Identification of Potential Parkinson's Disease Patient through Voice Features Based on BP Neural Network and Support Vector Machine

Jinshu Gu
Aerospace Research Institute of Materials & Processing Technology, Beijing, China

Abstract: Parkinson's disease (PD) is a common neurological disorder that is one of the most complex human diseases. The early detection and treatment of Parkinson's disease is of great significance to the remission of the disease process. In this paper, BP neural network (BPNN) and Support Vector Machine (SVM), which are two machine learning methods, were used to construct a classification model for identifying potential Parkinson's disease patients based on 754 features of patient and normal people's voice features. The experimental results show that the classification accuracy of BPNN is 96.34%, while that of SVM is 77.84%. Therefore, BPNN would be an effective method for early identification of Parkinson's disease patients through voice features.

Keywords: Parkinson's disease, voice, classification, BP neural network, support vector machine

1. Introduction

Parkinson's disease (PD) is a common neurological disease, which is mostly seen in the elderly, with the average age of onset being about 60 years old, and the onset of Parkinson's disease in young people under 40 years old is relatively rare [1]. Most people with Parkinson's disease have sporadic cases, and less than 10% have a family history. In 1817, James Parkinson, an English doctor, first described the disease in detail. Its clinical manifestations mainly include static tremor, bradykinesia, myotonia and postural gait disorder. At the same time, patients may be accompanied by depression, constipation and sleep disorders and other non-motor symptoms. The most important pathological change of Parkinson's disease is the degeneration and death of dopamine (DA) neurons in the substantia nigra, which leads to the significant reduction of DA in the striatum and causes the disease [2]. The exact cause of this pathological change is still unknown, and genetic factors, environmental factors, aging, oxidative stress and other factors may participate in the denaturation and death of PD dopaminergic neurons. Parkinson's disease is a chronic progressive disease with high heterogeneity. The disease progresses at different rates in different patients. In the terminal stage of the disease, due to the poor response to drugs, the symptoms cannot be controlled, the patients can be rigid, unable to take care of themselves, or even stay in bed for a long time, and eventually die of pneumonia and other complications. Therefore, early detection of preclinical patients and effective preventive measures to prevent the degeneration and death of dopaminergic neurons can prevent the occurrence and progression of the disease.

In recent years, aiming at the diagnosis of Parkinson's disease, scholars have proposed many clinical decision support systems based on machine learning [3]. These systems fall into two main categories: on the one hand, to distinguish between people with Parkinson's disease and those without, that is, to achieve the diagnosis of Parkinson's disease. On the other hand, building the measure used to assess Parkinson's disease, that is, to achieve an assessment of Parkinson's disease. Considering the application of machine learning in prediction and its features of high accuracy and short time, the diagnosis based on machine learning is not only easy to measure, short time, low price, but also easy to implement [4].

Therefore, in this paper, we analyzed the sound characteristics of Parkinson's disease patients and normal people, constructed feature sets, and utilized BPNN and SVM to carry out classification prediction, so as to find potential patients and provide help for the prevention of Parkinson's disease.

2. Methodology

2.1 Back Propagation Neural Network

Artificial neural networks (ANNs) could establish complex non-linear relationships between the inputs and outputs through the transfer functions of the neurons. The main idea of ANNs is to implement the training process on a representative set of input-output “data pairs” by some optimization algorithms, even if the inputs are noisy or incomplete. As the most classic and commonly used ANN, BPNN method has been used by many researchers to solve some complicated, multivariate, and non-linear problems, due to its self-adaptive, nonlinear mapping, generalization abilities, and so on [5]. BPNN method was proposed by Rumelhart et al, which is a kind of multilayer feed forward network trained by error back-propagation algorithm. The topology of BPNN is shown in Fig. 1.

The structure of BPNN is composed of three different kinds of layers, which are an input layer, an output layer, and one or more hidden layers between the input and output layers. As shown in Fig. 1, the notation BPNN m-r-s-n could stand for a neural network with m input variables, n output variables, and two hidden layers with r and s neurons, respectively. The BPNN algorithm includes forward propagation of signal and back propagation of error. That is, the error output is calculated in the direction from input to output, while the adjustment weights and thresholds are calculated in the direction from output to input. When forward propagation, the input signal would spread through the hidden layers and generates the output signal through nonlinear transformation. If the actual output is not consistent with the expected output, the process of reverse propagation of error is transferred. By adjusting the connection strength of the input node and hidden layer node, and the connection strength and the threshold value of hidden layer node and an output node, the error is reduced along the direction of the gradient. After repeated learning and training, the network parameters corresponding to the minimum error are determined, and the training is stopped. At this point, the trained BPNN could process the non-linear transformation information with the minimum output error automatically for the input information of similar samples.

