VOL.5 NO.2 September 2016 - ISSN 2233 – 1859

UOIuBIH
ORSinBIH
Operations Research Society in Bosnia and Herzegovina

Southeast Europe Journal of Soft Computing

Available online: http://scjournal.ius.edu.ba

IUS Soft Computing Research Group

Classification of Parkinson's disease – A comparison between Support Vector Machines and neural networks

Enes Akca

International University of Sarajevo,
Faculty of Engineering and Natural Sciences,
HrasnickaCesta 15, Ilidža 71210 Sarajevo, Bosnia and Herzegovina

eakca@ius.edu.ba

Article Info

Article history:
Article received on 1 May 2016
Received in revised form 7 Aug.2016

Keywords:
Parkinson's disease; Radial basis function; Neural Networks; Support Vector Machine

Abstract

Parkinson's disease (PD) is a chronic and progressive movement disorder, meaning that symptoms continue and worsen over time. The diagnosis of Parkinson is challenging because currently none of the clinical tests have been proven to help in diagnosis. In this paper, the main purpose was to classify the PD people (sick) and non-PD people (healthy). Recently the machine learning methods based diagnosis of medical diseases has taken a great deal of attention. The Support Vector Machine (SVM) and the Neural Network (NN) learning methods are used as base classifiers. The support vector machine is a novel type of learning machine, based on statistical learning theory, which contains radial basis function (RBF) as special cases. 100% / 80% accuracies are reported.

1. INTRODUCTION

Parkinson's disease (PD) is a slow, progressive degenerative disorder of the central nervous system (Rango et al., 2006). It is characterized by muscle rigidity, tremor, slowing of physical movement (Bradykinesia) and, in extreme cases, loss of physical movement (Jankovic, 2003).

The disease can be difficult to diagnose accurately, particularly in the early stages of the disease when symptoms resemble other medical conditions, and misdiagnosis occurs occasionally. There are currently no blood or laboratory tests that have been proven to help in diagnosing PD, and the prognosis depends on the patient's age and symptoms. The diagnosis is based on the medical history and neurological examination conducted by interviewing and observing the patient. Brain scans or laboratory tests may be used to help doctors exclude other medical conditions that produce symptoms similar to those of Parkinson's disease.

Scientific research on vocal recordings of patients that suffer from Parkinson's disease are not abundant. The data set used in this study was collected by Little et al. (2009) who used support vector machine techniques to distinguish between the people who have normal vocal signals and who suffer from Parkinson's disease. They achieve a classification accuracy of 91.4% but they do not report single class true positive rates. This is noteworthy because of the highly imbalanced sick to healthy ratio (3:1) data class distribution of the Parkinson's disease data set (Freddie and Koker, 2011).

In previous studies in which the possibility of developing an ''intelligent'' stimulator (Cassidy et al., 2002; Gasson et al., 2005) by analysing the STN LFP signals was first documented, different types of artificial neural network such as Multiple Layer Perceptrons (Pan et al., 2007) and Radial Basis Function Networks (Wu, Warwick, Ma, Burgess, et al., 2010; Wu, Warwick, Ma, Gasson, et al., 2010) were used to ''predict'' PD tremor based on LFP signals. Initial results showed that both RBN and MLP could be used as classifiers to obtain reasonable results when fed with such tailored LFP signal features as the power spectrum density up to 45 Hz. In such cases both networks were able to separate the features into pre-defined categories with a respectable success rate that clearly could be improved upon.

Support Vector Machine (SVM) is a supervised classification method derived from statistical learning theory that often yields good classification results from complex and noisy data (Vapnik, 1995). It separates the classes with a decision surface that maximizes the margin between the classes. The surface is often called the optimal hyperplane, and the data points closest to the hyperplane are called support vectors. The support vectors are the critical elements of the training set (Brown et al., 2000).

Neural networks are the tools that should be recalled for any classification job. They are developed enormously since the first attempts made modeling the perceptron architecture six decades ago (Lee, 1997).

