

Speaker identification analysis for SGMM with k -means and fuzzy C -means clustering using SVM statistical technique

K. Manikandan^{a,*} and E. Chandra^b

^a*Department of Computer Science, PSG College of Arts and Science, Coimbatore, India*

^b*Department of Computer Science, Bharathiar University, Coimbatore, India*

Abstract. Speaker Identification denotes the speech samples of known speaker and it identifies the best matches of the input model. The SGMFC method is the combination of Sub Gaussian Mixture Model (SGMM) with the Mel-frequency Cepstral Coefficients (MFCC) for feature extraction. The SGMFC method minimizes the error rate, memory footprint and also computational throughput measure needs of a medium-vocabulary speaker identification system, supposed for preparation on a transportable or otherwise. Fuzzy C -means and k -means clustering are used in the SGMM method to attain the improved efficiency and their outcomes with parameters such as precision, sensitivity and specificity are compared.

Keywords: k -means, fuzzy C -means, SGMFC, speaker identification, SVM

1. Introduction

The speaker identification method depends on options influenced by each the organic structure of associate individual's vocal tract and therefore the activity characteristics of the individual [1,9,13]. The objective of acoustic modeling is to take information from isolated data as an input system which are built using large quantities of training data. However, the performance of k -means and fuzzy C -means clustering differs for same input speech and the better performance achieved in Fuzzy C -means. In classification, SVM is implemented in SGMFC method. The fuzzy based speaker identification system increase the accuracy rate of the system and obtained efficient results.

On comparison with the k -means, the SGMFC fuzzy C -means method increases the accuracy rate. SVM classification leads to better training to the parame-

ters [24]. Acoustic modeling has following advantage over other acoustic systems.

- The system takes input from isolated data as an input and performs the speech processing.
- Feature Extraction, Clustering and Classification are done with the best approaches and the results obtained in speaker identification system.
- The Fuzzy based system has efficient speech process on comparing with the k -means based speech processing system.
- The overall speaker recognized processing accuracy rate has been increased in the SGMFC system. By implementing confusion matrix, the accuracy rate has been calculated and the results are produced.

2. Literature review

Reynolds et al. [1,3,4] proposes speaker identification and verification through Gaussian Mixture model for high recognition accuracy. The maximum likelihood classifier approach is described for the speaker identifi-

*Corresponding author: K. Manikandan, Department of Computer Science, PSG College of Arts and Science, Coimbatore 641015, India. E-mail: k.manikandan@gmail.com.

cation and the likelihood ratio hypothesis is discussed for the speaker verification using background speaker normalization.

Leon et al. [26] proposes GMM based speaker identification using simple *k*-means clustering results in greater speed up gains. The clustering method results in actual speed-up factors as high 74x with no loss in accuracy.

Verma and Khanna [28] proposes speaker identification using *k*-means algorithm for MFCC features with average classification accuracy of 81%. It provides automatic language identification system using SVM classification and *k*-means algorithm.

Povey et al. [27] propose Sub gaussian mixture model for symmetric approach. It avoids the likelihood evaluation and parameter estimation. It provides explicit factorization of speech and speaker information.

Baid and Talbar [31] propose the study of *k*-means, Gaussian Mixture Model (GMM) and Fuzzy *C*-means clustering for the problem in brain tumor segmentation analysis.

Sahu and Dharmale [30] observed voice communication using FCM and *k*-means algorithm. It provides better result for FCM (Fuzzy *C*-means) when compared to *k*-means.

Ghosh and Kumar Dubey [29] proposed where behavior pattern of *k*-means and FCM are analysed.

3. SGMFC methodology

In Gaussian distribution model, accuracy of the result is low since it depends upon the short segments of the single speaker [31]. To overcome this problem UBM model is introduced in speaker identification. Here improvement is shown better whereas the problem of speaker sub segment is not concentrated. Hence the proposed SGMFC method concentrates on the sub segment with efficient accuracy. The feature extraction of the SGMFC is based on Sub Gaussian Mixture Model (SGMM) and Mel frequency cepstral coefficient (MFCC). The clustering algorithms *k*-means and Fuzzy *C*-means are compared with the obtained features. The SVM classification provides better accuracy than the other classification. The SGMFC is discussed as follows.