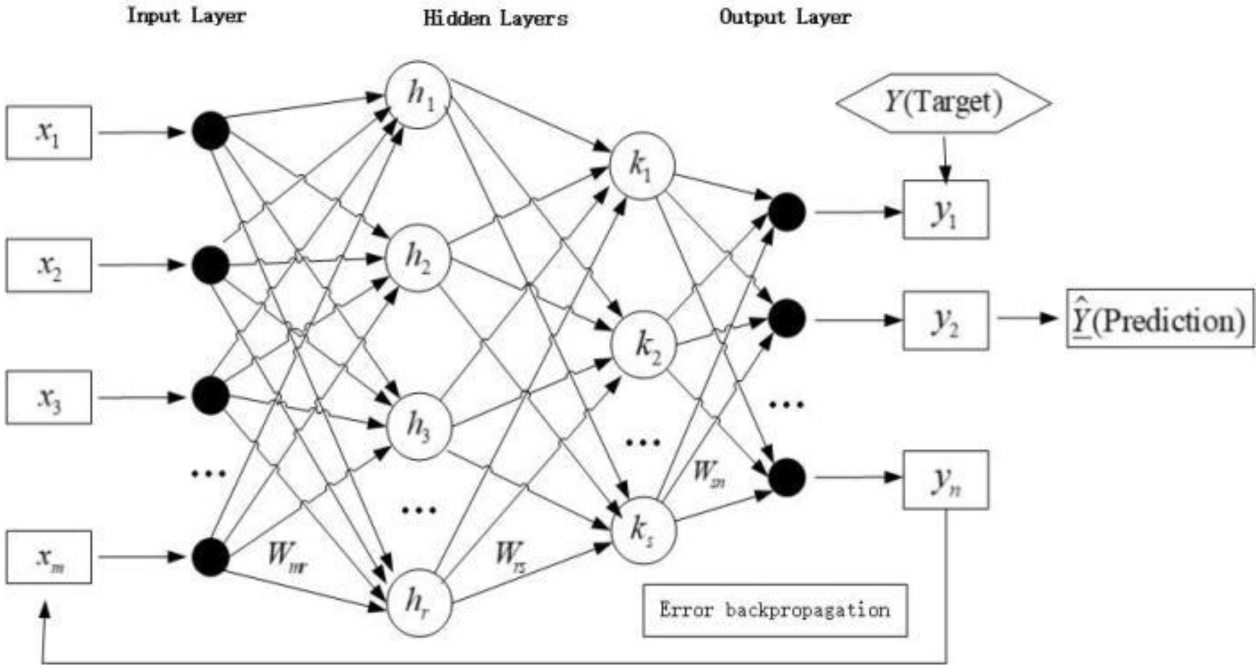

Figure 1. The Topology of Back Propagation Neural Network

2.2 Support Vector Machine

As a traditional classification method, SVM method has been widely used in several different fields, such as pattern classification, bioinformatics, and text categorization, due to its good generalization performance and strong theoretical foundations [6]. Facing the datasets with relatively small sample size and nonlinear characteristics, SVM shows its advantages than many other classification methods. In general, SVM adopts the principle of structural risk minimization (SRM), which could avoid the "dimension disaster" and has great generalization ability. The main objective of SVM is to estimate a relationship between input and output random variables under the assumption that the joint distribution of the variables is completely unknown.

The implementation of SVM model can be summarized by the following steps: (1) define the dataset for training and testing; (2) choose an appropriate kernel function (Linear, Gauss, Polynomial, and Sigmoid); (3) select a hyper-parameter optimization method (Grid Search, Evolutionary Strategy, Particle Swarm Optimization, and Simulated Annealing); (4) model training and testing.

The SVM model could be formalized as a problem of inferring a function $y = f(x)$ based on the training data $X = \{(x_i, d_i), i = 1,2, \dots, m\}$, where $x_i \in R^n$ is the i^{th} input vector for the i^{th} training data, $d_i \in R$ is the target value for the i^{th} training data and m is the number of training data. Furthermore, learning an SVM is equivalent to finding a regression function of the form:

$$f(x) = \sum_{i=1}^{m}(\alpha_i - \alpha_i^*)k(x_i, x) + b \quad (1)$$

Where $k(x_i, x)$ is a positive definite kernel function, $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_m)^T$, $\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_m^*)^T$ and b are the parameters of the model.

