International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249 – 8958 (Online), Volume-9 Issue-4, April 2020

Speaker Diarization based on Black-Hole Entropy Fuzzy Clustering using Cepstral Features

V. Subba Ramaiah, S. Srinivasa Rao, V.S.N.Kumar Devaraju

Abstract*: Speaker diarization is the process of identification of the speaker in an audio sequence. This paper proposed a speaker diarization method using the Black-hole entropy fuzzy clustering and multiple kernel weighted Mel frequency cepstral coefficient (MKMFCC) parameterization. Initially, the MKMFCC descriptor extracted the cepstral features from the input audio signal. These features are used for clustering the speakers as groups for which the BHEFC is used. The feature parameter uses the audio signal containing both the high and low energy frame for speaker indexing that resulted in accurate separation of speaker. The performance evaluation of the proposed speaker diarization system is analyzed using the measures, such as F-measure, diarization error rate, and false alarm rate. The proposed MKMFCC with BHEFC obtained a minimum diarization error rate of 0.2447, maximum F-measure of 0.8526 and minimum false alarm rate of 0.4299, respectively while changing the wavelength and obtained a minimum diarization error rate of 0.2447, maximum F-measure of 0.8526 and minimum false alarm rate of 0.4298 when compared to the existing methods for the change in the frame length.

Keywords*: Black-hole entropy fuzzy clustering, multiple kernel weighted Mel frequency cepstral coefficient, Speaker diarization.

I. INTRODUCTION

Speaker diarization plays a significant role in providing the auxiliary information and speaker indexes for improving the speech-to-text transcriptions. The speaker diarization operates in an open set manner and there is no constant model. The speaker diarization does not have the knowledge regarding the identity of the speaker and the number of speaker is partitioned into homogeneous speech region [16]. The components involved in speaker diarization are voiced activity detection (VAD), clustering, segmentation, and re-segmentation. In the speaker diarization, the segmentation algorithm is categorized into segmentation based on model, segmentation that was guided by the decoder and segmentation that was based on metric. Clustering is one of the major parts in speaker diarization. The clustering groups the segments of the audio sources, such as music, speaker, and noise. The diarization procedure is classified into bottom up approach and top down approach depending on the clustering approach [1].

During the segmentation stage, the bottom-up clustering considers the individual segment that is obtained as separate clusters. The closest clusters are merged in the bottom-up clustering until the stopping criterion is reached. Hierarchical agglomerative approach is one of the examples of the bottom-up approach. The top-down clustering iteratively splits the single model formed from the entire audio into sub clusters. Divisive Hierarchical clustering is the example of top-down clustering. Besides the advantages, both the approaches have error propagation problem. The error propagation is overcome by minimizing the dispersion within the clusters through the clustering algorithms, like Integer Linear Programming (ILP) [15]. Even though, the ILP overcomes the problem, the attributes can be modeled by starting with the initial clusters that contains a sufficient numbers of samples thus, the bottom-up clustering is followed by ILP. The distance measures are also used for the determination of the segments of the same class. The distance measures, like generalized log-likelihood ratio (GLR) [8], Bayesian information criteria (BIC) [9], probabilistic linear discriminant analysis (PLDA) based distance [11], cosine distance score (CDS) [12] and Kullback-Leibler (KL) divergence [10]. Deep neural network has been applied for i-vector extraction [13] and speaker embedding feature extraction [14] in addition with the speaker clustering.

The main objective of the paper is the development of the speaker diarization technique using the BHEFC and MKMFCC parameterization. The MKMFCC descriptor helps in the extraction of cepstral features from the audio signal and accordingly, the speakers are clustered together as groups using the BHEFC algorithm.

II. LITERATURE SURVEY

Yu, C.et al.[1] developed a speaker diarization method using active learning. There are two active learning approaches in which the first active learning method was developed for acoustic clustering, whereas the second active learning method identified the speaker by the conversion of unsupervised task into semi-supervised task. Although this method reduced the diarization error rate, it failed to remove the human errors.

