Voice based gender classification using machine learning

IOP Conf. Series: Materials Science and Engineering **263** (2017) 042083 doi:10.1088/1757-899X/263/4/042083

Voice based gender classification using machine learning

A Raahul, R Sapthagiri, K Pankaj and V Vijayarajan
School of Computer Science and Engineering, VIT University, Vellore-632014, India

E-mail: vijayarajan.v@vit.ac.in

Abstract. Gender identification is one of the major problem speech analysis today. Tracing the gender from acoustic data i.e., pitch, median, frequency etc. Machine learning gives promising results for classification problem in all the research domains. There are several performance metrics to evaluate algorithms of an area. Our Comparative model algorithm for evaluating 5 different machine learning algorithms based on eight different metrics in gender classification from acoustic data. Agenda is to identify gender, with five different algorithms: Linear Discriminant Analysis (LDA), K-Nearest Neighbour (KNN), Classification and Regression Trees (CART), Random Forest (RF), and Support Vector Machine (SVM) on basis of eight different metrics. The main parameter in evaluating any algorithms is its performance. Misclassification rate must be less in classification problems, which says that the accuracy rate must be high. Location and gender of the person have become very crucial in economic markets in the form of AdSense. Here with this comparative model algorithm, we are trying to assess the different ML algorithms and find the best fit for gender classification of acoustic data.

1. Introduction

Dimorphism is the property of voice that is highly observed in human beings. Intonation, speech rate, and duration are certain characteristics that distinguish human voices, mainly male and female voices [1]. The perceived dimorphism accounts for 98.8% which consists of the gender of the speaker and the respective frequencies. Variation in gender, however, cannot be predicted by vocal speech. Some voice pitch may vary between male and female so it is difficult to predict male and female accurately.

With the help of R language, we can identify the gender of the respective speaker with the help of techniques used for speech processing in real time environment. Vocal fold thickness is the main reason behind the difference that can be calculated between the genders. Another reason that contributes to is the style in which the person speaks and the present physical situations. These anomalies are being exploited in such a way that we can classify a speaker as female or male. Previous work related to this has examined differences between male and female voices. These include different parameters. Studies show that main parameter for doing speech analysis is based on frequency and pitch which results in recognition and classification. Speech recognition helps in extracting the information about the gender, their age and the dialect in which they speak. A large amount of work has been done in this field. Certain speech statistics are being applied which uses over time span, mean and maximum value for the detection of gender.

The voice data set is transformed into different parameters like vocal strength, pitch, frequency, q21, q25, etc., these are then trained and tested with different algorithms to predict the gender based on the algorithms mentioned above.

In this paper, we propose a comparative model algorithm which classifies gender based on various prediction algorithms. Prediction is done basically on the dataset whose values are processed from speech. Results obtained are compared with previous research results and it is evaluated with other algorithms to conclude which algorithm results in better performance of classifying gender based on certain parameters. Prediction of how accurately this comparative model algorithm detects the gender based on these algorithms is also found.

2. Literature Survey

Speech classification and processing and gender based recognition and classification have been used for long period of time. We used some concepts developed over time to implement gender recognition. Recent studies based on gender detection shows that voice is converted first into different parameters based on different parameters. Main parameters include pitch and frequency. Classification is done to differentiate male, female and children. For this system is first trained with the training data and test data are introduced and evaluated for the performance of the system for these data. The results obtained are different for different algorithms and seems to produce different results at different times. Gender based classification using F0 frequency [1] and pitch with various training and test data shows that Logistic Regression is best and its accuracy is 92% when compared to other algorithms like regression, Random Forest, Ada Boost and for voice data who speak the same language works better in Random Forest and it has accuracy of 93%. This shows Random Forest suits well for speech recognition based on F0 frequency and pitch to classify male, female and a child [2]. Additional tuning this based on a binning technique to improve the efficiency of the results produced.

Voice based word extraction lab view [3] shows that algorithm works better for this classification and it results better to extract the vowels in male samples. As the samples are trained and tested it efficiently produces a result. It is also observed that by increasing the unvoiced part in speech similar to the sound of 's' value of pitch increases hampering the gender detection in case of samples of the male. Similarly, by increasing the voice part of speech like 'a', decreases the value of the pitch. It fails to identify it when the speaker speaks two different tones.

