ATLANTIS PRESS

Advances in Intelligent Systems Research, volume 147

International Conference on Network, Communication, Computer Engineering (NCCE 2018)

A Classification Diagnosis of Liver Medical Data Based on Various Artificial Neural Networks

Yong Qi [a)], Haozhe Liu [b)], Wentian Zhang [c)], Qiaosheng Zhu and Zhijian Zhao [d)]

Shaanxi University of Science and Technology, Shaanxi, 710021, China.

[a)] qiyong@sust.edu.cn
[b)] liuhaozhest@gmail.com
[c)]zwtmachinelearning@gmail.com
[d)] corresponding author: messhiro@gmail.com

Abstract. This paper presents a method for the identification and classification of medical data of hepatic pathological changes by using SVM FNN and KNN. The liver lesion classifier is a neural network model trained by experts' hand-divided samples and cross-validated to optimize the results. It can achieve better identification performance with medical data of hepatic pathological changes by training a variety of different neural network structures.

Key words: health care; liver complaint; machine learning; deep learning; regression.

INTRODUCTION

The liver is one of our vital organs, controlling the synthesis of proteins, detoxifying the food we consume. Once diseased, the liver's normal function will be affected and the body health will be threatened. The most common liver diseases include fatty liver, cirrhosis, liver cancer. In modern medical technology field, medical imaging and medical data sets are indispensable for the diagnosis and recognition of liver diseases. The two methods are widely used in clinical diagnosis and have important clinical application in various aspects, such as surgical planning, surgical navigation, the tracking of pathological changes and the evaluation of therapeutic effects.

In the field of medical data analysis, it is common to use artificial neural network to realize auxiliary diagnosis. Lee, sun, et al. put forward the application of machine learning method to establish breast cancer diagnosis model, for the purpose of improving the identification precision of the breast cancer diagnosis. They adopted three methods based on machine learning: BP (Back Propagation) network, LV (Learning Vector Learning Vector Quantization Network) and SVM (Support Vector Machine) as the training algorithm when setting the model. The simulation results showed that the diagnostic models seen in the three machine learning methods had high recognition rates. Cao, Ying, et al established a diagnostic prediction model based on BP neural network, logistic regression and stochastic forest algorithm. They compared the diagnostic value of the three models for prostate cancer, and eventually verified that all three models had higher diagnostic validity.

Due to the large amount of medical data sets and the wide range of difficulties, doctors' manual analysis of medical data sets has become increasingly time-consuming and inefficient. In order to improve this situation, the methods using computationally intensive (such as deep learning analytic medical data sets) prevail currently. This paper studied various neural network models such as SVN, FNN, KNN etc. to find the method of automatically extracting the characteristics of cervical medical data. Meanwhile, we determined whether the patient is diseased according to the classification of data based on the semantics. The proposed diagnosis combining traditional machine learning and deep learning network models ensured the accuracy and reliability of computer aided diagnosis.

EXPERIMENTS

Our team used the data set of Indian patients with liver disease for experiments and conducted experiments on the network model. The data set is a CSV file with a total of 583 records. It contains 416 liver patient records and 167 non-liver patient records. The attribute information of the data set contains the data of the 10 related test samples, which are age, gender, tot bilirubin, direct bilirubin, and tot proteins, albumin, ag ratio (albumin and globulin ratio), SGPT (aminotransferase), SGOT (serum aspartate aminotransferase) and Alkphos (alkaline phosphatase), The label indicates whether the test sample belongs to a liver disease patient.

We adopted machine learning and deep learning algorithms to achieve the effect of automatic classification and identification of liver diseases after network training through the relevant feature attributes.

(1) SVM (Support Vector Machine) is a common method of discrimination. In machine learning, it is a supervised learning model that is generally used for pattern recognition, classification, and regression analysis. The SVM maps the sample space to a high-dimensional or even infinite dimensional feature space through a nonlinear mapping, so that the problem of nonlinear separability in the original sample space is transformed into a linearly separable problem in the feature space. In short, it is dimension raising and linearization. Dimension raising is a method of mapping the samples to a high-dimensional space. As for regression and other issues, it is very likely that the sample set cannot be linearly processed in the low-dimensional sample space; while in the high-dimensional feature space, it can be mapped by a linear super-dimension plane to realize linear partitioning. The SVM method cleverly solves the complex computational problem caused by dimension-raising: it applies the expansion theorem of kernel function without knowing the explicit expression of nonlinear mapping; to some extent, it avoids the "dimensionality Disaster "problem.

We used RBF (Radial Based Function) as the kernel function of the SVM to perform experiments.

(2) KNN is one of the simplest classification methods in data mining classification technology. The so-called k-nearest neighbor indicates that the number of the nearest neighbors is k. That is to say each sample can be represented by its closest k neighbors.

The idea of this method: if a sample has the k most similar (i.e. the nearest neighbor in the feature space) samples in a feature space belonging to a certain category, the sample also belongs to this category. In the KNN algorithm, the selected neighbors are already correctly classified objects. The method determines the category to be sub-sampled according to the category of the nearest one or several samples in the class-decision. The KNN method, although also theoretically dependent on the limit theorem, is only relevant for a very small number of adjacent samples in class decision making. Because the KNN method mainly depends on the limited neighboring samples instead of the discriminant domain, it is more efficient than other methods for the cross-over or overlap of the sample sets to be divided.

In the third step, we trained the preprocessed datasets through deep learning and FNN (feedforward neural network).

FNN (Feedforward Neural Network) is the simplest neural network in which neurons are arranged in layers. Each neuron is connected to the neurons only in the previous layer. It operates by receiving the output of the previous level, and outputting to the next level. There is no feedback between layers, which can be represented by a directed acyclic graph. It is currently the most widely used and the one of the fastest growing artificial neural network.

