Hierarchical K-Means Algorithm Applied On Isolated Malay Digit Speech Recognition

S. A. Majeed*, H. Husain, S. A. Samad, and A. Hussain

Digital Signal Processing Lab, Department of Electrical, Electronic and System Engineering, Universiti Kebangsaan Malaysia, 43600 UKM, Bangi Selangor, Malaysia

*Sayf_alali@yahoo.com

Abstract. In recent years, there has been an increasing interest in speech recognition in terms of accuracy. In this paper, the implementation of a speech recognition system in a speaker-independent isolated Malay digit was discussed. The system is developed applying Hierarchical K-means clustering approach that combines the K-means and the Hierarchical algorithm. To recognize the Malay speech digits, the Mel Frequency Cepstral Coefficient technique (MFCC) is used to extract speech features, Hierarchical K-means used as a clustering technique for training and testing of the feature's vectors. The performance of the system was evaluated. And the overall speech recognition accuracy attained 87.5% which is considerably satisfactory.

Keywords: Clustering, Hierarchical K-Means, MFCC, Speech Recognition

1. Introduction

Speech recognition can be approximately divided into two stages: feature extraction and classification. Feature extraction is defined as a step to minimize the dimensionality of the input data, a minimization which certainly causes to some information loss. In typical speech recognition systems, speech signals divide into frames and extract features from each frame. Through feature extraction, speech signals are changed into a sequence of feature vectors. Then these vectors are moved to the classification stage. Information loss during the transition from speech signals to a sequence of feature vectors must be kept to a minimum[1]. Feature extraction stage plays a vital essential role in the realization of speech recognition systems. This is the result of the fact that better signal feature extraction gives rise to better recognition performance[2].

Mel-Frequency Cepstral Coefficients (MFCC), and Perceptual Linear Prediction (PLP) techniques are famously used at the feature extraction stage because of their ability to estimate the process of human hear and perceive sounds with different frequencies. Initially, at the preprocessed stage which contains the steps of sampling, quantification, windowing, endpoint detection, filtering, etc. is applied to the input speech signal. MFCCs are the most commonly used acoustic features in speech recognition and have been used in this paper[3].

For classification stage several approaches have been recommended such as Hidden Markov Models (HMM), Dynamic time warping (DTW), Support Vector Classifiers, Artificial Neural Networks (ANN), which among them neural networks have proven to be very efficient[4]. Different neural network's architectures have been used in recent years for the speech recognition task[5]. A contemporary study on isolated Malay digit recognition applying dynamic time warping (DTW) has 80.5 % recognition accuracy[6]. Reem Sabah & Raja N. Ainon in 2009 used Adaptive Neuro-Fuzzy Inference System (ANFIS) to recognize the Malay speech digits has 85.24 % recognition accuracy[7].

In this work unsupervised classification or clustering is used to optimally classify the feature vectors which are extracted from Malay speech digits.

2. Clustering Process

In recent years, there has been an increasing interest in Clustering. Moreover clustering or unsupervised classification can be defined as a separation of data into groups (clusters) of similar objects. Within the same cluster, objects ought to be similar to one another and dissimilar from the objects in other clusters.

However, the better clustering is obtained when the greater the similarity within a group and the dissimilarity between the other groups[8]. In our implementation, clustering is applied on database containing feature vectors extracted from Malay digits utterances.

2.1 Hierarchical and K-means clustering

K-means approach which has developed by Mac Queen in 1967 is the most famous methods for clustering. Notwithstanding the simplicity of K-means, various fields were used this algorithm. In general, K-means is a division clustering process that isolates input data into *k* mutually excessive groups. By repeating such division, K-means reduces the sum of distance from each input data to its clusters. In addition, K-means is very well known process because of its capability to cluster massive data quickly and efficiently[9].

