LOVELY PROFESSIONAL UNIVERSITY

Lovely Professional University,
Jalandhar-Delhi, G.T. Road, Phagwara, Punjab

Report On Gender Detection Using Voice Frequency Using Deep Learning

Department of Computer science and Engineering
NAME:SOMESHWAR SINGH,GAURAV SHARMA
REG NO.: 11806479,11806823
SECTION:KM073

Abstract

Gender classification is one of the major problems in field of speech analysis now-a-days. Identification of gender from acoustic properties of frequency i.e., mean, median, frequency etc. is the highly important. Machine learning is used to solve this problem because it gives promising results for classification techniques. There are several algorithms that can be used to predict the gender using acoustic properties. In my project, I am evaluating classifiers using 6 different machine learning algorithms. These algorithms include K-Nearest Neighbor (KNN), Decision Tree (DT), Random Forest (RF), and Support Vector Machine (SVM), Support Vector Machine Using Poly kernel and Gradient Boosting (GB). The main parameter involved is the accuracy obtained using all these classifiers. I am trying to assess the accuracy, recall, precision, and F1 Score obtained after predicting on test data for all these classifiers and finally the best fit model will be generated for gender classification of acoustic data.

INTRODUCTION

Speech signal is the most common means of communication for human beings. 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. The task to identify a human's gender by voice is very easy when a human identifies it. On the other hand, it becomes difficult for a computer to identify whether the voice is of male or female. A human has instinct to identify the difference between the voice of male and female but when it comes to computer we need to train and make it learn by providing training data and various algorithms.

The aim of this project is to create a classifier which identifies the gender based on acoustic attributes of voice. The data set split into train and test data and model is build using train data. This model is trained using various machine learning algorithms which includes K-Nearest Neighbor (KNN), Decision Tree (DT), Random Forest (RF), and Support Vector Machine (SVM), Support Vector machine Using Poly kernel and Gradient Boosting (GB). The confusion matrices are created and results of different models are compared with each other. The model with the highest accuracy is considered as the best model. There are various applications where gender recognition can be useful. Some of these include:

For further identification of human sounds like male laughing, female singing

Categorizing audios/videos by adding tags and simplifying and reducing search space

Automatic salutations

Can help personal assistants like Siri, Google Assistant to answer the question

With female generic or male generic results.

The general block diagram of gender classification model using audio signal is given in as follows.

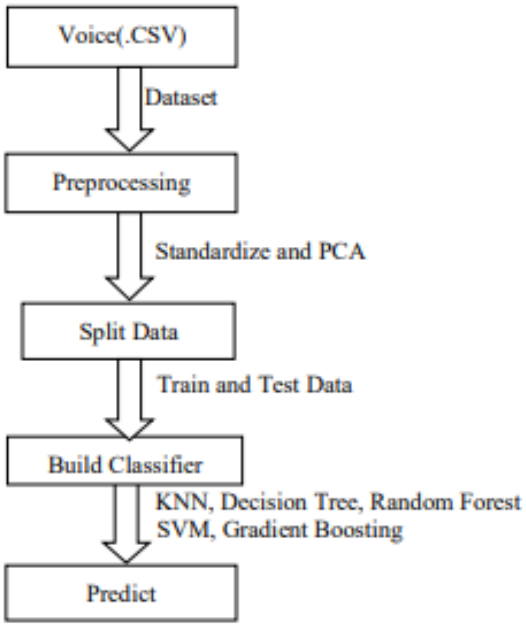

Fig. 1. Work flow to recognize gender using voice

Preprocessing Techniques

The dataset is already preprocessed. But still preprocessing techniques are performed to achieve the maximum accuracy. The preprocessing techniques used in my project are as follows:

Missing values

Handling missing values is first and the prior step towards building the model. Missing values occur in data due to various reasons, such as problems occurred during extraction or data collection process. So, I checked for the missing values in the dataset. There are no missing values found. If there were any missing values, I would have deleted the the missing values using list wise method. In list wise method, the complete row which contains missing value is erased. After checking for the missing value, the next and most important step is to encode the label column in the dataset. Label column contains the output of the dataset which can

either be male or female. The output male is encoded as 1 and the output female is encoded as 0. The reason for this encoding is that , further a classification model is to built up and it will be easy if the output column contains two classes 1 or 0 instead of male or female.

Standardization

Standardization is a technique used to remove mean or to scale the variance. If the individual features of the data do not look like standard normally distributed data i.e. zero mean and unit variance then fitting the model will not be accurate. It is always required to create the best fit model using machine learning algorithms. Standardization is basically subtracting the mean value of each feature, then scaling it by dividing the respective feature by its standard deviation. I am performing this technique to centralize the data and it is implemented using sklearn library

Principal Component Analysis

Principal component analysis (PCA) is a technique to reduce the learning space as it reduces many correlated variables into a smaller number of uncorrelated variables. It decreases the dimensions of dataset. The aim of this technique is to reduce those features which are uncorrelated with the output or highly correlated with other features. From those highly correlated features, one feature is selected. In my dataset, at first I performed PCA for reducing 20 features. As, the dataset used is already preprocessed, so performing this technique does not prove to be much efficient.