4. SGMFC process

The SGMFC method consists of various processing steps as a system flow design which are discussed as follows. The speaker identification analysis for SGMM with Fuzzy *C*-means and *k*-means are also discussed.

4.1. Input audio files

The SGMFC method with Fuzzy *C*-means algorithm takes input from 32 audio files and processes it. The audio files are stored along with .wav extension the isolated words taken. The audios are recorded by using sound recorder with closed mikes in a silent room.

4.2. Database

Data from THE CHAINS CORPUS – Speech database designed to help to characterize speakers individually. There are 16 individual sentences selected from CSLU Corpus and TIMIT corpus, for further processing of the input audio is converted as .wav format.

4.3. Feature extraction module

The variation in the mean and mixture weight of an acoustic model in a subspace is said to be as SGMM. After acquiring the input data, SGMM is applied with feature extraction which is efficiently by using MFCC method. Preprocessing, framing, windowing, Mel Filter Bank as well as Frequency Wrapping are done for the input audio files and logarithm values taken. After taking the values of logarithm to calculate Discrete Cosine Transform (DCT) and they are step into clustering process.

4.4. Clustering module

The clustering technique involves in grouping the distinct type of data from the extracted values center point is calculated and based on distance between data points clustering is performed. In existing system *k*-means algorithm is used. In the SGMFC method the Fuzzy *C*-means clustering.

4.5. Classification module

In this SGMFC work classification is done using Support Vector Machine (SVM). The classification involves two processes i.e., Training and Testing. In training phase, all the training datasets will be trained and placed in the template database. In testing phase, the test dataset available in the test database will be trained and is compared with template database for the decision to be made.

4.6. Decision module

The decision is made based upon the match scores

Table 1
SGMM for *k*-means clustering

Speaker	<i>k</i> -means	Positive		Negative		Precision	Sensitivity or recall	Specificity	Accuracy
		True	False	True	False				
1	6	6	0	52	2	1	0.75	1	0.87
2	5	5	2	53	0	0.71	1	0.96	
3	2	2	0	55	3	1	0.40	1	
4	7	7	2	51	0	0.78	1	0.96	
5	4	4	1	55	0	0.80	1	0.98	
6	6	6	0	54	0	1	1	1	
7	8	8	0	51	1	1	0.89	1	
8	5	5	1	54	0	0.83	1	0.98	
9	3	3	2	55	0	0.60	1	0.96	
10	6	6	0	52	2	1	0.75	1	

generated by the classifier in this module. After the classification, the predicted results by using the two clustering techniques are compared with actual results. The performance analysis is done by using confusion matrix. Accuracy rate is calculated and analyzed.

5. Performance evaluation

The performance results obtained through the SG-MFC method helps in measuring performance of speaker identification method with precision, recall and accuracy.

5.1. Accuracy

Accuracy rate could be calculated from formula given as follows,

$$\text{accuracy} = \frac{\text{no. of true positives} + \text{number of true negatives}}{\text{no. of true positives} + \text{false positives} + \text{false negatives} + \text{true negatives}} \quad (1)$$

5.2. Precision

Precision value with their true positive and false positive can be defined as,

$$\text{precision} = \frac{\text{number of true positives}}{\text{number of true positives} + \text{false positives}} \quad (2)$$