However, applying K-means clustering algorithm with the Euclidean distance measure has been given some providential results where the distance is calculated by finding the distance square between each score, then summing up the distance square and finally finding the square root of the sum[10-11]. Mathematically Euclidean Distance can be expressed as:

$$\text{Euclidean Distance } (X_i, Y_i) = \sqrt{\sum_{i=1}^{n} (X_i - Y_i)^2} \tag{1}$$

On the other hand, hierarchical clustering is used widely as a practical alternative for clustering. In addition hierarchical clustering nodes are arranged in particular fashion, and there is an option for users to evaluate and analyze data at different levels of abstraction by expanding and collapsing these nodes. On the positive side, hierarchical clustering algorithms do not necessitate to know the number of clusters in advance[12].

Therewith, hierarchical and K-means clustering are two main analytical methods for unsupervised learning. Anyhow, both have their natural limitations. Hierarchical clustering cannot signify obvious clusters with similar expression patterns. Furthermore, the actual expression patterns become less relevant when clusters expand in size[13].

In another hand, K-means algorithm is highly sensitive in the initial starting points. The initial cluster is created by K-means randomly. Whereas random initial starting points near to the latest solution, K-means has a high prospect to detect the clusters' centre. Furthermore, it will cause to inaccurate clustering results. Given that K-means does not guarantee the unique clustering results because of initial starting points created randomly[14].

In order to solve the cluster initialization for K-means, Kohei Arai and Ali Ridho Barakbah in 2007 proposed the hierarchical K-means algorithm which combines the K-means and Hierarchical algorithm. The clustering results by K-means are utilized. In this algorithm, initial centroids for K-means were determined[14].

In our application the hierarchical K-means algorithm is applied on database containing feature vectors extracted from Malay Digit Speech.

2.2 Hierarchical K-means

The better result of K-means clustering can be reached after working on some experiments. All the same, it is hard to determine the limitation of experiments which could reach the preferable result. There has been no prior knowledge about the numbers of experiments performed that are sufficient to achieve the best result or possibly whether the next experiment will reach the better result. This variety of uncertainness causes the K-means algorithm comparatively difficult to be utilized in actual clustering cases[14].

However, the result of K-means clustering can be treated as useful input to have the better result, whilst it achieves the local optima, because it reflects the partitioned feature space based on the specific initial points that were produced randomly[14].

Since K-means produces random initial starting points, therefore, the clustering results are not unique, and which the optimum results are anonymous. By calculating in specific periods, other clustering results can be created. For each calculation, the last centroids are recorded. Subsequently, all final centroids are gotten

from specific computations; the hierarchical clustering algorithms are implemented. The final centroids after using the hierarchical algorithm can be applied as initial centroids for K-means[14].

2.3 Algorithm

In this subsection, the implementation steps of Hierarchical K-means algorithm to determine initial centroids for K-means were presented. The algorithm is described as follows:

1. Let A = {ai}, i = 1,..., n. n-dimensional vector.
2. Let X = {xi}, i = 1,..., r: each data of A.
3. Set K as the predefined number of clusters.
4. Determine p as numbers of computation.
5. Set i = 1 as initial counter.
6. Apply K-means algorithm.
7. Record the centroids of clustering results as Ci = {cij} j = 1,…, K.
8. Increment i = i + 1.
9. Repeat from step 6 while i<p.
10. Assume C = {Ci} | i = 1,…, p as new data set, with K as predefined number of clusters
11. Apply hierarchical algorithm
12. Record the centroids of clustering result as D = {di} i = 1,…, K.

Then, applying D = {di} i = 1,…, K as initial cluster centers for K-means clustering[14].

3. Dataset Description

The experiments were performed using Spoken Malay Digit. A number of 50 individual Malay native speakers were involved to utter all the ten digits ten times. Depending on this, the database comprises of 5000 tokens (10 digits × 10 repetitions × 50 speakers). In this experiment, the data set is divided into two parts: a training set with 70% of the samples and test set with 30% of the samples.