Speech recognition in adult shows that they are capable of spontaneous and vocal length adjustments and they can sound like masculine and feminine. So, it is difficult to classify the male and female.

Some Female voices are hard to analyze based on the pitch [4] such as examination of one aspect of female voices does not fulfill our requirements.This paper pitch between male and female [4] shows that female voice has to be identified with a different method of parameters than male like High-pitch, Shrill, Emotional and Swoopy. These are different parameters where female users and it varies by a female to female hence, data set needs to be processed based on this before classification of male and female.

According to the perception of pitch [5]. Fundamental frequency [f0] has a combination of linguistic, paralinguistic and nonlinguistic information of the speaker and these three corresponds to male and female and it also depends on high pitch and tone of the speaker. This managed to place a frequency [f0] without any experience with range and no syllable-external information. This shows that the voice of the speaker varies between high and low pitches between speakers.

Identification of gender by SVM [8] shows that speech of gender is analyzed by various different speech mechanisms like compressed speech, speaking of telephone and difference in languages and so on. It conveys that male voice from pitch, period and Mel-frequency is in the range of about 100-146Hz and in the female of about 188-221Hz. Here, the voice is separated based on the frequency and it is extracted and it is analyzed.

GMM based classification of gender [9] propose that speech is analyzed by means of age, words, etc., a Mixture model is used and it identifies up to 98% accuracy and it is one of an efficient method for speech detection and gender analysis. Classification is based on combined parameters of pitch and relative spectral perceptual linear predictive (Rasta-Plp) coefficient to model male and female speech.

3. Algorithm

3.1 Linear Discriminant Analysis (LDA)

LDA use some datasets a mixture of topics that can choose words with probabilities. It concludes that documents are composed in the following manner: when writing each topic, you think on the number of words N the data set will have. We are taking a dataset which is a combination of the topic. Considering this type of model for a mixture of data set, LDA backtrack from the dataset to collect the set of topics that are likely to have generated the group of data.

This model is highly flexible and can be easily elongated. The main area of importance is modeling relations between topics. Hierarchical LDA wherein hierarchy fashions topics are grouped by using the nested Chinese restaurant process. LDA also be extended to the corpus, which involves two types of variables like words and names, as in the LDA-dual model. Nonparametric extensions of LDA include the hierarchical Dirichlet process mixture model, which confess the number of topics to be unbounded and learned from data and in nested Chinese restaurant process, topics to be arranged in a hierarchy whose design is learned from data.

3.2 K-Nearest Neighbour

K-Nearest Neighbour (KNN) is also a machine learning approach. KNN is easy because of its easiness and flexibleness to process various data variables. The objectivity is to estimate on a stable no. of observation, let’s say k that is nearer to the required outcome. It doesn't build any replica or task, but still, it predicts nearest k records from the data set that is including the most identity to the test. KNN is mention as lazy learning procedure because of this nature. It can be implemented on both continuous and discrete known as regression and classification respectively. Regression carries out the k neighbor average, whereas classification carries out a frequent neighbor. It is a supervised algorithm, training data n pair (xi ,yi) and y(x) is to find out the problem from a new input x. To implement the technique, it is considered to have a training set and a test sample. In order to find the value of k, and the formula for the distance between the instances. It can be given by, 1. Euclidean Distance, 2.Manhattan Distance, 3.Minkowski Distance.

3.2.1 Algorithm

for all test example x do

for all training example (x_i,y_i), do

computer distance(x,x_i);

end for

select the k-nearest neighbour of x;

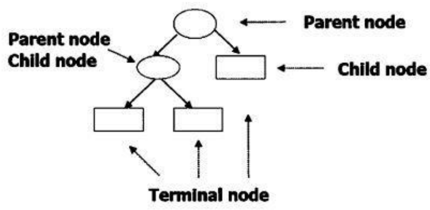
return the average output value among neighbours i.e. $1/k\ \boldsymbol{\Sigma}_i^k{=}_1 y_i$;

end for;

3.3. Classification and Regression Trees (CART)

Cart refers to Classification and Regression Trees, it is a classical stats and machine learning algorithm initiated by Leo, Jerome, Richard and Charles in 1984. It is used to present complex data in a decision tree manner. The topmost junction is a Root node, and this is split into child nodes, and again child nodes are treated as parent nodes for splitting nature of the nodes prior to stopping criteria happen. And then each node is considered as a terminal node and those nodes were assigned with a dataset. Normally, Classification results out the categorical whereas regression is a continuous outcome.

