Performance evaluation of machine learning algorithms in the classification of parkinson disease using voice attributes

Article *in* International Journal of Applied Engineering Research · January 2017

CITATIONS
26

READS
197

2 authors, including:

Rajagopalan S P
Dr.MGR Educational and Research Institute
182 PUBLICATIONS **940** CITATIONS

SEE PROFILE

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 21 (2017) pp. 10669-10675

Performance Evaluation of Machine Learning Algorithms in the Classification of Parkinson Disease Using Voice Attributes

J. Sujatha
Research Scholar, Vels University,
Assistant Professor, Post Graduate Department of Information Technology,
Bhaktavatsalam Memorial College for Women, Chennai-80, Tamil Nadu, India.
Orcid Id: 0000-0002-2518-6038

Dr. S.P. Rajagopalan
Professor, Department of Computer Science,
GKM College of Engineering & Technology, Chennai–63, Tamil Nadu, India.
Orcid Id: 0000-0003-2621-4069

Abstract

Nerve cells, the building blocks of the nervous system in the brain don't reproduce when damaged. On damage, the dopamine produced by these nerve cells are not produced which hinders motor skills and speech. Voice undergoes changes at an earlier stage before the brain cells are affected [1]. Voice changes helps to identify Parkinson disease at an initial stage thereby preventing damage to the brain cells which would result in reduced coordination and movement. The idea of this paper is to evaluate the performance of various data mining classification techniques used in the identification of Parkinson disease.

Keywords: Dopamine, Attributes, Sensitivity, Specificity, Mining

INTRODUCTION

Data Mining

Data mining is the evolution of intelligent algorithms to perform tasks and make decisions without human intervention. It involves three disciplines namely statistics, artificial intelligence and machine learning. Statistics involves data collection, study and analysis of existing data and interpretation of result. AI involves human thought like processing to statistical problems. ML combines statistics and AI. It learns from the input data and uses this knowledge to make intelligent decision for a new input data.

Data mining methods is classified as predictive data mining and descriptive data mining. Predictive data mining includes classification and prediction based on knowledge gained so far. Descriptive data mining includes clustering, association analysis which identifies patterns and relationship in the data.

Machine Learning Algorithms:

Machine learning is a novel concept that helps a computer program to automatically learn and adapt to new data without human interference and without being explicitly programmed. It analyzes the given data and helps the program to make decision based on the studied data. It learns and improves its knowledge from the previous result just as a human. It stimulates complex neural network in the human brain and derives its own rules after handling large amount of data. ML methods can be classified into three types as Supervised learning, semi supervised learning and unsupervised learning.

Different machine learning algorithms are available in WEKA. WEKA stands for Waikato Environment for Knowledge Analysis developed at the University of Waikato in New Zealand. The software is open source and it is written in Java Language. The dataset is in CSV(comma separated value) file which is converted to ARFF (Attribute Relational File Format) file using ARFF viewer. Click the Explorer view in WEKA. In the WEKA Explorer window, select the preprocess tab and select the ARFF file created. Now in the classify tab of the WEKA Explorer, click on choose button. Different machine learning algorithms are available.

LITERATURE SURVEY

Related works have been carried out by other authors. In Diagnosis and Classification of Parkinsons Disease Using Data Mining Techniques[2], sensitivity, specificity and accuracy were calculated for the voice features using Bayes Net, Naïve Bayes, MLP, SMO and J48 algorithm. MLP gave the highest accuracy, specificity and sensitivity with values of 97.8, 100% and 95.75%. In Diagnosing Parkinson Disease by using Artificial Neural Networks and Support Vector Machines[3], accuracy, sensitivity, specificity, positive predicted value and negative predicted value were calculated. MLP had a high value of "Positive predictive value" equal to 95.83%, SVM with linear kernel had the highest value of "Sensitivity" equal to 99.32% and "Negative predictive value" equal to 97.06%. Finally, SVM with kernel puk presented the highest values of "Classification of accuracy" equal to 93.33% and "Positive predictive value" equal to 96.53%. In An

Improved Approach for Prediction of Parkinson's Disease using Machine Learning Techniques[4], it was observed that all the classifiers performed reasonably well with boosted logistic regression giving the best performance with 97.16% accuracy and 98.9% area under the ROC(AUC). It was also found that the accuracy and area under the ROC curve are nearly the same among the different classifiers used. In A Survey of Parkinson's Disease Using Data Mining Algorithms [5], review of related work was done. Random tree gave highest classification accuracy of 100% which was done using Tanagra data mining tool.