4. Experimental Results

In this study, a system that recognizes isolated Malay digits from zero to nine is implemented. The first step in the system development is the collecting of the speech file corpus at 48 KHz with 16 bits. At the second step features of the input file are extracted by implementing MFCC technique to acquire feature vectors. Figure 1 shows the step of computing MFCC. Third, these features' vectors are divided into two groups for training and testing.
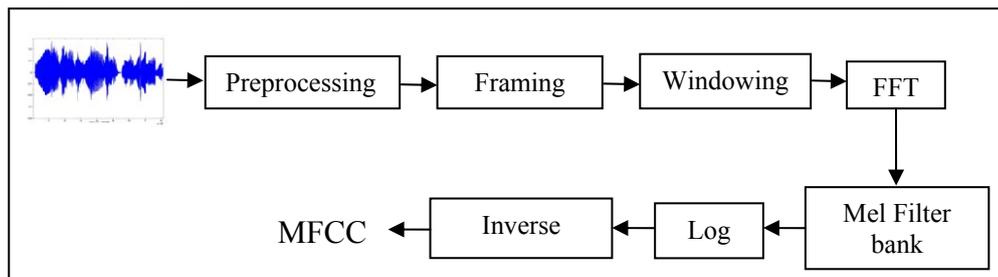

Figure 1. The steps involved in computing MFCCs

Fourth, use K-means algorithm on the Training data and Record the centroids of clustering results. Fifth, apply hierarchical algorithm and the final centroids will be used as initial centroids for K-means. Sixth, test the system and determine the recognition accuracy.

The recognition accuracy of the trained system is estimated according to the formula (1) below:

$$Recognition\ accuracy = \frac{Correct\ Recognition}{total\ Number of\ testing\ sample\ for\ each\ digit} \times 100\% \tag{1}$$

The values of K which is the branching factor and the number of clustering iterations have been selected as 3, 30 respectively. These values provide better recognition accuracy

Table (1) and Figure (2) present the recognition accuracy for each digit. The digit Tiga which means three has attained the poorest recognition accuracy (76%). Where it has been recognized as a digit Lima because of the syllable's similarity.

The speech recognition system designed in this work has provided satisfactory results. With the system, a small vocabulary consisting of Malay digits have been recognized with high accuracy. Recognition accuracy resulted in this work are 86.2%, and 87.5 % for branching factor K=2, and 3 respectively.

Table 1. Recognition accuracy for each digit

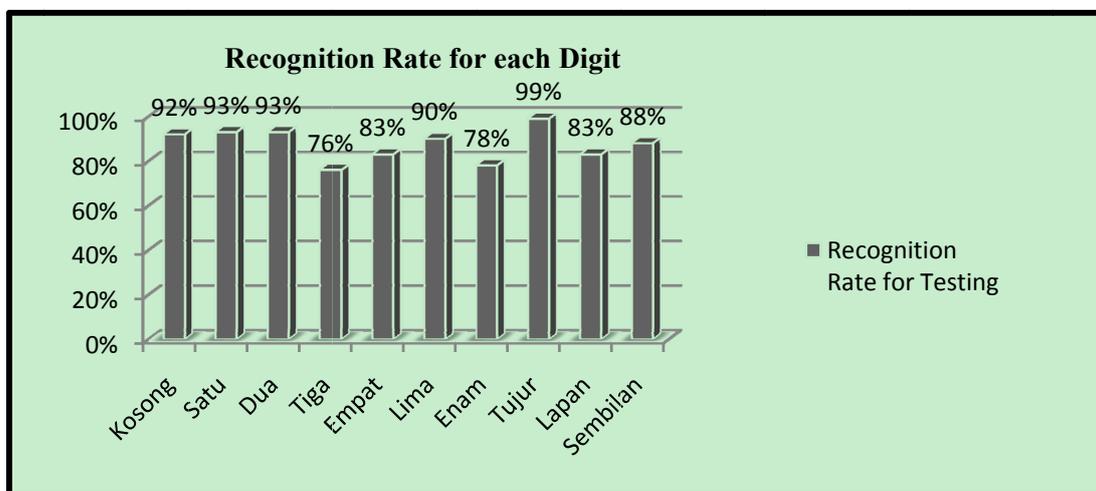| Digit	Recognition accuracy K=2	Recognition accuracy K=3
Kosong	90%	92%
Satu	92%	93%
Dua	93%	93%
Tiga	76%	76%
Empat	82%	83%
Lima	89%	90%
Enam	76%	78%
Tujur	96%	99%
Lapan	82%	83%
Sembilan	86%	88%
Total Average	86.2%	87.5%