This paper deals with the application of machine learning methods (Baldi and Brunak, 1998; Durbin et al., 1998), such as Neural Networks (NNs) and Support Vector Machines (SVMs) to a medical dataset concerning PD with the aim of automatically classify patients in PD or non-PD depending on their medical attributes. In order to test the performance and efficiency of the proposed method, the classification accuracy were used. The paper is organized as follows. Section 1 defines the related works carried out in the Parkinson Disease area. Section 2 deals with the Parkinson Dataset that is used in this research work. Section 3 gives an overview of the neural networks and Support Vector Machines, and explain architecture used for classification of the patient as PD or non-PD. Section 4 is dealt with the experimental results of the Algorithms. Lastly, Section 5 concludes research paper.

2. DATA SET OF PARKINSON'S DISEASE

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 (UCI). The features of dataset are given in Table 1.

Table 1: Attribution information

No	**Attribute Name**	**Rem.**	**Description**
1	MDVP:Fo(Hz)	Yes	Average vocal fundamental frequency
2	MDVP:Fhi(Hz)	Yes	Maximum vocal fund. freq.
3	MDVP:Flo(Hz)	Yes	Minimum vocal fun. freq.
4	MDVP:Jitter(%)	Yes	Variation in fund. frequency
5	MDVP:Jitter(Abs)	No	Variation in fund. frequency
6	MDVP:RAP	Yes	Variation in fund. frequency
7	MDVP:PPQ	Yes	Variation in fund. frequency
8	Jitter:DDP	No	Variation in fund. frequency
9	MDVP:Shimmer	Yes	Measures of var. in amp.
10	MDVP:Shimmer(dB)	Yes	Measures of var. in amp.
11	Shimmer:APQ3	Yes	Measures of var. in amp.
12	Shimmer:APQ5	Yes	Measures of var. in amp.
13	MDVP:APQ	No	Measures of var. in amp.
14	Shimmer:DDA	No	Measures of var. in amp.
15	NHR	No	Ratio of noise to tonal comp.
16	HNR	No	Ratio of noise to tonal comp.
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 was created by Max Little (2007) of the University of Oxford, in collaboration with the National Centre for Voice and Speech, Denver, Colorado. This organization recorded the speech signals. This dataset is composed of a range of biomedical voice measurements from 31 people, 23 with Parkinson's disease (PD). Each column in the table is a particular voice measure, and each row corresponds one of 195 voice recording from these individuals ("name" column). The main aim of the data is to discriminate healthy people from those with PD, according to "status" column which is set to 0 for healthy and 1 for PD.

3. OVERVIEW OF MACHINE LEARNING

3.1. Support vector machines classification

Support Vector (SV) machine is a novel type of learning machine, based on statistical learning theory, which contains polynomial classifiers, neural networks, and radial basis function (RBF) networks as special cases. In the RBF case, the SV algorithm automatically determines centers, weights and threshold such as to minimize an upper bound on the expected test error.

The support vector machine is an elegant and highly principled learning method for the design of a feedforward with a single hidden layer of nonlinear units.

In machine learning, the radial basis function or RBF kernel, is a popular kernel function used in various kernelized learning algorithms. In particular, it is commonly used in support vector machine classification (Yin-Wen et al. 2010).

The RBF kernel on two samples x and xı is defined (1) and represented as feature vectors in some input space (Vert et al. 2004).

$$K(x, x') = \exp\left(-\frac{\|x-x'\|^2}{2\sigma^2}\right) \quad (1)$$

Since the value of the RBF kernel distance and ranges between zero (in the limit) and one (when x=x′), it has a ready interpretation as a similarity measure. The feature space of the kernel has an infinite number of dimensions; for σ=1, its expansion (2) is (Shashua, 2009):

$$exp\left(-\frac{1}{2}\|x-x'\|^2\right) = \sum_{j=0}^{\infty}\frac{(x^{\mathrm{T}}x')^j}{j!}exp\left(-\frac{1}{2}\|x\|^2\right)exp\left(-\frac{1}{2}\|x'\|^2\right) \quad (2)$$

3.2. Artificial Neural Networks

Nervous systems existing in biological organism for years have been the subject of studies for mathematicians who tried to develop some models describing such systems and all their complexities. Artificial Neural Networks emerged as generalizations of these concepts with mathematical model of artificial neuron due to McCulloch and Pitts (1943) described definition of unsupervised learning rule by Hebb (1949), and the first ever implementation of Rosenblatt's perceptron (1958). The efficiency and applicability of artificial neural networks to computational tasks have been questioned many times, especially at the very beginning of their history the book "Perceptrons" by Minsky and Papert (1969) caused dissipation of initial interest and enthusiasm in applications of neural networks.

