosis decision support. By introducing LFDA, it can effectively maintain the local features of data and find a set of optimal projection transformations, which can make the data samples belonging to different categories more separable in the constructed feature space. Compared with traditional dimension reduction methods, LFDA can better embed high-dimensional data into a low-dimensional space. Then, combining it with SVM, a high precision prediction model LFDA-SVM is proposed for disease diagnosis decision support.

V. Conclusions

Machine learning, with its powerful ability to automatically learn data and extract complex patterns from them to make intelligent decisions, is gaining more and more demands for biomedical and financial research, providing a new way to effectively solve decision support problems in these fields. This paper mainly studies the construction of supervised classification model with strong generalization ability, focusing on the supervised classification learning methods such as SVM and FKNN, and puts forward the

diagnostic decision support methods of combination RS and SVM, the diagnostic decision support methods of combination LFDA and SVM, and other complex decision-making problems.

References

[1] Hui-Ling Chen. A three-stage expert system based on support vector machines for thyroid disease diagnosis, Journal of Medical Systems, 2012, 36(3):1953-1963.

[2] Bo Yang, Jie Liu. A support vector machine classifier with rough set based feature selection for breast cancer diagnosis, Expert Systems with Applications, 2011, 38(7):9014-9022.

[3] Zhou Weiming. Multi-core computing and programming[M]. Huazhong University of Science and Technology Press, 2009.

[4] Duda R O, Shortliffe E H. Expert systems research [J]. Science, 2013, 220(4594): 261-268.

[5] Zadeh L A. Fuzzy sets [J]. Information and control, 2015, 8(3): 338-353.