5.3. Sensitivity or true positive rate

The sensitivity is also called as recall, where there are two different incorrect conclusions which could be drawn in a statistical hypothesis test and it can be inappropriate. An analyzed data for a positive test which

accurately reflects the tested for activity. Let us consider p be the prediction, TPR be recall or True Positive Rate, TP be a true positive and FN be false negative which could defined as,

$$\text{True positive rate} = \frac{\text{True Positive}}{\text{Prediciton}} \quad (3)$$

$$\text{Prediciton} = \text{True Positive} + \text{False negative} \quad (4)$$

5.4. Specificity or true negative rate

The negative value results are obtained for TN. Let the TNR be the True Negative Rate, TN be true negative, FN be False negative and it can be defined as,

$$\text{True Negative Rate} = \frac{\text{True Negative}}{\text{Negative value}} \quad (5)$$

$$\text{Negative value} = (\text{True Negative} + \text{False Positive}) \quad (6)$$

Let FP be the false positive, TN be the true negative and α be the False Positive Rate (FPR) which can be defined as,

$$\text{False Positive Rate } (\alpha) = \frac{\text{False Positive}}{\text{False Positive} + \text{True Negative}} \quad (7)$$

5.5. False negative (FN)

The FN is obtained when the predicted output is n while the actual value is p . Let β be False Negative Rate (FNR), FN be False Negative, TP be True Positive defined by,

$$\text{False Negative Rate } (\beta) = \frac{\text{False Negative}}{\text{True Positive} + \text{False Negative}} \quad (8)$$

Table 2
SGMM for fuzzy *C*-means clustering

Speaker	<i>k</i> -means	Positive		Negative		Precision	Sensitivity or recall	Specificity	Accuracy
		True	False	True	False				
1	8	8	0	52	0	1	1	1	0.93
2	5	5	0	55	0	1	1	1	
3	3	3	2	53	2	0.60	0.60	0.96	
4	6	6	0	53	1	1	0.86	1	
5	4	4	0	56	0	1	1	1	
6	6	6	1	53	0	0.86	1	0.98	
7	9	9	0	51	0	1	1	1	
8	5	5	0	55	0	1	1	1	
9	2	2	0	57	1	1	0.67	1	
10	8	8	1	51	0	0.89	1	0.98	

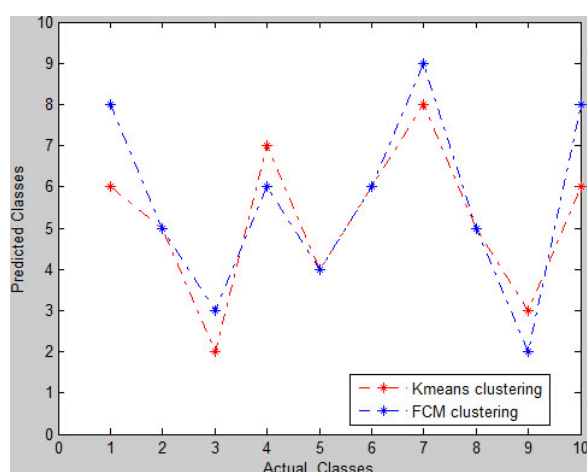


Fig. 1. Classes comparison.

6. Implementation result

The implementation results for the SGMFC method of SGMM with the *k*-means and Fuzzy *C*-means clustering algorithm are shown in Tables 1 and 2. The results are obtained by selecting various input of 10 speakers which could outcome with the precision, sensitivity, specificity and accuracy for comparison of outputs.

The classes comparison for *k*-means and Fuzzy *C*-means Clustering are shown in Fig. 1. The results are plotted by comparing the actual and predicted classes obtained from the table.

The accuracy of SGMFC with *k*-means and Fuzzy *C*-means clustering are plotted in Fig. 2. The accuracy (%) of the SGMFC method as shown in figure represents the Fuzzy *C*-means have high efficient results.

The error rate comparison between the *k*-means and FCM is shown in Fig. 3. From the obtained result it implicates that the FCM has low error rate than the *k*-means.

