

Classification of Parkinson's Disease by Comparing Multi Layer Perceptron Neural Network and Least Square Support Vector Machine

Choudhari Priyanka¹, Smt. T Jayakumari²
1M.Tech, Dept of Computer Science and Engineering,
BTL Institute of Technology, Bangalore, India

2Assistant Professor, Dept of Computer Science and Engineering,
BTL Institute of Technology, Bangalore, India

Abstract: Parkinson disease (PD) is the second most common neurodegenerative disorder. Here, the performance of Multi-Layer Perceptron Neural Network (MLPNN) and Least Square Support Vector Machines (LS SVM) regression methods are used to classify the effective detection of Parkinson's disease (PD). Classification technique is used to predict whether the particular patient is Parkinson's disease affected or not. In many cases, not all the tests contribute towards effective diagnosis of a disease. This study comprises to classify the presence of Parkinson's disease with reduced number of attributes. Originally, 23 attributes are included in the classification. After applying correlation filter, 12 attributes are taken away and 11 attributes are used. Results indicate that the MLPNN provides the best performance among the other and accuracy for MLPNN yields 60% when compared with LS SVM which yields 41.05%.

Keywords: Parkinson's disease, Multi Layer Perceptron Neural Network, Least Square Support Vector Machine

I INTRODUCTION

Data Mining is defined as extracting the information from the large amount of data. In other words, data mining is mining the knowledge from data. There is huge amount of data available in Information Industry. This data is of no use until it is converted into some useful information. So analyzing this huge amount of data and extracting useful information from it is necessary.

Classification is a data mining (machine learning) technique used to predict group membership for data instances. Classification technique is used to predict whether the particular patient is Parkinson's disease affected or not. Classification is the process of finding a model that describes the data classes. The purpose of the model is to predict the class of objects whose class label is unknown. This derived model is based on analysis of set of training data. Classification predicts the class of objects whose class label is unknown. Its objective is to find a derived model that describes and distinguishes data classes. The Derived Model is based on the analysis of set of

training data that is data object whose class label is well known.

II RELATED WORK

(Tsanas et al., 2010a) [5]. In this research paper linear regression least squares (LS), iteratively re-weighted least squares (IRLS), least absolute shrinkage and selection operator (LASSO), and a nonlinear regression method (CART) are applied to the prediction of Parkinson's disease progression. It is reported that CART method achieved the smallest prediction error, and tracked the linearly interpolated UPDRS more accurately. AIC and BIC methods are applied on the features for optimal subset selection, and reported low testing errors with the six features (FS-1) of dysphonia. In a recent paper (Tsanas et al., 2010b) [2]; another feature set (logtransform of FS-2) using the LASSO linear regression algorithm was introduced for telemonitoring of PD progression.

CORINNA CORTES and VLADIMIR VAPNIK [2]. In this research paper the support-vector network combines 3 ideas: the solution technique from optimal hyperplanes (that allows for an expansion of the solution vector on support vectors), the idea of convolution of dot-product (that extends the solution surfaces from linear to non-linear), and the notion of soft margins (to allow for errors on the training set) The algorithm has been tested and compared to the performance of other classical algorithms. Despite the simplicity of the design in its decision surface the new algorithm exhibits a very fine performance in the comparison study. Other characteristics like capacity control and ease of changing the implemented decision surface render the support-vector network an extremely powerful and universal learning machine.

M.A. Razi, K. Athappilly [14]. In this research a three-way comparison of prediction accuracy is performed involving nonlinear regression, NNs and CART models. NNs and CART models, produced better prediction accuracy than non-linear regression model. Brown et al. (1993) shows

that NNs do better than CART models on multimodal classification problems where data sets are large with few attributes. However, the authors also point out that CART outperforms NNs models where data sets are smaller with large numbers of irrelevant attributes. NNs and CART models are shown to outperform linear discriminate analysis on problems where the data is non-linear (Curram& Mingers, 1994).

III PARKINON’S DISEASE

Parkinson’s disease is a chronic neurodegenerative disorder of unknown etiology, which usually affects people over the age of fifty (Tsanas, Little, McSharry, & Ramig, 2010b) [2].It is reported that it affects over 1 million people in North America, apart from thousands of undetected cases (Lang & Lozano, 1998) [3]. As the worldwide population is growing older, number of PD affected patients are even expected to increase. Although medication is available to reduce the symptoms, there is no complete treatment for PD (Das, 2010) [4]. Therefore, early diagnosis is critical and important to help patients to improve and maintain their quality of life (Tsanas, Little, McSharry, & Ramig, 2010a)[5]. However, PD may be difficult to diagnose accurately, especially at the early stages of the illness, due to symptom overlap with other disease (Ene, 2008) [6].