Classification Techniques:

After performing the prep reprocessing techniques, the dataset is ready to be processed. For performing classification, the dataset is divided into two parts. 75% of dataset is in train data and 25% of dataset is in test data. There should always be more data in training data as compared to the test data set. It is so because more data is required to train the model as compared to test the model. Both training and testing data are further categorized into two parts. One is X which includes all the independent predictors, and the other is Y which includes the output i.e., the label column which tells whether the voice is of male or female for the corresponding values of X. The following are the types of classifiers:

K Nearest Neighbors Classifier

K Nearest Neighbors (KNN) is the intuitive machine learning algorithm. KNN. It is an unsupervised technique which predicts the output while searching for the nearest and most similar K records from the dataset. To implement this technique, K Neighbors Classifier from sklearn library is used. The fitted model i.e. classifier is build using the training set (including both X and Y). Prediction is done on this model for the test sample. Finally, the predicted output is compared with the actual output and the accuracy, recall, precision and F1 score are calculated.

Decision Tree Classifier

Decision Tree (DT) is a non-parametric supervised learning technique used for classification. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. A condition is checked at the root node and based on the output an appropriate branch is selected. It can further lead to another internal node where a new test condition is checked, or to a leaf node where class is assigned to the record [21]. This classifier is applied with the tree function in sklearn library. The model is trained with the help of X and Y values of train dataset. Then this model is predicted on X values of test dataset and the corresponding Y values are obtained. These Y values are compared with actual Y values e the accuracy, recall, precision and F1 score are recorded.

Random Forest Classifier

Random forest (RF) is an robust ensemble classifier which consists a number of decision trees. The output is obtained by the collective decision of output from individual trees. The root node is split using features in the dataset and each node is further divided using other features. I implemented it using Random Forest Classifier in sklearn library. The model is fitted using the X and Y of train dataset. Prediction of this model is done on X of test dataset and the output is predicted. This output belongs to class 0 or 1 which means the voice belongs to either male or female. The output is compared with Y of testing dataset and the accuracy, recall, precision and F1 score are obtained.

Support Vector Machine Classifier

Support vector machine (SVM) is a binary classifier which separates the data into two different classes at the same time. The decision boundary is built such that the distance between hyperplane and the training data is maximum. This is done to minimize the noise because if hyper plane

passes close to the points, it will be noise sensitive. After determining the best decision boundary, it is easy to determine the class of test data by checking the side of the hyperplane it belongs [20]. I used SVM because there are 2 classes 1 and 0 i.e., male and female. Two parameters involved in SVM are regularization parameter or C parameter that talks about the SVM optimization and the gamma parameter. To implement it, SVM from sklearn library is used. SVC class along with radial basis function default kernel , value of C as 10 and value of gamma as 0.1 is used to fit the model from train data. The fitted model is predicted on the test data and the output is obtained. This output is compared with actual output and the accuracy, recall, precision and F1 score are calculated.

Support Vector Machine Classifier using poly kernel

In Support vector Machine the default kernel used is radial basis function. But there are many other kernels that can also be used which include poly and sigmoid. This technique is used to find the better kernel than radial basis function and to implement it. At first, I performed cross validation technique to find the scores associated with each kernel. The kernel with highest score is better and chosen. For creating a better model, C and gamma values associated with kernel should also be calculated. For calculating these values, again cross validation techniques are performed, and accurate value of C and gamma is chosen. Using the kernel with high score and its associated C and gamma values, the classifier is created. The model is trained using this classifier along with the X and Y values of train data. Then this model is tested on X values of test data and corresponding Y values are obtained.

Gradient Boosting Classifier

Gradient boosting is a machine learning technique which produces a classifier in the form of an ensemble of weak prediction models. It builds an additive model in a forward stage-wise fashion and is generally used when individual classifiers do not provide accuracy. It is implemented using Gradient Boosting in ensemble-based methods from sklearn library.

RESULTS AND DISCUSSIONS

Classifiers which include K Nearest Neighbors, Decision Tree, Random Forest, Support Vector Machine, Support Vector Machine using poly kernel and Gradient Boosting are compared with each other based on Accuracy, Precision, Recall and F score. Accuracy is the fraction of predictions that are predicted correctly. Precision is the fraction of relevant

instances among the retrieved instances and is also called positive predictive value. It is the ratio of correctly predicted positive observations to the total predicted positive observations. Recall is the fraction of relevant instances that have been retrieved over the total amount of relevant instances and is also called sensitivity [22]. It is the ratio of correctly predicted positive observations to the all observations in actual class - yes. F1 Score is the weighted average of Precision and Recall [23]. To calculate accuracy, precision, recall and F1 score, I am using metrics function in sklearn library.

CONCLUSION

For the dataset I used, Support Vector Machine can be considered as best classifier for recognizing gender using voice as it provides best result. Support Vector Machine gives accuracy of 98.48%, precision of 98.71%, recall of 98.21% and F1 Score of 98.46%. In this case, I applied Principal Component analysis, but the results were not satisfactory. While creating the model, all the classifiers use different features. Therefore, while creating the best fit model, feature selection plays an important role. No classifier can always act as best classifier. It depends on the problem and the solution to it.

REFERENCES

A Raahul, R Sapthagiri, K Pankaj and V Vijayarajan, “Frequency based gender classification using deep learning”, IOP Conf. Series of Materials Science and Engineering, 2017.

K. Becker, “Identifying the Gender of a frequency using Machine Learning”, 2016, unpublished.

J. M. Hilbe, “Logistic Regression Models”, CRC Press, 2009.

L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, “Classification and Regression Trees”, CRC Press, 1984.