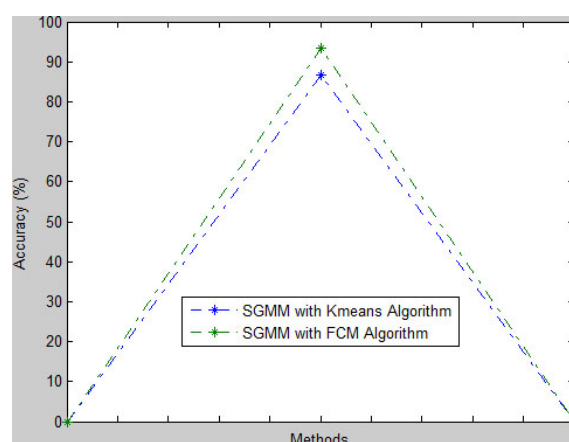


Fig. 2. Accuracy comparison.

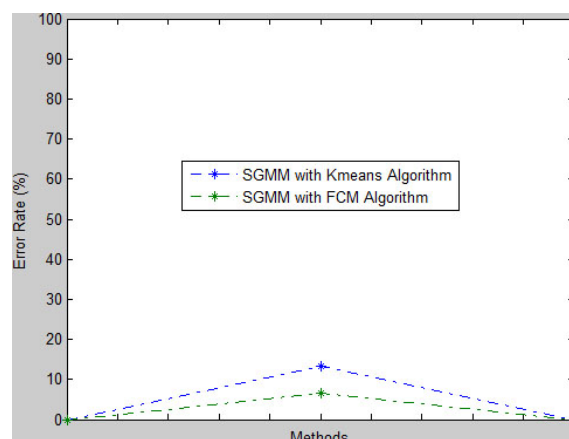
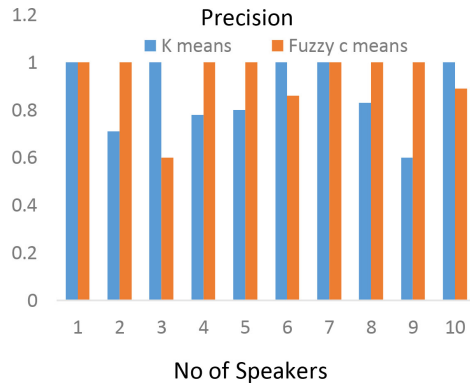
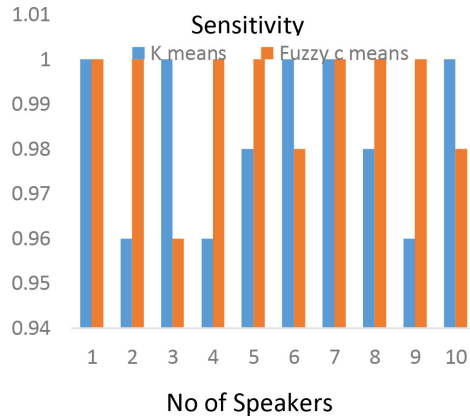
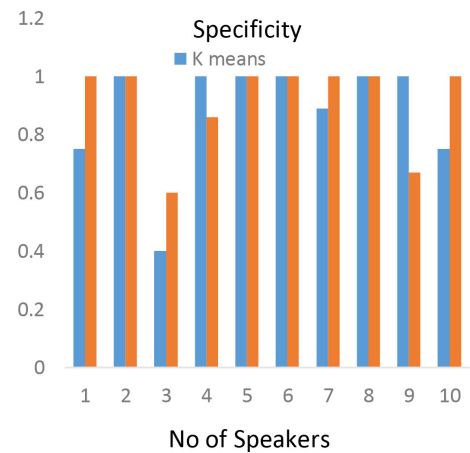


Fig. 3. Error rate comparison.

The precision can be calculated from the Eq. (2) for the SGMFC system and the obtained result is shown in Fig. 4.

The sensitivity can be calculated from the Eq. (3) for the SGMFC system between *k*-means and FCM as shown in Fig. 5.

Fig. 4. Precision chart for k -means and fuzzy C -means algorithm.Fig. 5. Sensitivity chart for k -means and fuzzy C -means algorithm.Fig. 6. Specificity chart for k -means and Fuzzy C -means algorithm.

The specificity can be calculated from the Eq. (5) for the SGMFC system between k -means and FCM as shown in Fig. 6.