VOICE AND PARKINSON'S DISEASE

Speaking process involves 4 stages - Respiratory stage, Phonation stage, Resonation stage, Articulation stage. Respiratory stage involves breathing and consists of 2 phases' inhalation and exhalation. In phonation stage, voice is produced in speaking as the expiratory air stream from lungs goes up through the trachea or wind pipe to the larynx (voice box). Voice produced in phonation is weak. So it is amplified by resonators. Resonation is voice amplification and modification. Tone thus produced in larynx is changed to specific sound by the articulators. Articulators help in production of different sound.

The Parkinson's disease dataset consists of biomedical voice measurement from the patients. Various voice measures are selected and each row corresponds to a voice recording of a patient. The voice sound produced during the speech test is recorded using a microphone and recorded signals are analyzed.

Traditional measures are calculated in Praat software and the rest were computed by the Kay Pentax Multi-Dimensional Voice Program (MDVP). The jitter and period perturbation derived from the sequence of frequencies for each vocal cycle, the shimmer and amplitude perturbation measures derived from the sequence of maximum extent of the amplitude of the signal within each vocal cycle, the noise-to-harmonics (and harmonics-to-noise) ratios derived from the signal-to-noise estimates are explained clearly by Little Max.[6]

The aim of the data is to classify the patients as healthy and PD and evaluate the performance. Change in voice is the simplest cost effective way to identify the disease at a very early stage before it damages the brain cells and produces visible symptoms. The features used to classify Parkinson's disease using voice detection are fundamental frequency, amplitude, jitter, Noise to Harmonic ratio, Harmonic to noise ratio, DFA and spread. Most of the attributes shown in the table below are produced during the phonation stage.

Classifiers Used

A total of 16-17 classification algorithms have been used in this work. The classifiers in Weka are divided into different groups such as Bayes, Functions, Lazy, Rules, Tree based classifiers etc. Algorithms from each group have been chosen in this work.

Parkinson's dataset attributes

These are the vocal attributes that are used to identify Parkinson's disease.

Table 1: Parkinson's dataset attributes

Name	**ASCII subject name and recording number**
MDVP: Fo(Hz)	Fundamental frequency of standard vocal
MDVP: Fhi(Hz)	Fundamental frequency of greatest vocal
MDVP: Flo(Hz)	Fundamental frequency of lowest vocal
MDVP: Shimmer (dB)	Amplitude of peak-to-peak in terms of decibels
Shimmer:APQ3	Quotient of amplitude perturbation in 3-point.
Shimmer:APQ5	Quotient of amplitude perturbation in 5-point.
MDVP: Jitter(%)	Inconsistency of period-to-period.
MDVP: Jitter (Abs)	Fundamental frequency variation in cycle-to cycle
Jitter (relative)	Difference between the average period average and consecutive periods.
MDVP: RAP	Comparative Perturbation
MDVP: PPQ	Perturbation Quotient of Period in 5-point.
NHR	Ratio of Noise-to-Harmonic
HNR	Ratio of Harmonic-to- Noise
DFA	Detrended fluctuation analysis-based on casual walk
Spread1, Spread2, PPE	Quantify the fundamental frequency in variation.
Status	Health status of the subject 1- Parkinson's 0- healthy

PERFORMANCE EVALUATION

The efficiency of any machine learning is determined using measures such as true positive rate, false positive rate, true negative rate and false negative rate. Sensitivity and specificity is used to explain clinical diagnostic test and to estimate how good the test was. Training and evaluating statistical performance on same data yields over optimistic result. Cross validation improves performance accuracy. Cross validation is a statistical method that compares machine learning schemes by dividing data into train and test set. The train set is used to train the data and test set is used to validate the model. In k-fold cross validation, data is partitioned into k equal size folds. The k iterations are trained and validated such that within each iteration, different fold of the data is held out for validation and remaining k-1 fold is used for learning.