We used three layers of neurons, the first two layers are activated using the relu function and the last is bisected using the sigmoid function.

ANALYSIS

The accuracy rate of each network after experimental training shows as Table 1.

TABLE 1. Accuracy of different networks

	SVM	KNN	FNN
Accuracy	0.7896	0.7886	0.9766

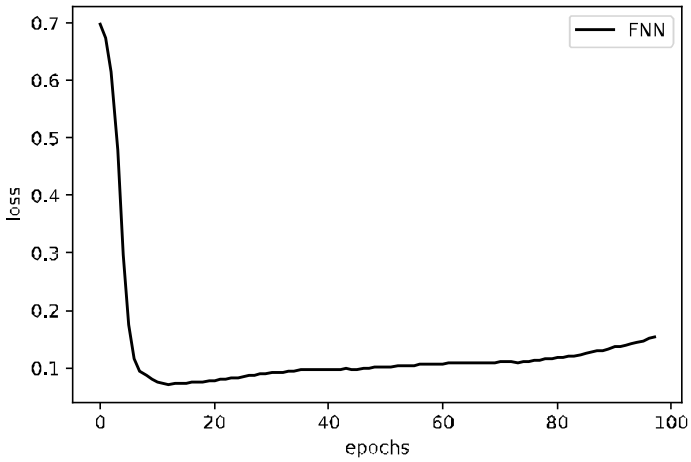

FIGURE 1. The loss of FNN

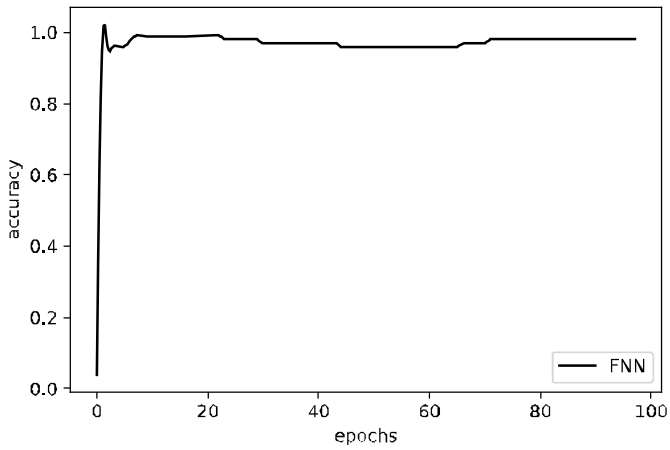

FIGURE 2. Accuracy

The accuracy of SVM is 0.7896, the accuracy of KNN is 0.7886, and the accuracy of FNN is 0.9766. By comparison, the accuracy of the feedforward neural network is higher than the other two machine learning algorithms.

The accuracy of FNN and the training results of loss function (cost function) are as Fig.1 and Fig.2.

The figure shows that the designed FNN network structure performed smoothly and steadily.

CONCLUSION

Multiple artificial neural network provides an effective method for medical diagnosis. The three kinds of artificial neural network models in this paper (SVM, FNN, KNN) can promote the application of data feature extraction and auxiliary diagnosis of liver disease. In a certain sense, it can simplify the medical diagnosis process and reduce the professional requirements of diagnosticians, providing a basic way to gradually improve the diagnostic intelligence level.

ACKNOWLEDGEMENTS

We are grateful to the following funds:
Student's Platform for Innovation and Entrepreneurship Training Program (1303)
Education Department of Shaanxi Province (15JK1086)
Natural Science Foundation of Shaanxi Province (2015GY009)
Shaanxi University of Science and Technology Dr. Foundation (BJ14-07)

REFERENCES

1. Experimental data: https://www.kaggle.com/loveall/cervical-cancer-risk-classification
2. Huang Shujing, Lou gemtek. Technology application of [J]. staining analysis and clinical diagnosis of hepatocellular carcinoma, 2016, 23 (4): 466-469.
3. Li Rong, Sun Yuan. The application of machine learning in the diagnosis of thoracic cancer [J]. science and technology and engineering, 2011, 11 (20): 4730-4733.
4. Cao Wenzhe, Ying Jun, Zhang Yahui, et al. Research on the diagnostic model of prostate cancer based on machine learning algorithm [J]. Chinese medical equipment, 2016, 31 (4): 30-35.
5. Vaidya J, Yu H, Jiang X. Privacy-preserving SVM classification[J]. Knowledge & Information Systems, 2008, 14(2):161-178.
6. Haobin Shi*, Xuesi Li, Wei Pan, Kao-Shing Hwang, A Novel Fuzzy Three-dimension Grid Navigation Method for Mobile Robots, International Journal of Advanced Robotic Systems, Volume 14, Issue 3, PP: 1-16, 2017(SCI, IF: 0.987, area 4)
7. Haobin Shi*, Zhiqiang Lin, Shuge Zhang, Xuesi Li, Kao-Shing Hwang, An adaptive Decision-making Method with Fuzzy Bayesian Reinforcement Learning for Robot Soccer, Information Sciences, Volume 436-437, pp:268-281, 2018(SCI, IF:4.832, area 1, TOP Journal)
8. Y, X., M. T and F. Q, DEEP LEARNING OF FEATURE REPRESENTATION WITH MULTIPLE INSTANCE LEARNING FOR MEDICAL IMAGE ANALYSIS, in Acoustics, Speech and Signal Processing (ICASSP). 2014, IEEE. p. 1626-1630.
9. Li Y, Dorai C. SVM-based audio classification for instructional video analysis[C]// IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004. Proceedings. IEEE, 2004: V-897-900 vol.5.