In our research, the Gauss kernel function which is a nonlinear kernel would be used. Its form is given by the following expression:

$$k(x_i, x) = \exp(-\frac{\|x_i - x\|^2}{\sigma^2}) \quad (2)$$

where $\sigma > 0$ is the width of the kernel.

In total, the essence of SVM is to find an n-1 hyperplane with the best tolerance in the n-dimensional space with x samples to separate the two kinds of samples and make the distance from any sample to the hyperplane greater than or equal to 1. The establishment of hyperplane is only affected by the vectors on the decision boundary. The vector on the decision boundary is called support vector, so the algorithm is called support vector machine. When the SVM encounters linearly indivisible samples, it will project the samples to a higher dimensional space to make the samples linearly separable.

3. Experiment Results

3.1 Data Description

In this paper, the Parkinson's Disease Classification Data Set from UCI Machine Learning Repository was used, which consists of a total of 756 sample data of Parkinson's patients and normal persons. The total number of features is 754, including several different voice features such as APQ3, APQ5, DDA, harmonic noise ratio, cyclic density entropy, etc.

For the purpose of model training, the dataset was randomly divided into a training set with 605 samples and a testing set with 151 samples. To make the evaluation more credible, the dataset division process was repeated ten times and finally computed the average prediction accuracy of different models. Besides, the min-max normalization process was implemented on both datasets, which could eliminate the impact of data magnitude through transferring all the data to [0, 1].

3.2 Hyper-parameter Optimization

Generally, the prediction accuracy of machine learning methods would be significantly influenced by hyper-parameter. Thus, for each model, the number and the range of variation of its hyper-parameter should be determined in advance.

For BPNN, Grid Search method was also implemented to conduct a parametric space search process, through which the optimal combination of different hyper-parameters could be obtained. That is, the number of hidden layers is 1, with the network structure of [40, 25, 1]. And the learning rate is 0.1. For SVM, Grid Search method was also implemented and the optimal combination of different hyper-parameters was finally obtained. That is, the kernel function is Gaussian, the penalty factor is 128, and the kernel function coefficient is 128.

In this paper, Python 3.6 was used to perform all the computation.

3.3 Classification Accuracy

After the training and testing process, the evaluation results of two models were obtained, respectively. The average prediction accuracy of different models is shown in Table 1.

Table 1. The Average Prediction Accuracy of Different Models

	T1	**T2**	**T**
BPNN	96%	96.67%	96.34%
SVM	80%	76.68%	77.84%

According to the results, the BPNN model is superior to the SVM model in this problem, with the accuracy improved by about 19%.

4. Conclusion

In this study, BPNN and SVM were mainly utilized to identify potential Parkinson's disease patients. Through the experiment results, the following major conclusions could be obtained. Firstly, BPNN could obtain relatively high prediction accuracy, which is a very effective method to make prediction. Secondly, voice features were proved to be useful features which could be utilized in machine learning methods.

In the future, we will expand the dataset to further verify the effectiveness of the method, and we will also try to use deep learning methods to make classification.

References

[1] W. Dauer and S. Przedborski. "Parkinson's Disease: Mechanisms and Models." *Neuron*, 2003: 889-909.

[2] C. R. Freed et al. "Transplantation of Embryonic Dopamine Neurons for Severe Parkinson's Disease." *New England Journal of Medicine*, 2001.

[3] C. O. Sakar et al. "A comparative analysis of speech signal processing algorithms for Parkinson’s disease classification and the use of the tunable Q-factor wavelet transform." *Applied Soft Computing*, 2018: 255-263.

[4] N. Abramson, D. J. Braverman, and G. S. Sebestyen. "Pattern recognition and machine learning." *IEEE Transactions on Information Theory*, 1963: 257-261.

[5] J. Yi et al. "BP neural network prediction-based variable-period sampling approach for networked control systems." *Applied Mathematics & Computation*, 2007: 976-988.

[6] J. A. K. Suykens and J. Vandewalle. "Least Squares Support Vector Machine Classifiers." *Neural Processing Letters*, 1999: 293-300.