World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:6, No:4, 2012

Parkinsons Disease Classification using Neural Network and Feature selection

Anchana Khemphila and Veera Boonjing

Abstract—In this study, the Multi-Layer Perceptron (MLP)with Back-Propagation learning algorithm are used to classify to effective diagnosis Parkinsons disease(PD).It's a challenging problem for medical community.Typically characterized by tremor, PD occurs due to the loss of dopamine in the brains thalamic region that results in involuntary or oscillatory movement in the body. A feature selection algorithm along with biomedical test values to diagnose Parkinson disease.Clinical diagnosis is done mostly by doctor's expertise and experience.But still cases are reported of wrong diagnosis and treatment.Patients are asked to take number of tests for diagnosis.In many cases,not all the tests contribute towards effective diagnosis of a disease.Our work is to classify the presence of Parkinson disease with reduced number of attributes.Original,22 attributes are involved in classify.We use Information Gain to determine the attributes which reduced the number of attributes which is need to be taken from patients.The Artificial neural networks is used to classify the diagnosis of patients.Twenty-Two attributes are reduced to sixteen attributes.The accuracy is in training data set is 82.051% and in the validation data set is 83.333%.

Keywords—Data mining , classification , Parkinson disease , Artificial neural networks , Feature Selection , Information Gain

I. Introduction

The use of classifier systems in medical diagnosis is increasing gradually. Recent advances in the field of artificial intelligence have led to the emergence of expert systems and Decision Support Systems (DSS) for medical applications. Moreover, in the last few decades computational tools have been designed to improve the experiences and abilities of doctors and medical specialists in making decisions about their patients. Without doubt the evaluation of data taken from patients and decisions of experts are still the most important factors in diagnosis. However, expert systems and different Artificial Intelligence (AI) techniques for classification have the potential of being good supportive tools for the expert. Classification systems can help in increasing accuracy and reliability of diagnoses and minimizing possible errors, as well as making the diagnoses more time efficient [1].

Parkinson's Disease (PD) is the second most common neurodegenerative affliction after Alzheimer's disease (AD).PD is a progressive neurological disorder characterised by tremor, rigidity, and slowness of movements. It is associated with progressive neuronal loss in the substantia nigra and other brain structures. Non-motor features, such as dementia and dysautonomia, occur frequently, especially in advanced stages of the disease.Diagnosis depends on the presence of two or more cardinal motor features such as rest tremor, bradykinesia, or rigidity [2].Having so many factors to analyze to diagnose PD, specialist normally makes decisions by evaluating the current test results of their patients. Moreover, the previous decisions made on other patients with a similar condition are also done by them. These are complex procedures, especially when the number of factors that the specialist has to evaluate is high (high quantity and variety of these data). For these reasons, PD diagnosis involves experience and highly skilled specialists. Our work is an attempt to introduce a classification approach using Multi-Layer Perceptron (MLP)with Back-Propagation learning algorithm and a feature selection using information gain with Parkinson disease patients.

In the next section,we review the the data mining technique.Section 3 explains the data set used and discusses on reducing the number of attributes using Information Gain .Section 4 is result of experiment. Finally, section 5 contains concluding.

Anchana Khemphila is a PhD student in Computer Science at King Mongkuts Institute of Technology Ladkrabang Thailand (e-mail: s0067103@kmitl.ac.th)

Veera Boonjing is an Associate Professor Dr. in Computer Science, Department of Computer Science,Faculty of Science at King Mongkuts Institute of Technology Ladkrabang, Thailand. Also, he is working in National Centre of Excellence in Mathematics, PERDO, Bangkok, Thailand 10400, (e-mail: kbveera@kmitl.ac.th)

II. Data mining technique

Few works have been published M.A. Little et al.(2008) study Suit- ability of dysphonia measurements for telemonitoring of Parkinson's disease [3]. E. Tolosa, G. Wenning, and W. Poewe.(2006)study the diagnosis of Parkinson's disease. Lancet Neurology [4].GA Ivanitsky and RA Naumov. recognite(2008) of ongoing mental activity with artificial neural network [5]. In this research,we would like to mention about classification using Artificial neural networks with feature selection.

A. Artificial neural networks(ANNs)

Artificial neural networks is inspired by attempts to simulate biological neural systems. The human brain consists primarily of nerve cells called neurons,linked together with other neurons via stand of fiber called axons. Axons are used to transmit nerve impulses from one neuron to another whenever the neurons are stimulated. A neuron is connected to the axons of other neurons via dendrites,which are extensions from the cell body of the neurons.The contact point between a dendrite and an axon is called a synapse. Neural networks provide a very general way of approaching problems.When the output of the network is continuous,such as the appraised value of a home,then it is performing prediction.When the output

has discrete values,then it is doing classification.A simple re-arrangement of the neurons and the network becomes adept at detecting clusters.The fact that neural networks are so versatile definitely accounts for their popularity.The effort needed to learn how to use them and to learn how to massage data is not wasted,since the knowledge can be applied wherever neural networks would be appropriate [6].