IOP Conf. Series: Materials Science and Engineering **263** (2017) 042083 doi:10.1088/1757-899X/263/4/042083



3.4. Random Forest

Tin KamHo(Bell Labs) in 1995 proposed the concept of a random forest. It is an ensemble classifier consisting many decision trees, further providing the output by means of class's output by individual trees outputs of the class. The method couples the bagging idea and the random selection. Each tree is build up using a variant bootstrap data samples. In addition to this random forest changes it also constructs a classification of regression trees. In case of standard trees, the best node is split using its variables. Moreover, each node is divided using the best subset of predictors which are randomly chosen at that node. This method seems to be robust and stable among the other algorithms including discriminate analysis, support vector machines and neural networks and is robust against over fitting.

3.5. Support Vector Machine (SVM)

Supervised learning is the machine learning task for deducing functions from a labeled training data that can be occupied for both classification and regression. Support vector machines are a binary classification algorithm. Support vectors are the data points adjacent to the hyperplanes if the dataset is removed it will change the position of the dividing hyperplane. In SVM, each example set is a pair having an input object and a preferred output value. Supervised learning algorithm analyses the data and result out the inferred function, which result in mapping new outcomes.

We can say, SVM is a machine learning tool used for classification, approximation, etc.., it can be used in a generalized manner that leads to the success in many fields. Metrics of SVM is to minimize and generalization of upper bound error upon maximizing the margin which is separating hyperplane and data set. Advantages of model selection by means of both optimal number as well as the location of functions are obtained automatically during training.

4. Dataset

Dataset used here is voice gender dataset and it has 3169 records out of which 1584 Male records and 1985 Female and for convenience labels of male and female are converted into 0 and 1. Each record has different acoustic properties like mean freq, sd, median, etc.

5. Result

5.1 Scatter plot

For a set of data variables (X1, X2, ... ,Xk,) the scatter plot matrix shows all the pairwise scatter plots of the variables on a single view with multiple scatter plots in a matrix format. It suggests different correlations between variables which may be positive, negative or null. The function of scatter plot is a plot(x, y) and this function is available under gg plot or plot libraries. In this, it is used to plot values for classification of gender and it shows that Support Vector Machine performs better for this voice gender dataset. The scatter plot is shown in Figure 1.

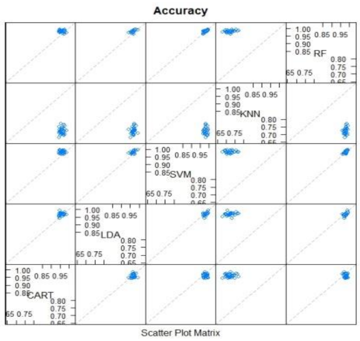
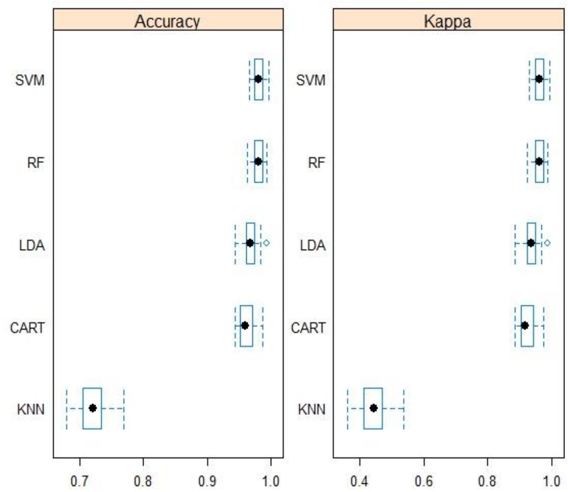


Figure 1 Scatter plot

5.2. R-plot

It is a generic function used to plot XY values. Various types like the chart, graph, line chart, a bar graph can be produced by using this rplot and it is available under ggplot library and the function used is plot(x, t), line(x, y). It is used to compare metrics based on input supplied. In this accuracy of gender prediction is done between CART and SVM and it results that SVMs accuracy is more than CART. The R-Plot is shown in Figure 2 and Figure 2.1.