Revised Manuscript Received on April 11, 2020.
* Correspondence Author
Dr. V. Subba Ramaiah*, CSE, Mahatma Gandhi Institute of Technology, JNTUH, Hyderabad, India. Email: vsubbaramaiah_cse@mgit.ac.in
Dr. S. Srinivasa Rao, ECE, Mahatma Gandhi Institute of Technology, JNTUH, Hyderabad, India. Email: ssrinivasarao_ece@mgit.ac.in
V.S.N.Kumar Devaraju, ECE, Mahatma Gandhi Institute of Technology, JNTUH, Hyderabad, India. Email: dvsnkumar_ece@mgit.ac.in

V. Subba Ramaiah and R. Rajeswara Rao [2] designed Tangent weighted Mel frequency cepstral coefficient (TMFCC) for speaker diarization. The LION algorithm was used for clustering the audio into particular speaker groups and the TMFCC was used as a feature parameter. This method had high tracking accuracy but the high computational complexity. Karim D. et al. [3] modelled a hybridization algorithm using K-means algorithm and differential evolution (DE) algorithm. Although this method accelerated the optimal classification search, it failed to evaluate the efficiency of the system in terms of variance ratio criterion (VRC) and trace within criterion (TRW). Le Lan G.et al. [4] developed a scalable unsupervised adaptation framework for speaker diarization. In this method, the computational requirements are low as the adaptation and the linking was based on the vectors. However, this method faced problems to link the process for bigger collections.

A. Challenges

The challenges faced during the speaker diarization are given below:

- In [4], the challenge was the implementation of the scalable unsupervised adaptation framework for bigger collections. For linking the process for the bigger collections, such as videos sharing platforms or daily shows particular attention was required.
- The active learning based speaker clustering method assumed that the perfect answers were provided by the human assistance to any query pair but there were human errors, which need to be removed for improving the performance [1].
- The main challenge in the hybridization algorithm using K-means algorithm and differential evolution (DE) algorithm was the evaluation of efficiency of the system in terms of VRC and TRW [3].

III. PROPOSED METHOD OF SPEAKER DIARIZATION USING BLACK-HOLE ENTROPY FUZZY CLUSTERING

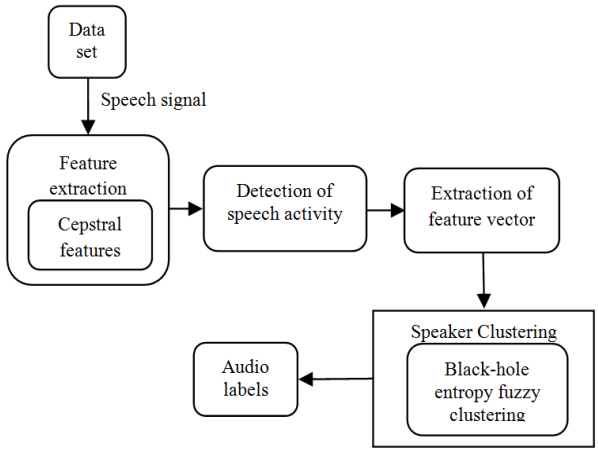

Fig. 1. Block diagram of the proposed MKMFCC based BHE fuzzy algorithm

In this paper, the speaker diarization is performed using the MKMFCC parameterization and BHEFC algorithm. Figure 1 depicts the block diagram of the proposed speaker diarization system using MKMFCC parameterization and BHEFC algorithm. Initially, the features from the input speech signal are extracted. Then, the speech activity is detected followed by the extraction of feature vector. The feature vector generated is used for clustering the speaker using BHEFC. Finally, the individual speakers in the audio signal are identified.