For accuracy, sensitivity and specificity, the terms TP, TN, FP and FN are used.

1) TP represents the number of people who are predicted positive and are actually positive.
2) TN represents the number of people who are predicted negative and are actually negative.
3) FP represents the number of people who are predicted positive and are actually negative.
4) FN represents the number of people who are predicted negative and are actually positive.

Confusion Matrix:

Confusion matrix also called as Contingency table is expressed as follows:

Table 2: Confusion matrix

Actual	Predicted	
	Positive	Negative
Positive	True Positive	False Negative
Negative	False Positive	True Negative

The number of correctly classified instances is the sum of diagonals in the matrix and all others are incorrectly classified.

Accuracy/Confidence/Success Rate:

This compares how close a new test value is to a value predicted.

$$ACC=\frac{TP+TN}{P+N}=\frac{TP+TN}{TP+TN+FP+FN}$$

Predictive accuracy is the proportion of correctly classified cases to all cases in the set.

True Positive Rate/Sensitivity/Recall:

This measures ability of a test to be positive when condition is actually present

$$\text{Recall /Sensitivity/TPR}=\frac{TP}{P}=\frac{TP}{TP+FN}$$

Where TPR is called the True Positive Rate which is the proportion of items classified as class X among all examples which truly have class X.

True Negative Rate/Specificity:

This measures ability of a test to be negative when condition is actually not present. It is the reverse of sensitivity.

$$\text{Specificity/TNR}=\frac{TN}{N}=\frac{TN}{TN+FP}=1-FPR$$

Where TNR is called the True Negative Rate which is the ratio of true negatives to actual negatives.

Positive Predicted Value/Precision:

$$\text{Precision} = \frac{TP}{TP+FP} \text{ or PPV}=\frac{TP}{TP+FP}$$

Where PPV is the Positive Predicted Value. Precision is the fraction of predicted diseases that are diseased while recall is the fraction of diseased that are predicted diseased.

Negative Predicted Value:

$$NPV=\frac{TN}{TN+FN}$$

Where NPV is the negative predicted value.

False Positive Rate:

$$FPR=\frac{FP}{N}=\frac{FP}{FP+TN}=1-TNR \text{ OR FALL OUT}$$

False Positive Rate is the ratio of false positive to actual negatives ie., what fraction of those that are actually negative were found to be positive.

False Negative Rate:

$$FNR = \frac{FN}{N} = \frac{FN}{FN+TP} \text{ OR MISS RATE}$$

F-Score:

F-Score is the Harmonic mean of recall and precision

$$F= 2*\frac{Precision*Recall}{Precision+Recall}=\frac{2TP}{2TP+FP+FN}$$

Mathew Correlation Coefficient:

$$MCC= \frac{TP*TN-FP*FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

Kappa Score:

The kappa statistic measures the agreement of prediction with the true class -- 1.0 signifies complete agreement

$$Kappa=\frac{Total\ Accuracy-Random\ Accuracy}{1-Random\ Accuracy}$$

Mean Absolute Error (MAE):

The MAE measures the average magnitude of the errors in a set of forecasts, without considering their direction. It measures accuracy for continuous variables. The MAE is a linear score which means that all the individual differences are weighted equally in the average. It calculates the closeness between the predictions to the actual outcome and is the average of the absolute errors.

$$MAE=\frac{1}{N}\sum_{i=1}^{N}|\hat{\theta}_i-\theta_i|$$

where $\hat{\theta}_i$ is the prediction value and θ_i is the true value.

Root mean squared error (RMSE):

Root mean squared error also called as root mean squared deviation is a measure to calculate the values predicted by a model when compared to the actual observed values.

$$RMSE=\sqrt{\frac{1}{N}\sum_{i=1}^{N}|\hat{\theta}_i-\theta_i|}$$

Relative Absolute error (RAE):

Relative Absolute error is just the total absolute error with the same kind of normalization. The relative absolute error takes the total absolute error and normalizes it by dividing by the total absolute error of the simple predictor.