PD symptoms include (Sakar & Kursun,2009 [7]; Skodda & Sclegel, 2008 [8]; Tsanaset al.. 2010a)[9].

- tremor of the limbs
- muscle rigidity
- slowness of movement
- difficulty with walking
- balance and coordination
- difficulty in eating and swallowing
- vocal impairment
- cognition and mood disturbances

The data is collected at the patient’s home, transmitted over the internet, and processed appropriately in the clinic. The data is collected using the Intel At-Home Testing Device (AHTD), which is a telemonitoring system designed to facilitate remotely located, Internet-enabled measurement of a variety of PD-related motor impairment symptoms. Each patient specified a day and time of the week during which they have to complete the protocol, prompted with an automatic alarm reminder on the device. The collected data is encrypted and transmitted to a dedicated server automatically when the USB stick is inserted in a computer with internet connection. Further details of the AHTD apparatus and trial protocol can be found in the work of Goetz et al [10].

IV DATASET

Voice measurement has shown a great progress in the advancement of Parkinson Disease detection. About 90%

of people with Parkinson’s disease present some kind of vocal deterioration. And hence, in this paper dataset which mainly focus on the speech signals is chosen. This dataset is taken from UCI machine learning database [1]. The features of dataset are given in Table I.

TABLE I
TABLE DESCRIBING THE ATTRIBUTES THAT ARE REMOVED AFTER APPLYING THE CORRELATION FILTER

No.	Attribute Name	Rem.	Description
1	MDVP-Fo(Hz)	Yes	Average vocal fundamental frequency
2	MDVP-F1i(Hz)	Yes	Maximum vocal fundamental frequency
3	MDVP-F1o(Hz)	Yes	Minimum vocal fundamental frequency
4	MDVP-Jitter(%)	Yes	Variation in fundamental frequency
5	MDVP-Jitter(Abs)	No	Variation in fundamental frequency
6	MDVP-RAP	Yes	Variation in fundamental frequency
7	MDVP-PPQ	Yes	Variation in fundamental frequency
8	Jitter:DDP	No	Variation in fundamental frequency
9	MDVP-Shimmer	Yes	Measures of variation in amplitude
10	MDVP-Shimmer(dB)	Yes	Measures of variation in amplitude
11	Shimmer:APQ3	Yes	Measures of variation in amplitude
12	Shimmer:APQ5	Yes	Measures of variation in amplitude
13	MDVP-APQ	No	Measures of variation in amplitude
14	Shimmer:DDA	No	Measures of variation in amplitude
15	NHR	No	Ratio of noise to tonal components
16	HNR	No	Ratio of noise to tonal components
17	status	No	The status of the patient (1)-Parkinson’s Disease, (0)-Healthy
18	RPDE	No	Dynamic complex measurement
19	DFA	No	Signal fractal scaling exponent
20	spread1	Yes	Non-linear measure of fundamental frequency
21	spread2	Yes	Non-linear measure of fundamental frequency
22	D2	No	Dynamic complex measurement
23	PPE	No	Non-linear measure of fundamental frequency

The dataset is composed of a range of biomedical voice measurements, there are 195 instances comprising 48 normal and 147 PD cases in the dataset. The purpose of the data is to discriminate PD unaffected patients from those with PD affected. Thus, the dataset is divided into two classes according to its "status" column which is set to 0 for PD unaffected patients and 1 for those with PD affected patients. It is a two-decision classification problem.

A correlation filter is applied and out of 23 attributes 12 are removed. The correlation coefficient, which are less than 0.95 are not considered for classification accuracy, thus that attribute is removed. Likewise, a total of 11 attributes are kept after the correlation filter is has been applied. Table II indicates which features are kept. First 10 attributes are used as inputs to the classifiers.

TABLE II FEATURES USED IN THIS PAPER

Index	Attribute Name	Description
1	MDVP:Jitter(Abs)	Variation in fundamental frequency
2	Jitter:DDP	Variation in fundamental frequency
3	MDVP:APQ	Measures of variation in amplitude
4	Shimmer:DDA	Measures of variation in amplitude
5	NHR	Ratio of noise to tonal components
6	HNR	Ratio of noise to tonal components
7	RPDE	Dynamic complex measurement
8	DFA	Signal fractal scaling exponent
9	D2	Dynamic complex measurement
10	PPE	Non-linear measure of fundamental frequency
11	status	The status of the patient (1)-Parkinson’s Disease, (0)-Healthy

V ALGORITHMS

A. MLPNN (Multi Layer Perceptron Neural Networks)

In this study the MLPNN regression method is used. For this the number of input nodes is determined by the number of attributes decided i.e. 10 neurons; the number of hidden nodes is determined through trial and error; and [tansig, purelin, trainlm] transfer functions are used for input, hidden, output layers. The number of output nodes is represented as a class label that is 0 for PD unaffected or 1 for PD affected. Each neuron at one layer receives a weighted sum from all neurons in the previous layer and provides an input to all neurons for the later layer. For MLPNN regression method, Levenberg Marquardt (LM) backpropagation learning algorithm has been used in the feed forward single hidden layer neural network.

