Assessment of Rehabilitative Speech treatment in Parkinson's Disease Using Non-Linear Speech analysis algorithms and Statistical Machine Learning.

Authors

Hassan, Gasser - Asaad, Mostafa – Fawzy, Muhammed – Hisham, Abdelrahman

1. Abstract

The speech based classification for Parkinson disease is an effective, non-invasive method for diagnosis of the disease. This has made a boost in speech pattern analysis in recent years to pave the way towards building a predictive tele-diagnostic model. The obstacles around this technique appears in the need for obstacles in optimizing classifications is to reduce noise within the collected speech samples, thus ensuring better classification accuracy and stability.

In this study, a PD classification algorithm was proposed and examined that utilizes nearest-neighbor algorithm and an ensemble learning algorithm. Also Statistical Models used in machine learning as Naïve Bayes algorithm was used to serve the same aim of assessing the applicability of speech treatment for PD.

2. Introduction

Parkinson's disease (PD) is a well-known progressive neurological disorder characterized by a large number of motor and non-motor features that can impact on function to a variable degree. It is the second most common age-related neuro-degenerative disease among the world, with an estimated number of patients ranging from 7~10 Million. The incidence of the disease –also – cannot be under estimated, as it affects 41 people per 100,000 in the fourth decade of life; however, this boosts to 1900 people per 100,000. Gender is also an influential factor, as the likelihood of men to have Parkinson is 1.5 more than women. Parkinson and Alzheimer patients can have clinical and pathological features of both diseases, so overlap might occur in diagnosis of both of them.

Till the recent moment, there is no exact causes behind loss of brain cells observed in PD patients. Drugs available in market can only delay degeneration of neurons causing damage of motor functions. Diagnosis of PD has not recorded a distinctive test, and only clinical features as (Rest

Tremor, Rigidity, Loss of postural reflexes…) are used for diagnosis. Effective treatment and early diagnosis for PD are hindered by a lack of quantifiable biomarkers and clear measures for the disease and its progression. Thus, if we can quantify this features, then we can have reliable method to diagnose – and may predict – Parkinson.

Scientific community has proposed a new model for diagnosis of PD. The model applies techniques of speech recognition and concepts of machine learning. This approach requires recording sustained voice from patients. The recording then undergoes analysis, and some attributes detect specific variations in sound vibrations linked to vocal tremors, breathlessness, and weakness. Calculated attributes are then compared to databases of recordings of Parkinson's patients and non-Parkinson's patients that serve as a control. Such voice changes are indicative of the neurological degeneration accompanying Parkinson.

The proposed modality is perceived to be more beneficial than conventional methods, it is planned to be ultra-low-cost and all analysis will be computerized with no intervention by any professionals or administrators. Also this technique will be of a great impact on those areas with low health infrastructure; since voice recordings can be obtained readily and transferred by telecommunication system. In addition to that, voice-based diagnosis will lead to early intervention and better management of PD; that's because voice changes in PD precedes physical disabling symptoms.

In this study concepts of "Machine Learning" were used; ML refers to the changes in systems that perform tasks associated with artificial intelligence (AI). The AI can involve many tasks within as, diagnosis, planning, robot control, prediction, etc. Machine learning depend on various techniques; one of them depend on giving the machine certain analyzed data to build its learning background, this is called Supervised ML, on the other side exists the non-supervised ML which depends on the system's ability to analyze data input and obtain its learning Background from. The aim of our study is to utilize certain machine learning algorithms to able to quantify some clinical features and hence, be able to diagnose cases suspected to have PD, and to discuss the accuracy of the used algorithm in comparison to others used worldwide.

3. Methodologies

3.1 Study Design:

- **Grouping**

This study was carried on a sample of 126 patients who have undergone voice test and some attributes were calculated accordingly. This study is a descriptive cross-sectional study, dealing with the samples from a retrospective point of view. The sample was taken at random to avoid any bias in the results.

- **Processing:**

We have chosen two algorithms to process our data**; Naïve Bayes statistical model & Key Nearest Neighbor algorithm**. Those two algorithms were utilized to serve the aim of our study through applying ML techniques. Our goal is to modify the code and assess the results in order to optimize the diagnosis of PD using voice recognition.

Since the used technique involves supervised ML, therefore, the given dataset is split into two groups: train dataset and test data set. The train dataset is used to provide the algorithm with the prior learning background, on the other side, the test dataset is to test the accuracy of the algorithm.