6.1. Performance analysis comparison with literature paper using GMM method and proposed paper using sgmmfc method

Table 3
Comparison for SGMFC using fuzzy C -means clustering

Comparison with GMM and SGMFC	Sensitivity	Specificity	Accuracy
k -means in GMM	0.67	0.90	0.74
Fuzzy C -means with GMM	0.69	0.92	0.82
k -means using SGMFC	0.75	0.96	0.87
Fuzzy C -means with SGMFC	0.86	0.98	0.93

6.2. Advantages of proposed system

This approach is normally used to extract the unknown pattern from large set of data for business as well as real time application. It is a computational intelligence discipline which emerged as valuable tool for data analysis. The outcome of clustering process and efficiency of identification of speaker is generally determined through algorithms. This algorithm is widely applied in agriculture, engineering, astronomy, chemistry, geology, image analysis, medical diagnosis, and shape analysis and target recognition. Finally, the error rate is reduced by augmenting the standard feature vector with the cluster classification component.

6.3. Limitations and future direction

The study was limited to the text dependent speaker verification task. In future text independent speaker environment can be choose in an intended investigating proposed on various fusion methods. Modern hybrid.

7. Conclusion

The SGMFC method isolate SGMM data features and implemented in Fuzzy C -means for the known speech samples with the best matches of input model using SVM statistical method for speaker identification. Speaker identification is implemented by using FCM and k -means clustering. From the CHAINS CORPUS and Timit database 32 individual data's are taken as the input for the speaker characterization. After the characterization obtained, the Feature Extraction is done effectively. The results predicted are compared with Fuzzy C -means and k -means clustering for accuracy which obtained result as 93% and 87% respectively.

The experimental results shows that Fuzzy C -means clustering has produced high efficient and low error rate results when compared with k -means clustering. The proposed system using SGMMFC with Fuzzy C -means clustering provides better results.

