Matthias Vassallo Pulis

ICS2203/ARI2203 – Vowel Analysis and Classification Report

Link to ABI-1 Corpus (as it is too large to fit into submission ZIP file)

https://drive.google.com/file/d/18nQPfL1uU6qThxy7t3rukJlarBFkUN9L/view?usp=drive_link

Feature Extraction and Data Collection

The dataset that is used is a part of the ABI-1 Corpus, which contains 14 different regional accents of the British Isles, whereby each accents includes voices several male and female speakers. Each speaker says out the following words: "heed", "hid", "head", "had", "hard", "hudd", "hod", "heard", "hoard", "hood", "who'd", "hid", "hoid", "hoed", "howd", "heered", "hared", "hured", "heed".

Throughout the feature extraction process, 3 different vowel sounds from 5 male and 5 female speakers from 5 different accents were extracted. The words used for this process are "had", "hood" and "hid". These words were chosen in order to have 3 different phonemes, which in this case are AE, UH and IH respectively. The values for Formant 1, Formant 2 and Formant 3 for these phonemes were retrieved using Praat. This was because the process to extract the formant values is much easier in this program compared to Ocenaudio. The speaker label, gender, word, and vowel phoneme were also entered. All the data was written into a CSV file with the following fields: Accent Label, Speaker Label, Gender, Word, Vowel Phoneme, Class Number, Formant 1 (Hz), Formant 2 (Hz), Formant 3 (Hz).

Classifier Implementation

In this task, a k-Nearest Neighbour Classifier, which tries to determine what phoneme of speech is being detected based on the formant values that were setup during the Feature Extraction process, was built. For this classifier, the data is split up such that 75% of the extracted data for each gender is used as the training set, while the remaining 25% for each gender is used as the test set. This was done for further analysis which will be explained later. The split data for each gender is then combined together accordingly so as to test on the entire data.

To determine the correct value of k, all the odd values between 1 and 21 were used, making sure to calculate their F1-Score. Several distance metrics were also tested, including Euclidean distance, Manhattan distance and cosine distance in order to determine which metric gives the best results.

Evaluation

After testing all possible values of k, along with all possible distance metrics, it was determined that the optimal value for k is 5, and that Euclidean distance provides the highest performance, as this combination produced the highest F1-Score of 1. In fact, other distance metrics returned a slightly lower F1-Score when k=5. Despite this, there was a range of values between 5 and 11 that also resulted in a perfect F1-Score, using the same distance metric. This means that using a k-Nearest Neighbour Classifier with k=5, with Euclidean distance, leads to only correct predictions.

Once the classifier has been trained and all predictions are made on the test set, a confusion matrix, which can be seen in Figure 1 below, was generated in order to visualize the number of True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN). As expected, using the parameters mentioned above resulted in 0 False Positives or False Negatives. This means that there are only True Positives and True Negatives, depending on which phoneme class is being analysed.

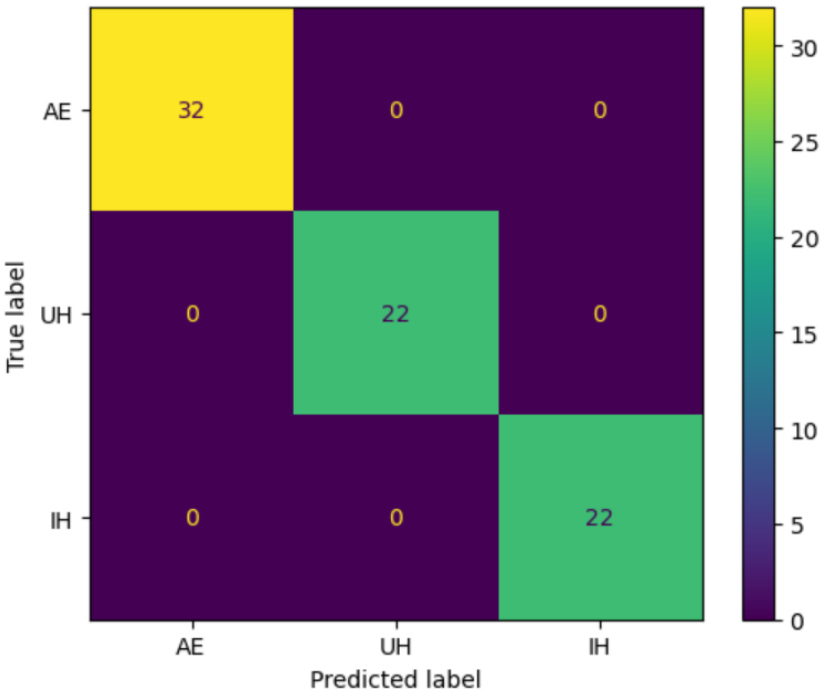

Figure 1 - Confusion Matrix for kNN Classifier using k=5 and Euclidean Distance

As part of further analysis, two separate k-Nearest Neighbour Classifiers were built, each taking the data from each separate gender, to test if there is similar performance when classifying the data for only a single gender, instead of combining both genders together. The same value for k (k=5) and the same distance metric (Euclidean) were used. As expected, this resulted in exactly the same F1-Score of 1. This means that there is perfect performance of the classifier on the data, regardless of whether the data is split by gender or not. Afterwards, the value for k and the distance metric were changed to test this further. The result was that the range of values for k for which a perfect F1-Score of 1 for each gender is produced changed to between 3 and 5. The values between 7 and 11 resulted in slightly lower F1-Scores. Changing the distance metric, while keeping the same value of k=5 resulted in the F1-Score for each gender being slightly lower or higher depending on the metric used.

Lastly, the original confusion matrix was checked to detect any vowel-based phoneme pairs that produce the most confusion, if any. As expected, since there are 0 False Positives or False Negatives, there are no pairs of phonemes that produce confusion. However, this becomes increasingly apparent when changing the value of k or the distance metric.

Conclusion

In conclusion, the creation and evaluation of a dataset containing data about three different vowel phonemes using a k-Nearest Neighbour classifier has yielded some important insights and interesting observations. The evaluation of this project indicated that, by using the correct value of k, along with the correct distance metric, there can be up to perfect performance.

Despite this, there were several noteworthy limitations. One such limitation is that the dataset is quite small. This can be improved by increasing the number of male and female speakers per accent or increasing the number of accents to analyse. This is a major reason why there are perfectly correct predictions when using k=5 and Euclidean distance.

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

Declaration

Plagiarism is defined as “the unacknowledged use, as one’s own work, of work of another person, whether or not such work has been published” (<u>Regulations Governing Conduct at Examinations</u>, 1997, Regulation 1 (viii), University of Malta).

I / ~~We~~*, the undersigned, declare that the [assignment / ~~Assigned Practical Task report / Final Year Project report~~] submitted is my / ~~our~~* work, except where acknowledged and referenced.

I / ~~We~~* understand that the penalties for making a false declaration may include, but are not limited to, loss of marks; cancellation of examination results; enforced suspension of studies; or expulsion from the degree programme.

Work submitted without this signed declaration will not be corrected, and will be given zero marks.

* Delete as appropriate.

(N.B. If the assignment is meant to be submitted anonymously, please sign this form and submit it to the Departmental Officer separately from the assignment).

Matthias Vassallo Pulis

Student Name | Signature

Student Name | Signature

Student Name | Signature

Student Name | Signature

ARI2203/ICS2203
Course Code

Statistical NLP - Vowel Analysis & Classification Assignment
Title of work submitted

27/05/2024
Date