Multilayer is feed-forward neural networks trained with the standard back-propagation algorithm.It is supervised networks so they require a desired response to be trained.It learns how to transform input data in to a desired response,so they are widely used for pattern classification.With one or two hidden layers,they can approximate virtually any input-output map. It has been shown to approximate the performance of optimal statistical classifiers in difficult problems.The most popular static network in the multilayer.The multilayer is trained with error correction learning,which is appropriate here because the desired multilayer response is the arteriographic result and as such known.Error correction learning works in the following way from the system response at neuron j at iteration t, $y_j(t)$,and the desired response $d_j(t)$for given input pattern an instantaneous error $e_j(t)$is defined by

$$e_j(t) = d_j(t) - y_j(t) \tag{1}$$

Using the theory of gradient descent learning, each weight in the network can be adapted by correcting the present value of the weight with a term that is proportional to the present input and error at the weight, i.e.

$$w_{jk}(t+1) = w_{jk}(t) + \eta\delta_j(t)x_k(t) \tag{2}$$

The $\eta(t)$ is the learning-rate patameter.The $w_{jk}(t)$ is the weight connecting the output of neuron k to the input neuron j at iteration t.The local error$\delta_j(t)$ can be computed as a weighted sum of errors at the internal neurons.

III. Classification accuracy among data mining technique.

A. Description of the data.

The Parkinson database used in this study is taken from the University of California at Irvine (UCI) machine learning repository [7].The dataset was created by Max Little of the University of Oxford, in collaboration with the National Centre for Voice and Speech, Denver, Colorado, who 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.This study reviewed the literature and used the following 23 variables as explanatory variables in Table.1.Before building models,the data set were randomly split into two subsets,60 %(n=117) of the the data for training set and 40 %(n=78) of the data for validation set.

TABLE I
Parkinsons Patient Data.

Attribute	Description
MDVP:Fo(Hz)	Average vocal fundamental frequency
MDVP:Fhi(Hz)	Maximum vocal fundamental frequency
MDVP:Flo(Hz)	Minimum vocal fundamental frequency
MDVP:Jitter(%) MDVP:Jitter(Abs) MDVP:RAP MDVP:PPQ Jitter:DDP	Several measures of variation in fundamental frequency
MDVP:Shimmer MDVP:Shimmer(dB) Shimmer:APQ3 Shimmer:APQ5 MDVP:APQ Shimmer:DDA	Several measures
RPDE D2	Two nonlinear dynamical complexity measures
DFA	Signal fractal scaling exponent
spread1 spread2 PPE	Three nonlinear measures of fundamental frequency variation
NHR HNR	Two measures of ratio of noise to tonal components in the voice
Concept Class	Healthy,Sick

B. Classifier evaluation measures.

In our work, context selection is determined by the weight of each low-level context.Instead of learning weights through a generic algorithm or other machine learning method, we use the information gain of each attribute, as its weight. The basic motivation for this study comes from the power of ANN classification algorithm and it uses the concept of information gain as the criterion to select an attribute.IG is usually used for feature set selection so the method applied consists of computing the IG for each field.IG give attribute X with respect to the class attribute Y is the reduction in uncertainty about the value of Y when we know the value of X,I(Y;X).The uncertainty about the value of Y when we know the value of X is given by the conditional entropy of Y given X,H(Y|X).

$$IG(Y;X) = H(Y) - H(Y|X) \tag{3}$$

When Y and X are discrete variables that take values in $y_1...y_k$ and $x_1...x_l$ then the entroy of Y is given by:

$$H(Y) = -\sum_{i=1}^{i=k} P(Y = y_i)log_2(P(Y = y_i)) \tag{4}$$

The condition entropy of Y given X is:

$$H(Y|X) = -\sum_{j=1}^{j=l} P(X = x_j)H(Y|X = x_i)) \tag{5}$$

alternatively the information gain is given by:

$$IG(Y;X) = H(X) + H(Y) - H(X,Y) \quad (6)$$

The ranking of the attributes is then done with respect to the values of IG in a descending order, reflecting the intuition that the higher an IG value, the more information the corresponding attribute has to offer regarding the class. Note that to compute the information gain, data sets with numeric features are required to be discretized. Many alternative methods can be applied for this. In the present work, we simply handle the discretizaiton of continuous valued attributes by partitioning the range of values into a finite number of subsets.