References

- [1] D. Reynolds and R. Rose, Robust text-independent speaker identification using Gaussian mixture speaker models, *IEEE Trans Speech Audio Process* **3**(1) (1995), 72–83.
- [2] J. Makhoul, Linear prediction: A tutorial review, *Proc IEEE* **63**(4) (1975), 561–580.
- [3] D. Reynolds, Large population speaker identification using clean and telephone speech, *IEEE Signal Process Lett* **2**(3) (1995), 46–48.
- [4] D. Reynolds, Speaker identification and verification using Gaussian mixture speaker models, *Speech Commun* **17**(1–2) (1995), 91–108.
- [5] B. Pellom and J. Hansen, An efficient scoring algorithm for gaussian mixture model based speaker identification, *IEEE Signal Process Lett* **5**(11) (1998), 281–284.
- [6] A. Baraldi and P. Blonda, A survey of fuzzy clustering algorithms for pattern identification, *IEEE Trans Syst, Man, Cybern, B: Cybern* **29**(6) (1999), 778–785.
- [7] D. Reynolds, T. Quatieri and R. Dunn, Speaker verification using adapted Gaussian mixture models, *Digital Signal Process* **10**(1–3) (2000), 19–41.
- [8] C. Wang, Prosodic modeling for improved speech identification and understanding, *Ph.D. dissertation, Mass Inst of Technol*, Cambridge, MA, 2001.
- [9] H. Ezzaidi, J. Rouat and D. O'Shaughnessy, Towards combining pitch and mfcc for speaker identification systems, in: *Proc 7th Eur Conf Speech Commun Technol*, 2001.
- [10] U. Chaudhari, J. Navratil, G. Ramaswamy and S. Maes, Very large population text-independent speaker identification using transformation enhanced multi-grained models, in: *Proc IEEE Int Conf Acoust, Speech, Signal Process (ICASSP'01)* **1** (2001), 461–464.
- [11] A. De Cheveigné and H. Kawahara, Yin, a fundamental frequency estimator for speech and music, *J Acoust Soc Amer* **111** (2002), 1917.
- [12] D. Hosseinzadeh and S. Krishnan, Combining vocal source and MFCC features for enhanced speaker identification performance using GMMS, in: *IEEE 9th Workshop Multimedia Signal Process MMSP'07*, 2007, pp. 365–368.
- [13] M. Grimaldi and F. Cummins, Speaker identification using instantaneous frequencies, *IEEE Trans Audio, Speech, Lang Process* **16**(6) (2008), 1097–1111.
- [14] V. Apsingekar and P. De Leon, Speaker model clustering for efficient speaker identification in large population applications, *IEEE Trans Audio, Speech, Language Process* **17**(4) (2009), 848–853.
- [15] A. Sarkar, S. Rath and S. Umesh, Fast approach to speaker identification for large population using mlr and sufficient statistics, in: *Proc National Conf Commun (NCC)*, 2010, pp. 1–5.
- [16] R. Togneri and D. Pallella, An overview of speaker identification: Accuracy and robustness issues, *IEEE Circuits Syst Mag* **11**(2) (2011), 23–61.
- [17] N. Dehak, P. Kenny, R. Dehak, P. Dumouchel and P. Ouellet, Front-end factor analysis for speaker verification, *IEEE Trans Audio, Speech, Lang Process* **19**(4) (2011), 788–798.
- [18] Y. Hu, D. Wu and A. Nucci, Pitch-based gender identification with two-stage classification, *Security Commun Netw* (2011).
- [19] N. Wang, P. Ching, N. Zheng and T. Lee, Robust speaker identification using denoised vocal source and vocal tract features, *IEEE Trans Audio, Speech, Lang Process* **19**(1) (2011), 196–205.
- [20] M. Diez, M. Penagarikano, A. Varona, L. Rodriguez-Fuentes and G. Borel, On the use of dot scoring for speaker diarization, *Pattern Recogn and Image Anal* (2011), 612–619.
- [21] S. Nakagawa, L. Wang and S. Ohtsuka, Speaker characterization and identification-speaker identification and verification by combining mfcc and phase information, *IEEE Trans Audio, Speech, Lang Process* **20**(4) (2012), 1085.
- [22] Y. Hu, D. Wu and A. Nucci, Fuzzy-clustering-based decision tree approach for large population speaker identification, *IEEE Transactions On Audio, Speech, and Language Processing* **21**(4) (2013).
- [23] E. Chandra, K. Manikandan and M. Sivasankar, A proportional study on feature extraction method in automatic speech identification system, *Ijireeice* **2**(1) (2014).
- [24] J. Mashao, A hybrid GMM-SVM speaker identification system, *AFRICON, IEEE* **1** (2004), 319–322.
- [25] S. Madikeri, P. Motlicek and H. Bourand, Combining SGMM speaker vectors and KL-HMM approach for Speaker Diarization, *ICASSP, IEEE*, (2015).
- [26] V.R.A. Leon, Speaker model clustering for efficient speaker identification in large population applications, *IEEE Transactions on Audio, Speech, and Language Processing* **17**(4) (2009), 848–853.
- [27] D. Povey and L. Burget, The subspace gaussian mixture-model – a structured model for speech recognition, *Computer-Speech and Language* **25**(2) (2011), 404–439.
- [28] V.K. Verma and N. Khanna, Indian language identification using k -means clustering and support vector machine (SVM), *Engineering and Systems (SCES), IEEE*, (2013).
- [29] S. Ghosh and S. Kumar Dubery, Comparative analysis of k -means and fuzzy C -means algorithm, *IJACSA* **4**(4) (2013).
- [30] S. Sahu and N. Dharmale, Controlling the application via speech processing through mel frequency cepstral coefficients and back propagation neural method **3**(2) (2016).
- [31] U. Baid and S. Talbar, Comparative mixture, fuzzy C -means algorithms for brain tumor segmentation, *ICCCASP* **137** (2016), 592–597.

Copyright of International Journal of Knowledge Based Intelligent Engineering Systems is the property of IOS Press and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.