3.2 Naïve Bayes Algorithm:

Naïve Bayes algorithm – also known as Bayes Classifier – is a supervised learning method which allows us to capture uncertainty about the model in a principled way by determining probabilities of the outcomes. The algorithm can do both; diagnosis and prediction as well as being a useful perspective for understanding and evaluating many learning algorithms.

One of the most important prerequisites to apply Bayes Classifier is **independence** of events. Two or more events are said to be independent if the occurrence of any of them does not necessarily require the occurrence of the other. An example can show that:

Example: Suppose there are two events:

_ M: Manuela teaches the class (otherwise it's Andrew)

_ S: It is sunny

"The sunshine levels do not depend on and do not influence who is teaching."

This independence should be in our events in order to say that applying Naïve Bayes is allowed.

Bayesian reasoning is applied to decision making and inferential statistics that deals with probability inference. It is used the knowledge of prior events to predict future events.

The Bayes Theorem:

$$\mathbf{P(h/D) = \frac{P(D/h)\,P(h)}{P(D)}}$$

P(h): Prior probability of hypothesis h

P(D): Prior probability of training data D

P(h/D): Probability of h given D

P(D/h): Probability of D given h

Naive Bayes for Machine Learning:

Naïve Bayes algorithm is a powerful algorithm for predictive modeling. The reason behind the name "naïve" is that it uses simplified calculations that make it tractable. Rather than attempting to calculate the values of each attribute value P (d1, d2, d3|h), they are assumed to be conditionally independent given the target value and calculated as P(d1|h) * P(d2|h) and so on. Although this assumption is very strong one, but it probably unlikely to happen in real data, as most of data interact. Nevertheless, the approach performs surprisingly well on data where this assumption does not hold. This model is characterized by fast training; because only the probability of each class and the probability of each class given different input (x) values need to be calculated, so there is no need to calculate coefficients to optimize results.

Predictions for new input data can be carried out using Bayes theorem. For example, if we have the two constraints: whether and going for a picnic and we had a given that the weather is sunny, so we can calculate:

go-out = P(weather=sunny|class=go picnic) * P(class=go picnic)
stay-home = P(weather=sunny|class=stay-home) * P(class=stay-home)

The values output can then be changed to probabilities by normalizing:

P(go-out|weather=sunny) = go picnic / (go picnic + stay-home)
P(stay-home|weather=sunny) = stay-home / (go picnic + stay-home)

In case of having more input variables the equation is extended to predict the outcome with interference of more variables in our experiment. For example, if an extra attribute introduced which is CAR that could be either working or broken, in this case the calculation of class "go picnic", if the CAR is "working" will be:

go-picnic = P (weather=sunny| class=go-out) * P (CAR=working |class=go-out) * P(class=go-out)

To use the predictions then computation of the mean and standard deviation for input values [X] for each class is done, as well as probabilities for each class. In this case we have the selected attributes for the speech signal as inputs [X], and the classes are "Acceptable" if the patient speech analysis attributes indicate that he can undergo rehabilitative training, or "Un-Acceptable", if the patient case is deteriorated to an extent that he cannot improve by training.

This is as simple as calculating the mean and standard deviation values of each input variable (x) for each class value.

mean(x) = 1/n * sum(x)

Where n is the number of instances and x are the values for an input variable in your training data.

We can calculate the standard deviation using the following equation:

standard deviation(x) = sqrt (1/n * sum(xi-mean(x)^2))

For new input values (x) Gaussian Probability Density Function is used to calculate their probability, where the input x is standardized using its mean and standard deviation by computing this equation:

pdf (x, mean, SD) = (1 / (sqrt (2 * PI) * SD)) * exp(-((x-mean^2)/(2*sd^2)))

Then we insert probabilities into the equation to predict with real values, in this study's example the formula will be – for instance if we are to consider acceptable:

Acceptable = P(pdf(Attribute1) |class=Acceptable) * P(pdf(Attribute2) |class=Acceptable) * P(class=Acceptable)

<u>Comments on the algorithm:</u>

This Algorithm requires to prepare the data in certain manner.

- Inputs should be categorized and attributes should be labeled either binary, categorical or nominal.
- For real-valued inputs; assume Gaussian distribution and remove the outliers – if needed – in order to enhance the performance of the algorithm.
- Update the data always if new data are available for training or test, that may optimize the results accuracy.

3.3 Key Nearest Neighbor Algorithm

Key Nearest Neighbor Algorithm (KNN) is a method used for classification and regression. This method lies under the category of supervised ML where the input consists of the K-closest training examples in the features space. For KNN classifier, the output is membership of certain class (in this study: "Acceptable" or "Non-Acceptable" class) according to a vote of its neighbors, where the object is said to belong to the most common class among its K neighbors.