It was not until 1970s and 80s, when the back propagation algorithm for supervised learning was documented that artificial neural networks regained their status and proved beyond doubt to be sufficiently good approach to many problems. Artificial Neural Network can be looked upon as a parallel computing system comprised of some number of rather simple processing units (neurons) and their interconnections. They follow inherent organizational principles such as the ability to learn and adapt, generalization, distributed knowledge representation, and fault tolerance. Neural network specification comprises definitions of the set of neurons (not only their number but also their organization), activation states for all neurons expressed by their activation functions and offsets specifying when they fire, connections between neurons which by their weights determine the effect the output signal of a neuron has on other neurons it is connected with, and a method for gathering information by the network that is its learning (or training) rule (Can, 2013).

Activation or transfer function of a neuron is a rule that defines how it reacts to data received through its inputs that all have certain weights.

Among the most frequently used activation functions are linear or semi-linear function, a hard limiting thresh-old function or a smoothly limiting threshold such as a sigmoid or a hyperbolic tangent. Due to their inherent properties, whether they are linear, continuous or differentiable, different activation functions perform with different efficiency in task-specific solutions.

For classification tasks antisymmetric sigmoid tangent hyperbolic function is the most popularly used activation function shown in Figure 1.

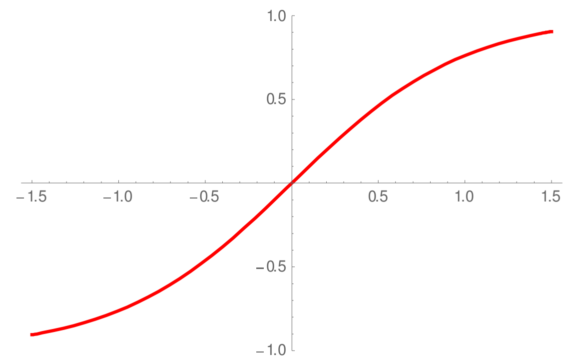

Figure 1: Antisymmetric sigmoid tangent hyperbolic activation function

4. EXPERIMENTAL RESULTS

The main goal of this paper was to understand how different classifiers would behave when encountering the chosen data and to compare their performance. In this study, neural network and support vector machine have been used for classification of Parkinson's disease.

The dataset is composed of a range of biomedical voice measurements from 31 people, 23 with PD. The data set used in this study is very unbalanced, where out of 195 samples, 147 are Parkinson's disease type and others 48 represent healthy people. The main problem with imbalanced data set is that it is very difficult to train to predict the presence of Parkinson's disease, since the ratio between classes is 3:1.

Table 2: Neural Network classification accuracy

o	Healthy	Sick	Correct (%)
Healthy	100	0	100
Sick	25	75	75

Table 1: Support Vector Machine classification accuracy

o	Healthy	Sick	Correct (%)
Healthy	83	17	83
Sick	22	78	78

It can be seen from neural network results that this method yields very good classification result, having accuracy of 75% from first training and 100% form second training, also Table 2 shows the accuracy of PD and non-PD, 75% and 100% respectively. Support vector machine results are determined as 70% from first test, 80% from second test, and Table 3 represents the accuracy of PD and non-PD, 83% and 78%, respectively. Comparing the presented results with those reported in other studies one can notice that the proposed method gives excellent results, considering the fact that is applied on vary imbalanced dataset with small number of samples.

5. CONCLUSIONS

It can be concluded from this study that neural networks and support vector machine, the results are reported as 100% for NN and 80% for SVM in an imbalanced data set, and it is seen that NN has a full accuracy of classification between PD and non-PD.

