Analysing Features Effects on Classifier on EDHV (Emotion Detection from Human Voice)

PHYS4038 Scientific Programming in Python / Final Report

Alpaslan Erdag
20405820

ABSTRACT

Here, it is discussed the effect of feature extraction methods and feature extraction methods parameter selections on accuracy performance, creating training data and test data and making measurements by giving these data as input to the classifiers. In this study, RAVDESS data set is used. KNN, SVM and MLP classifiers were used to make these measurements. Classification results were evaluated using the accuracy metric. The purpose of this application is to understand effects of the features to the accuracy metric.

1. Introduction

Various methods were used in this project. The theoretical explanations and code fragments of mfcc, zcr (zero crossing rate), Chroma, MEL, STFT feature extraction, normalization methods used in this study are given below.

Feature Scaling / Data Normalization:

Feature scaling is a method used to normalize the range of independent variables or features of data. In data processing, it is also known as data normalization and is generally performed during the data pre-processing step. Standardization refers to shifting the distribution of each attribute to have a mean of zero and a standard deviation of one (unit variance). [1]

$$z = \frac{x_i - \mu}{\sigma}$$

Equation 1: Standard Scaling

```
X_train = (X_train - mean)/std
X_test = (X_test - mean)/std
```

Code Representation

MFCC Feature

MFCC features return basic audio features to synthesis vectors and are the most commonly used spectral features. This set of coefficients is obtained by calculating the Fourier transform of the logarithmic power spectrum of the signal. In this study, 20 feature vectors were used.

```
mfcc_f = librosa.feature.mfcc(y=data, sr=sr, n_mfcc=20)
```

Code Representation of MFCC Feature Extraction

ZCR Feature: Zero crossing rate is the rate of sign change across a signal, that is, the rate at which the signal changes from positive to negative or backward. This feature has been used extensively in both speech recognition and musical information retrieval and is a key feature for classifying percussive sounds.[2]

Chroma Feature: The chroma feature is a descriptor, which represents the tonal content of a musical audio signal in a condensed form. In speech analysis it's shown that this feature is defective in emotion detection. [3]

STFT Feature: The Short-time Fourier transform (STFT), is a Fourier-related transform used to determine the sinusoidal frequency and phase content of local sections of a signal as it changes over time. [4]

K-NN Classifier: The k-nearest neighbours algorithm is a non-parametric method used for classification which is used for same purpose in this study. Input consists of the k(number is defined logically by user to get more accuracy) closest training examples in the feature space. The output depends on the purpose because KNN can be used for regression also.

When you use it for classification, the output refers to a class label. A label(or a data point) is classified by the majority of its neighbours; the data point is assigned the class most common among its nearest neighbours (k is usually a small positive integer). If k=1, then the object is only assigned to that nearest neighbour's class. Another important think about the "k" is that, when you define it it's better to be odd number to avoid equality.

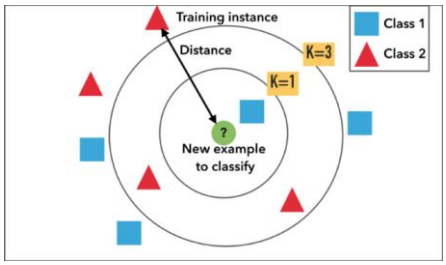

Figure 1: kNN Sample Representation

SVM: In machine learning, support-vector machines are supervised learning models with associated learning algorithms that analyse data for classification and regression analysis.

The advantages of support vector machines are:

- Still effective in cases where number of dimensions is greater than the number of samples.
- Uses a subset of training points in the decision function (called support vectors), so it is also memory efficient.
- Besides performing linear classification, SVMs can efficiently perform nonlinear classification using so-called kernel numbers and implicitly map their inputs to higher-dimensional feature spaces. Common kernels are provided, but it is also possible to specify custom kernels.[5]

The SVM classifier algorithm is designed for two classes. However, methods named one-vs-all, one-vs-one have been developed in order to be able to classify more than 2 classes. While training the one-vs-all method, it creates a model for each class separately. While creating this model, it operates by assigning 1 for the selected class and 0 for all other classes, reducing the multi-class system to 2 classes. The SVM method used here to get the results is one-vs-all. Because we are working with a multi-class dataset.

MLP: A multilayer perceptron is a class of feedforward artificial neural network (ANN). The term MLP term is used for feedforward ANN, sometimes strictly to refer to networks composed of multiple layers of perceptrons (with threshold activation).
Multilayer perceptron has be contain at least three layers. These layer are an input layer,hidden layer and an output layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. MLP is mostly used in supervised learning.[6]

2. Application

Flow diagram of code is given below. After examining the data set, I thought it would be more informative to do a study with 8 and 16 classes. After reading the dataset properly, I extracted two datasets with 8 and 16 classes using the functions of the Pandas library and extracted features from these two datasets separately with the functions of the librosa library. Then, I divided the data into two as test and training data with the train_test_split function of the scikit-learn library. I trained the classifier with the training data and evaluated the results with the test data.