KNN is classified as one of the "Lazy Learning" algorithm, that is not because of apparent simplicity, but because it does not apply certain statistical model to classify the training data, nut it memorizes the training dataset instead. This algorithm is also "Non-Parametric"; a non-parametric algorithm means that it does not assume any assumption about the given dataset, neither Gaussian distribution, linearity or any other special type of distribution. This makes this algorithm useful for real-life applications as most of the practical data does not obey the typical theoretical assumptions.

Computing K- Nearest neighbors in KNN algorithm follows some steps:

1- Determine the parameter "K" which indicates the number of nearest neighbors. It is preferable to choose an odd number to avoid voting tie.

2- Calculate the distance between the query instance and all training samples. The used distance is usually the Euclidian distance given by the formula:

$$d = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$

3- The data then is sorted ascendingly and nearest neighbors are determined according to the Kth minimal distance. This sorting process is the primal assessment of the instance relation to one of the classes

4- Gather the category of the nearest neighbors, to which class do they belong (“Acceptable” or “Non-Acceptable”).

5- Determine the majority category of the query according to the majority of neighbors.

4. Evaluation and Results:

To ensure the effectiveness of our proposed algorithm; a need to identify the accuracy emerges. Accuracy parameter as an important criterion to say that our technique is efficient. The way to compute the accuracy is the sum of the True Positive and True Negative over the sum of all false and true positives and negatives, simply this equation is applied:

$$Accuracy \ = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative}$$

Concerning the Naïve Bayes; using the C-partitioning Method resulted in 40% accuracy which is relatively low, while using the random shuffle function resulted an accuracy around 70%.

The KNN algorithm was accompanied by feature selection algorithms to improve the accuracy of the built system, those feature selection methods resulted the following:

Feature Selection Algorithm	**K Parameter**	**Accuracy**
Using all features (312 feature) (created function)	21	72%
Using 312 + 4 calculated attributes (created function)	21	68%
Built In KNN function	51	69.23%
Removing Low Variance Features.	21	75%
Recursive Feature Elimination	51	69.31%
Principle Components analysis	19	92.31%

5. Discussion:

The previous similar algorithms used for classification of Parkinson's disease utilizing multi-edit nearest-neighbor and ensemble learning algorithms with speech samples resulted in some accuracy. It is perceived it was improved in this study by using feature selection algorithms and increasing the K parameter in KNN algorithm as well as proper calculation of best features and using the randomizing functions.

Previous studies using Naïve Bayes produced 80% accuracy, which was not met in the used algorithm that resulted best accuracy of 70%. On the contrary, the used KNN algorithms improved the previous accuracy which was recorded as 77.5% with K=7, while using Principle Component analysis method for feature selection and K=21 produced high accuracy of 92.31%.

6. Conclusion and Future Work:

This study continues to prove the ability of machine learning and speech recognition methods to provide a proper reliable method to diagnose Parkinson and recommend patient for rehabilitation before reaching late stages that cause physical Disabilities. The Statistical methods of Naïve Bayes Theory and the Key Nearest Neighbor classifier can provide acceptable accuracy.

While for the future work we recommend the SVM (Support Vector Machine) classifier to be applied for the dataset in order to result in much improved accuracy which shall give us more confidence in our technique.

7. References

Introduction to Machine Learning - Nils J. Nilsson Robotics Laboratory Department of Computer Science-Stanford University -Stanford, CA 94305e-mail: nilsson@cs.stanford.edu /November 3, 1998.

J Neural Neurosurgery Psychiatry: first published as 10.1136/jnnp.2007.131045 on 14 March 2008. Downloaded from http://jnnp.bmj.com/ on 16 December 2018 by guest. Protected by

https://parkinsonsnewstoday.com/parkinsons-disease-statistics/

https://web.stanford.edu/group/sjph/cgi-bin/sjphsite/the-parkinsons-voice-initiative-early-diagnosis-for-parkinsons-disease-through-speech-recognition/

https://machinelearningmastery.com/naive-bayes-for-machine-learning/

DATA MINING Concepts and Techniques, Jiawei Han, Micheline Kamber Morgan Kaufman Publishers, 2003

https://saravananthirumuruganathan.wordpress.com/2010/05/17/a-detailed-introduction-to-k-nearest-neighbor-knn-algorithm/

Introduction to k Nearest Neighbor Classification and Condensed Nearest Neighbor Data Reduction -Oliver Sutton -February, 2012.