Figure 2. Recognition accuracy for each digit

5. Conclusion

In this paper, Hierarchical K-means has been used as a clustering technique for training and testing the isolated Malay digits feature vectors, which are extracted from MFCC. Both the speed of K-means algorithm and the precision of the hierarchical algorithm are given preference in Hierarchical K-means. The system showed promising results in recognition accuracy. The overall recognition accuracy reached was 87.5 % and is comparatively better than the DTW and ANFIS.

6. References

[1] Lee, C., Hyun, D., Choi, E., Go, J. and Lee, C., "Optimizing Feature Extraction for Speech Recognition," IEEE Transactions On Speech And Audio Processing 11 (1), 80-87 (2003).

[2] Anusuya, M. and Katti, S, "Front end analysis of speech recognition: a review" International Journal of Speech Technology 14 (2), 99-145 (2011).

[3] Xin-Guang, L., Min-Feng, Y. and Wen-Tao, H. "Speech Recognition Based on K-Means Clustering and Neural Network Ensembles." Paper presented at the Natural Computation (ICNC), 2011 Seventh International Conference 26-28 (2011).

[4] Dede, G. and Sazlı, M. H., "Speech recognition with artificial neural networks," Digital Signal Processing 20 (3), 763-768 (2010).

[5] Nakagawa, S., Okada, M., and Kawahara, T. [Speech, Hearing and Neural Networks Models], IOS Press, Amsterdam (1995).

[6] Al-Haddad, S.A.R., Samad, S.A., Hussain, A. and Ishak, K.A., "Isolated Malay Digit Recognition Using Pattern Recognition Fusion of Dynamic Time Warping and Hidden Markov Models", American Journal of Applied Sciences 5 (6), 714-720 (2008).

[7] Sabah, R. and Ainon, R. N., "Isolated Digit Speech Recognition in Malay Language using Neuro-Fuzzy Approach", 3rd Asia International Conference on Modeling & Simulation, 336-340 (2009).

[8] Lamrous, S. and Taileb, M., "Divisive Hierarchical K-Means", Computational Intelligence for Modelling, Control and Automation, and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, International Conference, 18-18 (2006).

[9] MacQueen, J., "Some methods for classification and analysis of multivariate observations", Proceedings Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1 281-297 (1967).

[10] Oyelade, O. J., Oladipupo, O. O., and Obagbuwa, I. C., "Application of k Means Clustering algorithm for prediction of Students Academic Performance", International Journal of Computer Science and Information Security, IJCSIS 7 (1), 292-295 (2010).

[11] Fahim A. M., Salem A. M., Torkey F. A. and Ramadan M. A., "An efficient enhanced k-means clustering algorithm," Journal of Zhejiang University Science A., 1626–1633 (2006).

[12] Malik, H. H., Kender, J. R., "Classification by Pattern-Based Hierarchical Clustering", Data Mining, 008. ICDM '08. Eighth IEEE International Conference, 923 – 928 (2008).

[13] Chen, B., Tai, P. C., Harrison, R. and Yi, P., "Novel hybrid hierarchical-K-means clustering method (H-K-means) for microarray analysis", Computational Systems Bioinformatics Conference, 2005. Workshops and Poster Abstracts. IEEE, 105-108 (2005).

[14] Arai, K., and Barakbah, A. R., "Hierarchical K-means: an algorithm for centroids initialization for K-means", Reports of the Faculty of Science and Engineering, Saga University 36 (1), (2007).