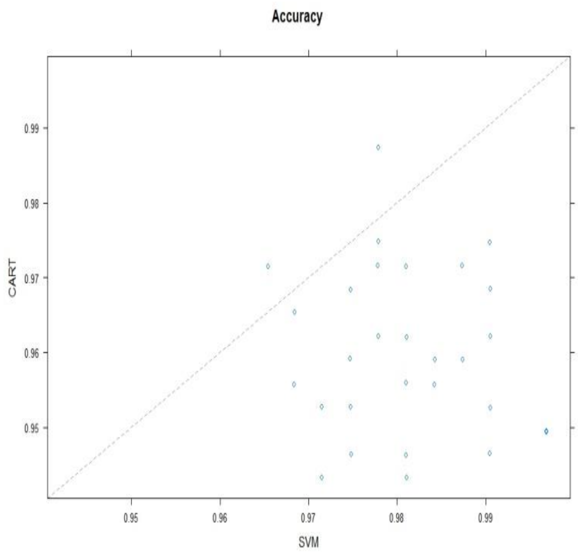


Figure 2 R plot **Figure 2.1** R plot

5.3. Parallel plot

This is similar to Rplot but it compares several algorithms at a time for a particular metrics. The function for the parallel plot is paracord(x, col, lty, var.label) or parallel.plot(tmatch, outcome)and it is imported from ggplot and lattice libraries.Here it compares the accuracy and shows that accuracy at SVM is better than others. Accuracy at svm point is same and tends to predict accurately. The parallel plot is shown in Figure 3.

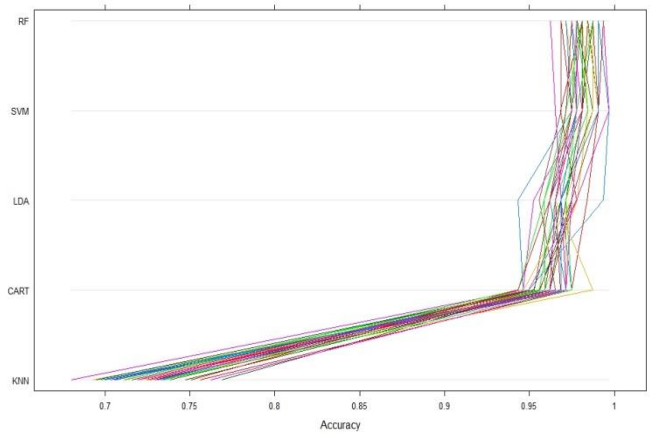


Figure 3 Parallel plot

5.4. Pairwise plot

The pairwise plot is the process of comparing entities in pairs to judge which entity has greater effects. It is also similar to scatter plot where the points are scattered along the graph and the accuracy is calculated based on the points lying above and below the boundary lines. The function of the pairwise plot is paired(cols, values) and it is available under lattice and ggplot library. In this SVM is compared with KNN, RF, LDA and the accuracy is predicted at different levels and thus shows SVM has greater effects on this dataset of voice gender. The pairwise plot RF and SVM is shown in Figure 4.1 and accuracy is shown in Figure 4.2. The same plot is shown for KNN-SVM is shown in Figure 4.3.

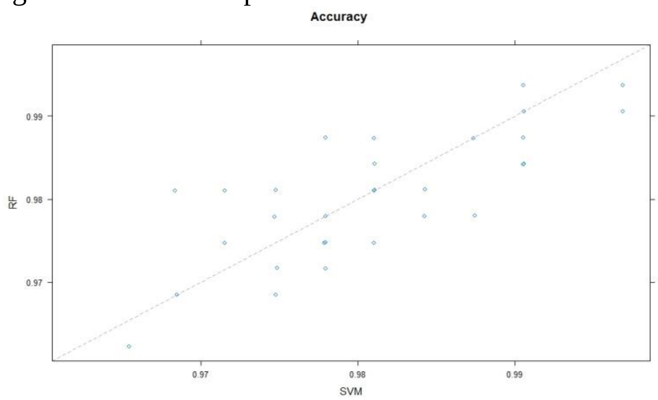


Figure 4.1 Pairwise plot RF and SVM.