Root Relative Squared error (RRSE):

Relative squared error takes the total squared error and normalizes it by dividing by the total squared error of the default predictor.

Area Under Curve (AUC):

Area Under Curve measures the performance of a binary classification. A higher AUC is good.

$$AUC=\frac{1}{2}\left(\frac{TP}{TP+TN}+\frac{TN}{TN+TP}\right)$$

Receiver Operation Characteristics (ROC):

ROC curve compares the classifiers performance across the entire range of class distributions. ROC curve is created by plotting the true positive rate against the false positive rate. It has the following property.

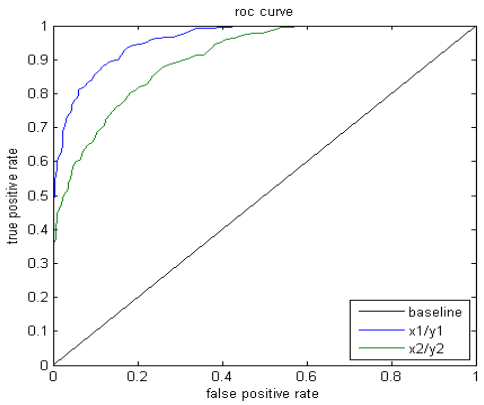


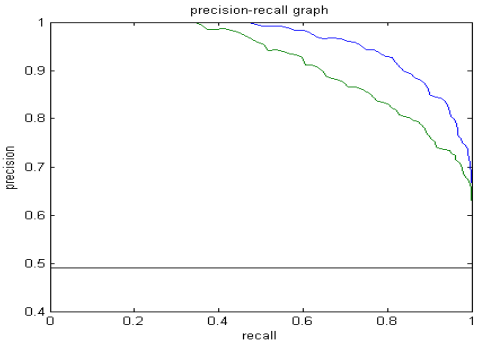
Table 3: ROC Values

Value	ROC Area
1	perfect prediction
0.9	excellent prediction
0.8	good prediction
0.7	mediocre prediction
0.6	poor prediction
0.5	random prediction
<0.5	Something wrong

Precision Recall Curve (PRC):

Precision-Recall is a useful measure of success of prediction when the classes are very imbalanced. In information retrieval, precision is a measure of result relevancy, while recall is a measure of how many truly relevant results are returned.

The precision-recall curve shows the tradeoff between precision and recall for different threshold. A high area under the curve represents both high recall and high precision, where high precision relates to a low false positive rate, and high recall relates to a low false negative rate. High scores means that the classifier is returning accurate results (high precision), as well as returning a majority of all positive results (high recall).



RESULTS AND DISCUSSION

Precision, Recall, F-Score, MCC:

Algorithm	Confusion Matrix		Correctly Classified Instances	Percentage	InCorrectly Classified Instances	Percentage	TP Rate	Weighted TP Rate Average	FP Rate	Weighted FP Rate Average	Precision	Weighted Precision Average	Recall	Weighted Recall Average	f SCORE	Weighted f score Average	MCC	Weighted MCC Average
Bayes Net	118	29	156	80	39	20	0.8		0.21		0.92		0.8		0.86		0.54	
	10	38					0.79	0.8	0.2	0.21	0.57	0.83	0.79	0.8	0.66	0.81	0.54	0.5
Zero R /one R	147	0	147	75.38	48	24.62	1		1		0.75		1		0.86		0	
	48	0					0	0.75	0	0.75	0	0.57	0	0.75	0	0.65	0	0
RBF Classifier	140	7	166	85.13	29	14.87	0.95		0.46		0.86		0.95		0.91		0.57	
	22	26					0.54	0.85	0.05	0.36	0.79	0.85	0.54	0.85	0.64	0.84	0.57	0.6
HMM	147	0	147	75.38	48	24.62	1		1		0.75		1		0.86		0	
	48	0					0	0.75	0	0.75	0	0.57	0	0.75	0	0.65	0	0
Naïve Bayes	91	56	135	69.23	60	30.77	0.62		0.08		0.96		0.62		0.75		0.46	
	4	44					0.92	0.69	0.38	0.16	0.44	0.83	0.92	0.69	0.59	0.71	0.46	0.5
Logistics	135	12	169	86.67	26	13.33	0.92		0.29		0.91		0.92		0.91		0.64	
	14	34					0.71	0.87	0.08	0.24	0.74	0.86	0.71	0.87	0.72	0.87	0.64	0.6
RBF Network	136	11	162	83.08	33	16.92	0.93		0.46		0.86		0.93		0.89		0.51	
	22	26					0.54	0.83	0.07	0.36	0.7	0.82	0.54	0.83	0.61	0.82	0.51	0.5
MLP Classifier	129	18	162	83.08	33	16.92	0.88		0.31		0.9		0.88		0.89		0.55	
	15	33					0.69	0.83	0.12	0.27	0.65	0.83	0.69	0.83	0.67	0.83	0.55	0.6
RBF Classifier	140	7	166	85.13	29	14.87	0.95		0.46		0.86		0.95		0.91		0.57	
	22	26					0.54	0.85	0.05	0.36	0.79	0.85	0.54	0.85	0.64	0.84	0.57	0.6
SGD	141	6	169	86.67	26	13.33	0.96		0.42		0.88		0.96		0.92		0.62	
	20	28					0.58	0.87	0.04	0.32	0.82	0.86	0.58	0.87	0.68	0.86	0.62	0.6
SMO	145	2	172	88.21	23	11.79	0.99		0.44		0.87		0.99		0.93		0.66	
	21	27					0.56	0.88	0.01	0.33	0.93	0.89	0.56	0.88	0.7	0.87	0.66	0.7
IBK	142	5	188	96.41	7	3.59	0.97		0.04		0.99		0.97		0.98		0.91	
	2	46					0.96	0.96	0.03	0.04	0.9	0.97	0.96	0.96	0.93	0.96	0.91	0.9
LWL	134	13	164	84.1	31	15.9	0.91		0.38		0.88		0.91		0.9		0.56	
	18	30					0.63	0.84	0.09	0.3	0.7	0.84	0.63	0.84	0.66	0.84	0.56	0.6
Ada Boost	137	10	166	85.13	29	14.87	0.93		0.4		0.88		0.93		0.9		0.58	
	19	29					0.6	0.85	0.07	0.32	0.74	0.85	0.6	0.85	0.67	0.85	0.58	0.6
Decision Table	134	13	163	83.59	32	16.41	0.91		0.4		0.88		0.91		0.89		0.54	
	19	29					0.6	0.84	0.09	0.32	0.69	0.83	0.6	0.84	0.64	0.83	0.54	0.5
J48	129	18	157	80.51	38	19.49	0.88		0.42		0.87		0.88		0.87		0.47	
	20	28					0.58	0.81	0.12	0.34	0.61	0.8	0.58	0.81	0.6	0.8	0.47	0.5
Random Forest	145	2	172	88.21	23	11.79	0.99		0.44		0.87		0.99		0.93		0.66	
	21	27					0.56	0.88	0.01	0.33	0.93	0.89	0.56	0.88	0.7	0.87	0.66	0.7

Kappa Statistics, MAE, RMSE, RAE and RRSE

Algorithm	kappa statistics	mean absolute error	root mean squared error	relative absolute error	root relative squared error
BAYES NET	0.53	0.2	0.44	54.16	101
Zero R	0	0.37	0.43	100	100
RBF Classifier	0.55	0.27	0.35	72.9	80.59
HMM	0	0.5	0.5	134.18	116.1
Naïve Bayes	0.39	0.31	0.55	82.62	126.6
Logistics	0.64	0.18	0.34	48.02	78.06
RBF Network	0.51	0.23	0.35	60.742	82.02
MLP Classifier	0.55	0.2	0.36	52.62	84.12
SGD	0.6	0.13	0.36	35.8	84.8
SMO	0.63	0.12	0.34	31.7	79.7
IBK	0.91	0.04	0.19	11	43.7
LWL	0.56	0.24	0.36	63.5	84.2
Ada Boost	0.57	0.17	0.33	46.8	77.2
Decision Table	0.54	0.24	0.36	63.5	84.2
J48	0.47	0.2	0.43	54.2	99
Random Forest	0.63	0.24	0.31	64.9	71.8