A. Feature Parameterization for speaker diarization

The properties related to the speaker are extracted for the proper diarization of the speaker. In this paper, the acoustic feature, like MFCC [7] is considered for the feature extraction rather than the sound source features as the formant information are preserved in the MKMFCC. The MKMFCC is selected as the feature parameter as the Mel frequency scale cannot perceive the frequencies over 1kHz even though it can capture the characteristics of phonetic components in the speech signal effectively. The performance of the diarization is improved by considering the weighted sum of the spectral components using the exponential and tangential function. The first step in the feature parameterization is pre-emphasis. In pre-emphasis, the high frequencies are highlighted by passing the input audio signal into the filter. The energy of the signal increases at high frequency and it is given as,

$$C(p) = B(p) - A * B(p-1) \quad (1)$$

Where A is a constant value for making the audio sample originate from the previous sample, the input and the output signal is denoted as B and C. The samples in the audio signal are represented from $p-1$ to p. The second step in the feature parameterization is framing. Framing is the process of segmenting the sample speech signal into small frames. In the audio streams, the adjacent frames are separated by the factor G, where $G < P$. The third step in the feature parameterization is hamming windowing. In the audio streams, the close frequency lines are integrated to perform the procedure of hamming window. The hamming window helps in the extraction of feature.

The hamming window function is represented as,

$$V(p): 1 \le p \le P-1 \quad (2)$$

After the hamming window function, the signal is represented as

$$C(p) = B(p) \times V(p) \quad (3)$$

The hamming window function is denoted as $V(p)$. The fourth step is the conversion of the audio signal to the frequency domain from the time domain using the discrete Fourier transform. The framed signal applied with the discrete Fourier transform is given as,

$$B_b(h) = \sum_{p=1}^{P} C(p).e^{-v2\pi hp}; 1 \le h \le L \quad (4)$$

Here the length of the discrete Fourier transform is given as h and the value compromising the analysis of the sample long window P, is given as $C(p) = B(p) * V(p)$. The fifth step in the feature parameterization is the Mel Filter Bank Processing. In the audio stream, the wide range of frequency makes the frequency range in a non-linear scale. Hence,

the unwanted information in the spectral is prevented in this step using the bank of filter. The signal frequencies are filtered using the triangular filter and then, the output from the triangular filter is approximated using the Mel scale for the estimation of weighted sum of the spectral components. The high and the low frequency components of the spectral from the periodogram is denoted as, R_H and R_L. FFT is used for the calculation of the Mel Filter bank equation, which is described as, $J(s) = (pFFT+1) * d(s) / Samplerate$. The filter bank creation is calculated using the Mel spaced frequencies values and the Mel scale values, which is given as

$$P_c(h) = \begin{cases} 0 & h < J(c-1) \\ \dfrac{h - J(c-1)}{J(c) - J(c-1)} & J(c-1) \le h \le J(c) \\ \dfrac{J(c+1) - h}{J(c+1) - J(c)} & J(c) \le h \le J(c+1) \\ 0 & h > J(c+1) \end{cases} \quad (5)$$

The number of Mel Filters varies from $c = 1 \; to \; R$. $R+2$ Mel spaced frequencies is given as $a()$. The filter bank energy is the multiplication of the power spectrum with the filter bank followed by the addition of coefficients. The filter bank energy is represented by

$$H(b) = \sum_{p=0}^{\frac{P}{2}} \log |B(p)| C_p\left(L.\frac{2\pi}{P}\right) * \mathbf{VW_p} \quad (6)$$

The MK weighted function is represented as $\mathbf{VW_p}$. The Multiple Kernel (MK) weighted function is the sum of Tangential weighted function and Exponential weighted function.

