Speech Phoneme Analysis and Classification Assignment

Charlton Sammut
0490200L
ICS2203

Index

Section 1: Introduction

Problem Analysis

This assignment is composed of two main parts. The first part requires one to extract a number of features from the ABI-1 Corpus.[1] The features that where extracted are:

- Speaker label
- Gender
- Word
- Vowel Phoneme
- Class number
- Format 1 (Hz)
- Format 2 (Hz)
- Format 3 (Hz)
- Time marker (s)

The second part requires one to build a k-nearest neighbour classifier that determines what phoneme of speech is being analysed based on its format and frequency values.[2]

Section 2: Selected Solution and Design

Files and environment

To successfully extract all 9 features Praat was used.[3] These features where recorded inside of a spreadsheet then converted into a csv file. These features can be found inside of the Data/ABI-1 Corpus features .csv file.

To build the classifier Python 3.7.7 was used. Python was chosen for a multitude of reasons. One of the main reasons is it's simple syntax and transparent semantics make it an excellent choice for any natural language processing task.[4] Another reason is that one can make use of its variety of libraries. For this project the pandas and sklearn libraries were used.[5][6] Before running and executing the assignment, the libraries have to be installed. This can be easily done by running the below commands in the command line.

```
pip install pandas
```

```
pip install numpy
```

```
pip install scipy
```

```
pip install sklearn
```

To make use of the KNN classifier the classifier.py has to be ran. classifier.py takes a total of 3 parameters as follows:

```
python classifier.py k numberOfLoops groupByGender
```

k:

Type: int

Description: Represents the number of closest data points the algorithm to locate.

Default Value: 15

numberOfLoops:

Type: int

Description: Represents the number of times the algorithm should iterate.

Default Value: 10

groupByGender:

Type: Bool

Description: When true compares all the data points with every other data point. If false compares all the data points with every other data point, given that the data points have the same gender.

Default Value: False

Section 3: Evaluation and Possible Improvements

Evaluation

To properly show the performance of the classifier, three results are outputted:

1. Classifier average score
2. Confusion Matrix → Diagonals represents correctly identified cases, while the other values are incorrectly identified cases.[7]
3. Average F1 score → Show's the classifier's accuracy.[8]

A number of tests were conducted to properly test what results the classifier produces under different. All the results for these tests can be found in Results .xlsx.

N.B: If a value is not specified below, then the default value was used.

Score in correlation with different values of numberOfLoops

Regardless of the number of iterations, the average classifier and F1 scores both where unchanged, and shifted between 0.9 to 0.91.

Score in correlation with different values of k

The best results produced where for smaller values of k. As the value for k got bigger, the results started getting worse. This is expected as with larger values of k the clusters generated won't be as clustered when compared to the clusters generated by smaller values of k.

Score in correlation with different values of groupByGender

Both the average classifier score and average F1 score are smaller when the classification is done on a single gender alone. It should be noted that the female average score tends to be higher (around 5% to 10%) then the male average score.

Score in correlation with different distance metrics

The classifier's distance metric is one such that the weight points are the inverse of their distance. To put it differently, closer neighbours of a query point will have a greater influence than neighbours which are further away.[9] However another distance metric was also used for testing purposes. In this different metric, all the points in each neighbourhood are weighted equally.[9] This however consistently produced worse results, and hence the first distance metric was chosen.

Words in correlation with number of misclassification

Consistently the word 'hoard' had the most number of misclassification. It's most often classified as the word 'hard'.

Possible Improvements

The biggest improvements that can be applied to this assignment is if it was repeated for different accents and different words, then these new results can be compared to the results evaluated from this assignment. From this comparison one could be able to generalise the results for more accents and words.

Additional features can also be added to this assignments, such as plotting graphs to better illustrate the outputted results.

References

[1] = Thespeechark.com. n.d. The Accents Of The British Isles (ABI-1) Speech Corpus. [online] Available at: <http://www.thespeechark.com/abi-1-page.html> [Accessed 7 April 2020].

[2] = D. Cheng, S. Zhang, Z. Deng, Y. Zhu and M. Zong, Advanced Data Mining and Applications: 10th International Conference. China, 2014, pp. 499-512.

[3] = P. Boersma and D. Weenink, "Praat: doing Phonetics by Computer", Fon.hum.uva.nl. [Online]. Available: http://www.fon.hum.uva.nl/praat/. [Accessed: 07- Apr- 2020]

[4] = D. Kozaczko, 8 best Python Natural Language Processing (NLP) libraries. 2018 [Online]. Available: https://sunscrapers.com/blog/8-best-python-natural-language-processing-nlp-libraries/. [Accessed: 07- Apr- 2020]

[5] = "pandas - Python Data Analysis Library", Pandas.pydata.org. [Online]. Available: https://pandas.pydata.org/. [Accessed: 07- Apr- 2020]

[6] = "scikit-learn: machine learning in Python", Scikit-learn.org. [Online]. Available: https://scikit-learn.org/stable/#. [Accessed: 07- Apr- 2020]

[7] = S. Visa, B. Ramsay, A. Ralescu and E. Knaap, "Confusion Matrix-based Feature Selection", in Proceedings of The 22nd Midwest Artificial Intelligence and Cognitive Science Conference, Cincinnati, Ohio, USA, 2011 [Online]. Available: https://www.researchgate.net/publication/220833270_Confusion_Matrix-based_Feature_Selection. [Accessed: 07- Apr- 2020]

[8] = Y. Sasaki, "The truth of the F-measure", 2007 [Online]. Available: https://www.researchgate.net/publication/268185911_The_truth_of_the_F-measure. [Accessed: 07- Apr- 2020]

[9] = "sklearn.neighbors.KNeighborsClassifier — scikit-learn 0.22.2 documentation", Scikit-learn.org. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html. [Accessed: 07- Apr- 2020]

Other references:

H. Traunmüller and A. Eriksson, "The frequency range of the voice fundamental in the speech of male and female adults", 1995 [Online]. Available: https://www2.ling.su.se/staff/hartmut/f0_m&f.pdf. [Accessed: 07- Apr- 2020]