We have trained the classifiers to classify the medical data set as either healthy or sick. The accuracy of a classifier can be computed using sensitivity and specificity.For the given two classes, we consider in terms of positive tuples (diagnosis =healthy) versus negative tuples (eg., diagnosis = sick).True positives refer to the positive tuples that were correctly labeled by the classifier, while true negatives are the negative tuples that were correctly labeled by the classifier. False positives are the negative tuples that were incorrectly labeled by the classifier, while false negatives are the positive tuples that were incorrectly labeled by the classifier. The sensitivity and specificity measures can be used for above purpose and precision is used for the percentage of samples labeled as healthy. The sensitivity(SEN) is given by:

$$SEN = \frac{t-pos}{pos} \quad (7)$$

t-pos is the number of true positives (i.e healthy samples that were correctly classified) and pos is the number of positive (healthy) samples.The specificity(SPE) is given by:

$$SPE = \frac{t-neg}{neg} \quad (8)$$

t-neg is the number of true negatives (i.e sick samples that were correctly classified) an neg is the number of positive (sick) samples and f-pos is the number of false positives (sick samples that were incorrectly labeled as healty).The Accuracy(ACC) is given by:

$$ACC = SEN\frac{pos}{pos+neg} + SPE\frac{neg}{pos+neg} \quad (9)$$

The true positives, true negatives, false positives and false negatives are also useful in assessing the costs and benefits (or risks and gains) associated with a classification model.In Table 2,the alternatively the accuracy is given by:

$$Accuracy = \frac{n_{0,0}+n_{1,1}}{n} \quad (10)$$

TABLE II
CONFUSION MATRIX

	Predicted class	
Actual class	C_0	C_1
C_0	$n_{0,0}$	$n_{0,1}$
C_1	$n_{1,0}$	$n_{1,1}$

where $n_{0,0}$ is number of C_0 cases classified correctly
$n_{1,1}$ is number of C_1 cases classified correctly

The Lift Chart measures the effectiveness of models by calculating the ratio between the result obtained with a model and the result obtained without a model.In the lift chart,it is represented by the random curve.It shows you the lift factor to how many times it is better to use a model in contrast to not using a model.Lift Chart is well know in the data mining community specialized in marketing and sales applications [8].The greater the area between the model curve and the baseline curve,the better model

IV. EXPERIMENT AND RESULT.

Experiments were calculated information gain with Weka 3.6.6 tool. We adopt information gain to sort the features according to their importance for classification in the following experiment. To calculate information gain, the input must be discrete numbers. Since the inputs in our experiment are continuous real numbers, we handled the discretization of continuous valued attributes by partitioning the range of values into a finite number of subsets.A features information gain is proportional with its weight to deduction.The ranking information gains of 22 attributes are shown in table 3.We first directly used ANN with no information gain based feature selection function.The result is shown in table 4 is the accuracy.The accuracy of the main model is in training data set is 91.453%, in the validation data set is 80.769%.Then,we deduct feature which has the lowest IG and use ANN to classify.If the classify accuracy is higher or equal than the accuracy of main model.We deduct next feature which has the second lowest IG and use ANN to classify. We made the loop until the classify accuracy is less than the accuracy of main model.Table.4 shown that the left column is the feature number balance used in experiment.We deduct the features until it has 16 features.The accuracy is in training data set is 82.051%, in the validation data set is 83.333%.

Figure 1,2 show the Lift Chart output.The X-axis shows the commulative while the Y-axis shows the cases.The red line is commulative class when sorted using average.The black-dash line is commulative class when sorted using predicted values with 22 features.And the blue line is commulative class when sorted using predicted values with 16 features.After we reduce the feature, the graph lines are still similarly.

V. CONCLUSION.

So far,several studies have been reported focusing on PD diagnosis.In these studies different methods were applied to the given problems.We use Information gain to filter features which do not need to take from patients.ANN is used to

TABLE III
INFORMATION GAIN IN EACH FEATURE

Item	Attribute	IG
1	PPE	0.3296
2	spread1	0.3039
3	MDVP:Fo(Hz)	0.3009
4	MDVP:Flo(Hz)	0.2369
5	MDVP:APQ	0.2055
6	spread2	0.1975
7	Shimmer:APQ5	0.1837
8	MDVP:Shimmer	0.1815
9	MDVP:Shimmer(dB)	0.1582
10	Shimmer:DDA	0.1514
11	Shimmer:APQ3	0.1514
12	MDVP:Fhi(Hz)	0.1498
13	HNR	0.1483
14	DFA	0.1478
15	MDVP:Jitter(Abs)	0.1201
16	MDVP:PPQ	0.1168
17	D2	0.1147
18	RPDE	0.1118
19	MDVP:Jitter(%)	0.0983
20	MDVP:RAP	0.0868
21	Jitter:DDP	0.0868
22	NHR	0.0216

TABLE IV
ACCURACY OF ANN CLASSIFIER USED FEATURE SELECTION USING PARKINSONS DISEASE DATA SET

Features	Training	Validation
22	91.453%	80.769%
16	82.051%	83.333%

classify the Parkinson's disease.The experiment shows that feature selection helps increase computational efficiency while improving classification accuracy.Besides,they decrease the complexity of the system by reducing the dataset.The ANN can be improved by selecting good combination of input variables.It is shown that reduction of input variable reduces number of computational requirement,save repository size,save health checklist costing.