REFERENCES

Baldi P and Brunak S (1998). Bioinformatics: the Machine Learning Approach. MIT Press, Cambridge, MA.

Brown MPS, Grundy WN, Lin D, Cristianini N, Sugnet C, Ares M Jr, Haussler D (2000). Knowledge-based analysis of microarray gene expression data by using Support Vector Machines. Proc. Natl Acad. Sci. USA 97:262-267.

Can M (2013). Diagnosis of Parkinson's disease by Boosted Neural Networks. South East Europe Journal of Soft Computing 2: 7-13.

Cassidy M, Mazzone P, Oliviero A, Insola A, Tonali P, Lazzaro VD, et al. (2002). Movement-related changes in synchronization in the human basal ganglia. Brain, 125(6): 1235-1246.

Durbin R, Eddy S, Krough A, Mitchison G (1998). Biological Sequence Analysis. Cambridge University, New York.

Freddie A and Koker R (2011). A parallel neural netwoek approach to prediction of Parkinson's Diease. Expert System with Applications 38: 12470-12474.

Gasson MN, Wang SY, Aziz TZ, Stein JF, Warwick K (2005). Towards a demand driven deep-brain stimulator for the treatment of movement disorders. In: 3rd IEE International Seminar on Medical Applications of Signal Processing: 83-86.

Hebb DO (1949). The Organization of Behavior. New York: John Wiley & Sons. Introduction and Chapter 4 reprinted in Anderson & Rosenfeld (1988), pp. 45-56.

Jankovic J (2003). Pathophysiology and Clinical assessment of parkinsonian symptoms and signs. In Handbook of Parkinson's disease 3rd edition. Edited by: Pahwa R, Lyons K, Koller WC. Informa Health Care: 71-98.

Lee BJ (1997). Parallel neural networks for speech recognition. Neural Networks, International Conference: 2093-2097.

Little MA, McSharry PE, Roberts SJ, Costello DAE, Moroz IM (26 June 2007). Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection. BioMedical Engineering OnLine 6:23.

McCulloch WS, Pills W (1943). A Logical Calculus of the Ideas Immanent in Nervous Activity. Bulletin of Mathematical Biophysics 5:115-133. Reprinted in Anderson & Rosenfeld (1988), pp. 18-28.

Minsky ML, Papert SA (1969). Perceptrons, Expanded Edition. Cambridge, MA: MIT Press. Original edition, 1988.

Pan S, Warwick K, Stein J, Gasson MN, Wang SY, Aziz TZ, Burgess J (2007). Prediction of Parkinson's disease tremor onset using artificial neural networks. In Proceedings of the fifth IASTED International Conference: biomedical engineering: 341-345.

Rango M, Canesi M, Ghione I, Farabola M, Righini A, Bresolin N, Antonini A, Pezzoli G (2006). Parkinson's disease, chronic hydrocarbon exposure and striatal neuronal damage: A 1-H MRS study. Neurotoxicology27:164-168.

Rosenblatt E (1958). The Perceptron: A probabilistic model for information storage and organization in the brain. Psychological Review 65: 386-408.

Shashua A (2009). Introduction to Machine Learning: Class Notes 67577. Cornell University Library.

Vapnik V (1995). The Nature of Statistical Learning Theory. Springer, New York.

Vert JP, Koji T, Bernhard S (2004).A primer on kernel methods.Kernel Methods in Computational Biology.

Wu D, Warwick K, Ma Z, Burgess JG, Pan S, Aziz TZ (2010). Prediction of Parkinson's disease tremor onset using radial basis function neural networks. Expert Systems with Applications 37(4): 2923-2928.

Wu D, Warwick K, Ma Z, Gasson MN, Burgess JG, Pan S, et al. (2010). Prediction of Parkinson's disease tremor onset using a radial basis function neural network based on particle swarm optimization. International Journal of Neural Systems 20(02): 109-119.

Yin-Wen C, Cho-Jui H, Kai-Wei C, Michael R, Chih-Jen L (2010). Training and testing low-degree polynomial data mappings via linear SVM. Journal of Machine Learning Research 11: 1471-1490.