The MLPNN is a common neural network for both regression and process modeling and has often been used for diagnosis and tracking of disease (Kumar, 2005 [11]; Ubeyli, 2009) [12]. It consists of an input layer, hidden layer and output layer. Each neuron of a layer is connected to all the neurons of the following layer. The input vectors are fed to the input layer and get multiplied by interconnection weights, known as weight factor, as they are passed from the input layer to the hidden layer. Hidden layer then calculates the transfer function and sends its signal to the output layer. The output layer calculates the transfer function to produce the neural network output. It is also necessary to assign the transfer functions for each neuron layer. Backpropagation algorithm employed is a supervised learning method for MLPNN introduced by (Rumelhart, Hinton, and Williams (1986)) [13], for which Levenberg–Marquardt training function to update the weight and bias values are used. After training the network for a set of data, the network performance is evaluated with a new set of test data. The MLPNN outputs are then compare with the outputs obtained from the other regression methods. The detailed information about MLPNN architecture may be found in the literature (Kumar, 2005[11]; Razi & Athappilly, 2005[14]; Rumelhart et al., 1986) [13].

B. LS-SVM (Least Square-Support Vector Machine)

Support vector machine uses training samples to build a model that will classify information, usually visualized as a scatter plot with a wide space between categories. When new information is fed into the machine, it is plotted on the graph. The data are then classified based on which category the information falls closest to on the graph. This method works only when there are two options to choose from.

Least squares support vector machines (LS-SVM) are least squares versions of support vector machines (SVM), which are a set of related supervised learning methods that analyze data and recognize patterns, and which are used for classification and regression analysis. In this version one

finds the solution by solving a set of linear equations instead of a convex quadratic programming problem for classical SVMs. Least squares SVM classifiers, were proposed by (Suykens and Vandewalle)[15]. LS-SVMs are a class of kernel-based learning methods. LS-SVM is an alternate formulation of SVM introduced by (Suykens and Vandewalle (1999)) [15]. This model uses a set of linear equations instead of quadratic programming problem typical of classical SVM. LS-SVM method considers the regression problem as the following optimization problem:

$$\min_{w,b,e} Q(w, e) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{k=1}^m e_k^2,$$

Subject to constraints

$$y_k = w^T \varphi(x_k) + b + e_k, \quad k = 1, \dots, m,$$

where e_k is the prediction error term for data point k , γ is a regularization constant and $\varphi(x_k)$ denotes an infinite dimensional feature map. By using the Lagrangian method, the LS-SVM model can be expressed as:

$$y(x) = \sum_{k=1}^m \alpha_k K(x_k, x) + b,$$

where α_k ($k = 1, \dots, m$) is the introduced Lagrange multipliers and b is the bias term, Among the Kernel function known as linear kernel, multilayer perceptron kernel, and polynomial kernel, here the linear kernel is used and is given by,

$$k(x, y) = x^T y + c$$

VI RESULTS AND DISCUSSION

As proposed, the Multi Layer Perceptron Neural Network and Least Square Support Vector Machine are implemented in MATLAB R2013a. To evaluate this classification UCI Machine Learning Repository Dataset [1] is used. There are 195 instances comprising 48 are PD unaffected and 147 PD affected cases in the dataset. To classify the data between PD affected and unaffected people. Thus, the dataset is divided into two classes according to its "status" column which is set to 0 for PD unaffected cases and 1 for those with PD affected.

A. Performance Measures

To find the performance metrics such as sensitivity, specificity and accuracy, a distinguished confusion matrix is obtained based on the classification results from these regression methods. Confusion matrix is a matrix representation of the classification results as shown in Table III

TABLE III

	PD Affected	PD UnAffected
Actual Affected	TP	FN
Actual UnAffected	FP	TN

Accuracy is the percentage of predictions that are correct. The precision is the measure of accuracy provided that a specific class has been predicted.

Sensitivity is the percentage of positive labeled instances that were predicted as positive.