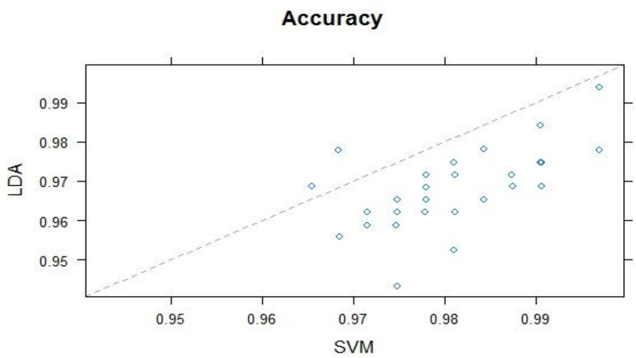


Figure 4.2 Pairwise plot LDA and SVM.

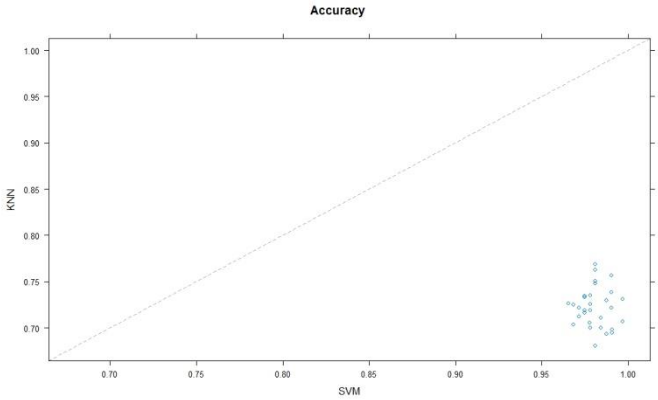


Figure 4.3 Pairwise plot KNN and SVM

5.5. Dot plot

Dot plot is a statistical simple plot suitable to small and medium sized datas. It is used for highlighting the values correctly and more precisely. It is a chart where data points are plotted on the surface and used for comparison. It is available under ggplot and the function for dotplot is dotplot(1) or dotplot(2). Here it is used fort comparing the accuracy and kappa for five different algorithms and it shows that SVMs accuracy and kappa is higher than others for this voice gender dataset. The dot plot is shown in Figure 5.

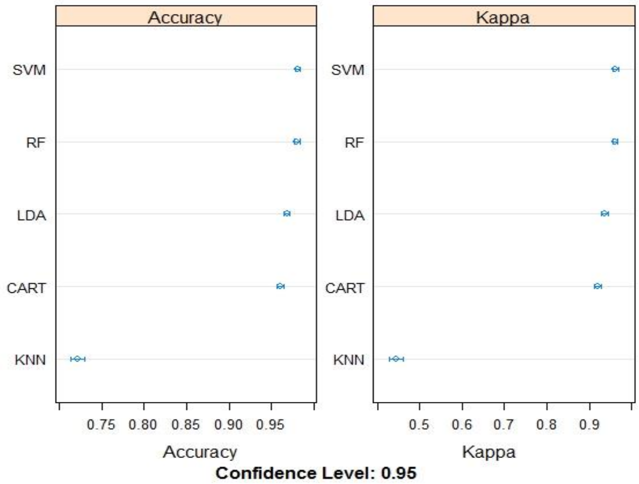


Figure 5 Dot plot

5.6. Density plot

Density plot constructs and graphs nonparametric density estimates possibly conditioned on a factor. It is a basic function available under ggplot and lattice libraries. The density plot is a much more effective way to view the distribution of a variable. Function for plotting density plot is a plot(density(x)) and it is used to create the histogram similarities. Here it is used for detecting accuracy and kappa and it shows high variation change. The density plot is shown in Figure 6.