AUC, ROC and PRC:

Algorithm	Confusion Matrix		AUC	ROC	PRC
BAYES NET	118	29	0.5	0.871	0.947
	10	38	0.5	0.871	0.752
Zero R	147	0	0.5	0.476	0.745
	48	0	0.5	0.476	0.237
RBF Classifier	140	7	0.5	0.871	0.957
	22	26	0.5	0.871	0.738
HMM	147	0	0.5	0.5	0.754
	48	0	0.5	0.5	0.246
Naïve Bayes	91	56	0.5	0.857	0.939
	4	44	0.5	0.861	0.735
Logistics	135	12	0.5	0.883	0.956
	14	34	0.5	0.883	0.728
RBF Network	136	11	0.5	0.857	0.932
	22	26	0.5	0.856	0.68
MLP Classifier	129	18	0.5	0.869	0.941
	15	33	0.5	0.869	0.682
SGD	141	6	0.5	0.771	0.871
	20	28	0.5	0.771	0.583
SMO	145	2	0.5	0.774	0.872
	21	27	0.5	0.774	0.631
IBK	142	5	0.5	0.967	0.985
	2	46	0.5	0.967	0.856
LWL	134	13	0.5	0.815	0.92
	18	30	0.5	0.815	0.6
Ada Boost	137	10	0.5	0.889	0.954
	19	29	0.5	0.889	0.793
Decision Table	134	13	0.5	0.833	0.918
	19	29	0.5	0.833	0.628
J48	129	18	0.5	0.769	0.881
	20	28	0.5	0.769	0.513
Random Forest	145	2	0.5	0.941	0.98
	21	27	0.5	0.941	0.873

The parameters discussed like confusion matrix, TP, FP, Precision, Recall, F1 score, MCC, kappa statistics, MAE, RMSE, RAE, RRSE were calculated for some of the important data mining algorithms like Zero R, One R,Bayes Net, RBF Classifier, Hidden Markow Model, Naïve Bayes, Logistics, Radial Basis Network, Multilayer Perceptron, Logistics, Ada Boost, Decision table, J48, Random forest etc The AUC, ROC and PRC values were calculated for the same.

FUTURE WORK

Performance evaluation of existing algorithm was discussed. A new algorithm using ANFIS classifier is to be implemented and the efficiency needs to be compared to the existing algorithm. This would help in identifying Parkinson patient with the help of the voice features. The proposed algorithm would act as a tool to identify Parkinson patient and helps the doctor in clinical decision support system.

CONCLUSION

The performance evaluation of machine learning algorithm is based on predictive accuracy and involves trade-off between true positive and true negative rate and between recall and precision. F1 score is the harmonic mean of recall and precision. ROC curve acts as graphical representation of trade off between false negative and false positive rate.

.

REFERENCES

[1] Kris Tjaden, Speech and Swallowing in Parkinson's Disease, Top Geriatr Rehabil 2008, 24(2) Pages 115-126.

[2] P. Durga et al, Diagnosis and Classification of Parkinsons Disease Using Data Mining Techniques, International Journal of Advanced Research Trends in Engineering and Technology, Vol. 3, Special Issue 14, March 2016.

[3] Diagnosing Parkinson Disease by using Artificial Neural Networks and Support Vector Machines, International Journal of Computer Science and Information Technologies, Global Journal of Computer Science and Technology, Page 63-71.

[4] An Improved Approach for Prediction of Parkinson's Disease using Machine Learning Techniques, International conference on Signal Processing, Communication, Power and Embedded System, 2016.

[5] Hariganesh S et al , A Survey of Parkinson's Disease Using Data Mining Algorithms, International Journal of Engineering and Innovative Technology (IJEIT), Volume 3, Issue 3, September 2013.

[6] M.A.Little, P.E. McSharry, E.J.Hunter, J.Spielmanand L.O.Ramig "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease" IEEE Transactions on Biomedical Engineering,2009