These performance criterions for the classifiers in the disease detection are evaluated as follows from the confusion matrix.

$$\text{Accuracy} = (TP+TN) / (TP+FP+TN+FN)$$

$$\text{Sensitivity} = TP / (TP+FN)$$

$$\text{Specificity} = TN / (FP+TN)$$

A.Results

For MLPNN regression method, Levenberg Marquardt (LM) backpropagation learning algorithm has been used in the feed forward single hidden layer neural network. 10 neurons are used for the input layer and tansig, purelin, trainlm transfer functions are used for input,hidden,output layer. 2 neurons for output layer is represented as a class label that is 0 for PD unaffected or 1 for PD affected. The outcome of MLPNN algorithm are tabulated in TABLE IV. MLPNN is developed using MATLAB Neural Network Toolbox generated the results as shown in TABLE V. There are 195 instances, out of which 48 patients are normal and 147 patients are PD Effected. MLPNN shows 60% accuracy.

TABLE IV

	PD Affected	PD UnAffected
Actual Affected	12	0
Actual UnAffected	8	0

TABLE V

Accuracy	60%
Sensitivity	100%
Specificity	0%

For LS-SVM, the Kernel Function, Linear is used. Linear kernel. In this case, svmtrain finds optimal separating plane in the space. The outcome of LS SVM algorithm are tabulated in TABLE V. The implementation of LS SVM in MATLAB generated the results as shown in TABLE VI. Accuracy for LS SVM shows 41.05%.

TABLE VI

Accuracy	41.025%
Sensitivity	87.500%
Specificity	25.850%

TABLE VII

	PD Affected	PD UnAffected
Actual Affected	42	6
Actual UnAffected	109	38

CONCLUSION

Early detection of any kind of disease is an essential factor. This helps in treating the patient well ahead. In this paper the performance of LS-SVM and MLPNN regression methods are used to classify the effective detection of Parkinson’s disease (PD). The results indicates that MLPNN regression method yields 60% accuracy when compared to LS SVM regression method which yields 41.05%. Thus MLPNN gives best result comparison with LS SVM regression method.

REFERENCES

- [1] UCI Machine Learning Repository- Center for Machine Learning and Intelligent System, <http://archive.ics.uci.edu>.
- [2] Tsanas, Little, McSharry, & Ramig, 2010b). Enhanced classical dysphonia measure and sparse regression for telemonitoring of Parkinson’s disease progression. International conference on Acoustics , speech and signal processing, 594-597.
- [3] Lang A.E., & Lozano, A.M. (1998) Parkinson’s Disease- First of two parts. New England journal Medicine, 339, 1044-1053.
- [4] Das, R. (2010). A comparison of multiple classification methods for diagnosis of Parkinson disease. Expert Systems with Applications, 37(2), 1568–1572.
- [5] (Tsanas, Little, McSharry, & Ramig, 2010a). Accurte telemonitoring of parkinson’s disease progression by non invasive speech tests. IEE Transactions on Biomedical Engineering, 57, 884-893.
- [6] Ene, M. (2008). Neural network-based approach to discriminate healthy people from those with Parkinson’s disease. Annals of the University of Craiova, Mathematics and Theoretical Computer Science, 35, 112–116.
- [7] Sakar, C, O., & Kursun, O, (2009). Telemonitoring of parkinson’s disease using measurements of dysphonia. Journal of medical system., 34(4), 591-599.
- [8] Skodda, S., & Schlegel, U. (2008). Speech rate and rhythm in parkinson’d disease. Movemen Disorders, 23,985-992.
- [9] Tsanaset al.. 2010a)
- [10] Goetz, C.G. et al. Testing objective measures of motor impairment in early Parkinson’s disease: Feasibility study of an at-home testing device.
- [11] Kumar, U. A. (2005) Comparison of neural network and regression analysis: A new insight. Expert systems with applications, 29, 424-430.
- [12] Ubeyli, E. D. (2009) Combined neural network for diagnosis of erythematosquamous disease. Expert systems with Applications(36), 5107-5112.
- [13] Rumelhart, D. E., Hinton, G. E., and Williams R. J. (1986). Learning international representation by error propagation . parallel and

distributed processing: Explorations in the microstructure of cognition: Foundations (vol 1). Cambridge, MA: MIT Press.

- [14] Razi, M. A., & Athappilly, K (2005). A comparative predictive analysis of neural networks(NNs), non linear regression and classification and regression tree (CART) models. Expert systems with Applications, 29, 65-74.
- [15] Suykens, J. A. K., and Vandewalle (1999). Least squares support vector machines classifiers. Neural processing Letters, 9(3), 293-300.
- [16] Cortes, C., & Vapnik, V. (1995). Support vector networks. Machine Learning, 20(3), 273-297.