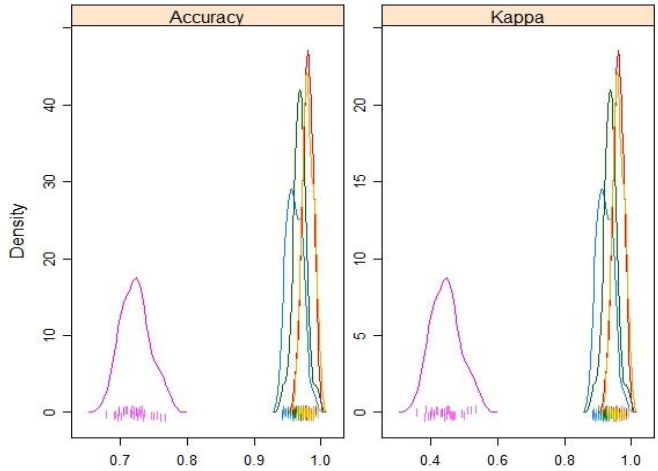


Figure 6 Density plot

5.7. Box plot

Box plot can be created for both individual and groups of the variable. The format for using boxplot is boxplot(x, data=), where data is providing a frame for data. It is imported from ggplot library and it shows the accuracy with the box style. Here it is used for evaluating accuracy and kappa where it results better for SVM over other algorithms. The box plot is shown in Figure 7.

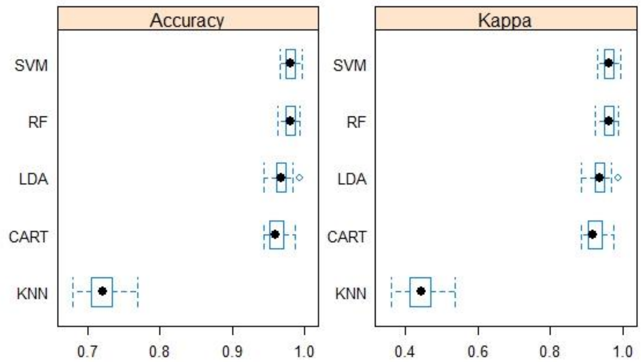


Figure 7 Box plot

6. Conclusion

By Comparative algorithm of comparing the above mentioned five different algorithms with eight parameters like plots and tests namely Box, Density, Parallel, Dot, Pair, and tests like Statistical significance. The results obtained shows that SVM algorithm performs better in classification and with reduced error rate. These results obtained using this comparative algorithm are only for this voice gender dataset and it may vary for another dataset. SVM tends to have more accuracy over another algorithm in classifying gender in spite of variations in pitch and frequency. Future work to add more algorithms to this Comparative model and to compare the performance with this work and to identify which algorithm in linear and Non-Linear performs better in the classification of gender in voice gender dataset.

References

[1] Ericsdotter C and Ericsson A M 2001 Gender differences in vowel duration in read Swedish: Preliminary results Working Papers - Lund University Department of Linguistics 34 – 37

IOP Conf. Series: Materials Science and Engineering **263** (2017) 042083 doi:10.1088/1757-899X/263/4/042083

[2] Whiteside S P 1996 Temporal-Based Acoustic-Phonetic Patterns in Read Speech: Some Evidence for Speaker Sex Differences, *J International Phonetic Association* **26** 23–40

[3] Byrd D 1992 Preliminary results on speaker-dependent variation in the TIMIT database *J AcoustSoc Am* **92** 593–596

[4] Henton C G 1989 Fact and fiction in the description of male and female pitch *Language and Communication* **9** 299-311

[5] Bishop J and Keating P 2012 Perception of pitch location within a speaker's range: Fundamental Frequency, voice quality and speaker sex *The Journal of the Acoustical Society of America* **32-2** 1100-1112

[6] Smith D R and Patterson R D 2005 The interaction of glottal-pulse rate and vocal-tract length in judgments of speaker size, sex, and age in *The Journal of the Acoustical Society of America,* **118-5** 3177-3186

[7] Muhammad G, AlSulaiman M, Mahmood A and Ali Z 2011 Automatic voice disorder classification using vowel formants, 2011 *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '11)* 1–6

[8] Gaikwad S, Gawali B, and Mehrotra S C 2012 Gender identification using SVM with combination of MFCC, Advances in Computational Research **4** 69–73

[9] Zeng Y M, Wu Z Y, Falk T and Chan W Y 2006 Robust GMM based gender classification using pitch and RASTA-PLP parameters of speech *Proceedings of the International Conference on Machine Learning and Cybernetics* 3376–3379