REFERENCES

[1] M.F. Akay. Support vector machines combined with feature selection for breast cancer diagnosis. Expert Systems With Applications, 2008.
[2] AJ Hughes, SE Daniel, L. Kilford, and AJ Lees. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. British Medical Journal, 55(3) 181184, 1992.
[3] M.A. Little, P.E. McSharry, E.J. Hunter, J. Spielman, and L.O. Ramig. Suit- ability of dysphonia measurements for telemonitoring of Parkinson? s disease. IEEE transactions on bio-medical engineering, 2008.
[4] E. Tolosa, G. Wenning, and W. Poewe. The diagnosis of Parkinson's disease. Lancet Neurology, 5(1):7586, 2006.
[5] GA Ivanitsky and RA Naumov. Recognition of ongoing mental activity with artificial neural network. International Journal of Psychophysiology, 69(3): 180180, 2008.
[6] Data Mining Techniques For Marketing,Sales, and Customer Support.John Wiley & Sons,Inc.
[7] A. Asuncion and D.J. Newman. UCI machine learning repository, 2007. URL http://www.ics.uci.edu/∼mlearn/MLRepository.html.
[8] Berry, M.J.A. and Linoff, G. (1999): Data Mining Techniques: For Marketing, Sales, and Customer Support. Morgan Kaufmann Publishers.

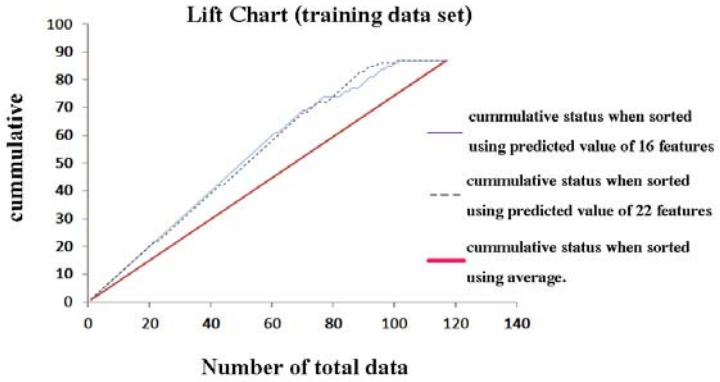

Fig. 1. Lift chart of Parkinson in Training Data set.

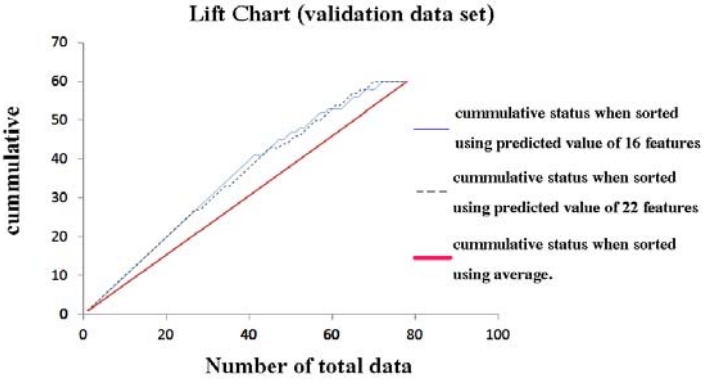

Fig. 2. Lift chart of Parkinson in Validation Data set.

Anchana Khemphila recieved her diploma in computer science from Mahasarakham University,a master degree in information engineering from King Mongkut's Institute of Technology Ladkrabang.

Veera Boonjing recieved his diploma in mathematic from Ramkhamhaeng University and in economics from Sukhothai Thammathirat Open University(second honors),a master degree in computer science from Chulalongkorn University. and Ph.D in Decision Sciences and Engineering Systems from Rensselaer Polytechnic Institute, USA.He is currently working as an Associate Professor Dr. in department of Computer Science,faculty of Science at King Mongkut's Institute of Technology Ladkrabang.Also, he is working in National Centre of Excellence in Mathematics, PERDO, Bangkok, Thailand 10400