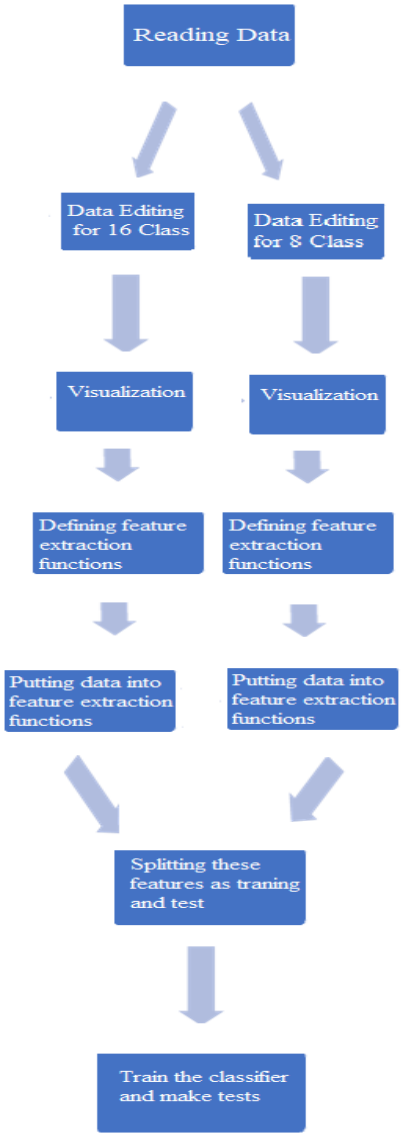

Figure 2. Flow Diagram of Application

3. Results

The accuracy values obtained as a result of each feature extraction are given in the tables below.
For example, for the first-row accuracy values obtained as a result of inputting the values as a result of MFCC feature extraction to the classifier.
The results showed that when the classifiers were trained with only MFCC and only STFT features, a slightly higher accuracy rate was achieved on 16-class data. When trained with the other three feature extraction, the classifiers achieved higher accuracy on the 8-class dataset.

16-CLASS SITUATION

FEATURES	MLP ACCURACY	SVM ACCURACY	KNN ACCURACY
MFCC	%72.57	%53.47	%74.82
Zero-Crossing Rate	%9.1	%6	%30.4
MEL	%33.68	%22.2	%49.65
CHROMA	%19.44	%20.83	%41.23
STFT	%67.36	%70.49	%62.84

8-CLASS SITUATION

FEATURES	MLP CLASSIFIER ACCURACY	SVM ACCURACY	KNN ACCURACY
MFCC	% 65.62	%46.18	%75.43
Zero-Crossing Rate	%16.32	%16.67	%36.1
MEL	%44.44	%34.03	%54.42
CHROMA	%24.31	%20.83	%42.36
STFT	%64.93	%58.33	%63.28

As you can see in the table below, detecting accuracy of gender from the voice is very high. I have not come across a recent study on this gender matter, but I wanted to evaluate the data I have from this point of view. It seems that gender discrimination from speech sound is highly accurate no matter what attribute you choose.

2-CLASS SITUATION

FEATURES	MLP CLASSIFIER ACCURACY	SVM ACCURACY	KNN ACCURACY
MFCC	% 97.92	%97.57	%99,1
Zero-Crossing Rate	%61.11	%61.81	%67.9
MEL	%83.68	%68.75	%83.15
CHROMA	%84.38	%85.76	%86.97
STFT	%98.61	%99,53	%98.43

Combinational Results: After getting these results, I also wanted to know "What if I combine some of these features and train the classifier with new data set?". For this purpose, I took 8 and 16-classes situations into account. Because there is a consistency in terms of "best accuracy values" between these two results.

Combination of MFCC and STFT features resulted as %64.93, %59,01 and %64,06 respectively in MLP, SVM and KNN classifiers.
On the other hand, combination of MEL and MFCC features resulted as %68.06, %44,11 and %73,52 respectively in MLP, SVM and KNN classifiers.

Combination of Chroma and MFCC features resulted as %59.38, %45,49 and %69.44 respectively in MLP, SVM and KNN classifiers.

Combination of MFCC and ZCR features changed accuracies very slightly.

References:

[1]https://medium.com/@fbanovax/feature-scaling-methods-7a19bb9fd7a6
[2] https://en.wikipedia.org/wiki/Zero-crossing_rate
[3]https://www.researchgate.net/publication/330796993_Chroma_Feature_Extraction
[4]https://en.wikipedia.org/wiki/Short-time_Fourier_transform
[5] https://scikit-learn.org/stable/modules/svm.html
[6 https://en.wikipedia.org/wiki/Multilayer_perceptron