$$\mathbf{VW_p} = \mathbf{VW_{p1}} + \mathbf{VW_{p2}} \quad (7)$$

The tangential weighted function is given by

$$\mathbf{VW_{p1}} = \mathbf{tanh}\left(-\mathbf{R}/\mathbf{2} + \mathbf{R} \cdot \left[\frac{\mathbf{c-1}}{\mathbf{R-1}}\right]\right) \quad (8)$$

and the exponential weighted function is given as

$$\mathbf{VW_{p2}} = \mathbf{exp}\left(-\mathbf{R}/\mathbf{2} + \mathbf{R} \cdot \left[\frac{\mathbf{c-1}}{\mathbf{R-1}}\right]\right) \quad (9)$$

The sixth step in the feature parameterization is discrete cosine transform (DCT). The log Mel spectrum values are converted into time domain using DCT. The MKMFCC are the results from DCT. For the input utterance, the acoustic feature vector is represented as the set of coefficients, which is given as

$$H(b) = \tilde{H}(h) \quad (10)$$

where $\tilde{H}(h) = \begin{cases} H(b) & ,h = h_b \\ 0 & ,Otherwise \end{cases}$. The cepstral coefficient is calculated using the computed energy, which is given as,

$$VA_m(p) = \frac{1}{P'} \sum_{h=0}^{P'-1} \tilde{H}(L)\, e^{vh\left(2\pi/p\right)p} \quad (11)$$

The MKMFCC is given as $VA_m(p)$. The seventh step is the Delta energy and Spectrum. For covering the phonetic components, the Cepstral features along with the energy features are included for the determination of acoustic feature vector. The robustness of the echo, noise and the speech recognition accuracy are increased using Spectrum and Delta energy. The final step is the cepstral normalization. The cepstral normalization decreases the residual mismatches in the feature vector. The normalization is obtained by subtracting the average of the coefficients and dividing the variance.

B. Detection of Speech activity for speaker diarization

One of the important steps in speaker diarization is the detection of the activity of the speech. The identity of the speaker and the segmentation of the signal that is related to the identity are recognized in speech activity. The Bayesian inference criteria (BIC) are used as selection criteria for the detection of activity. The log-likelihood is maximized using the BIC criterion. The model identification in the time series, statistical modeling, and linear regression are done using the BIC criterion. The audio signals are modeled with the MKMFCC feature for the detection of speech activity using the GMM. For the respective audio segment, the threshold is predefined for modeling the speaker. Based on the threshold and the BIC score, the activity of the speaker is detected. If the threshold value is greater than the computed BIC score, it means the activity is not detected and when the threshold is less than the BIC score, the activity of the speaker is detected.

C. Extraction of feature vector for speaker diarization

The speech activity detected segmented signal is used for the extraction of the feature vector. The i-vector extraction model is used along with the UBM model for the extraction and the GMM model for the statistical value calculation of the Gaussian mixture components. The first and zero order statistics from the UBM along with the MKMFCC feature trains the UBM. The zero and the first order Baum-Welch Statistics are represented by

$$\begin{aligned} Q_l &= \sum_{\tau} \alpha_\tau(l) \\ R_l &= \sum_{\tau} \alpha_\tau(l)(I_\tau - p_l) \end{aligned} \quad (12)$$

Zero order Baum-Welch Statistics is represented as Q_l and the first order Baum-Welch Statistics is represented as R_l. The sub vector with respect to the mixture components is given as p_l and the posterior probability at the time t is given as $\alpha_\tau(l)$. The UBM model is trained using the statistics for the extraction of the features. The GMM mean super vector M is modeled along with the total variability space that combines channel space and the speaker space using the Joint factor analysis [6], which is given as

$$M = M_o + Sg \tag{13}$$

Here the super vector of the UBM is given as M_o, the normal distribution along with the low dimension matrix is given as g and the vector with total variability is given as W. Using the i-vector extraction, the vectors extracted from the features are represented as

$$E = \{e_1, e_2, ..., e_q\} \tag{14}$$

Where e is the vectors formed from the extracted features.

D. Speaker clustering using BHEFC

In this paper, the proposed methodology is the implementation of BHEFC algorithm [5] for the clustering of the speaker. The BHE-based Fuzzy Clustering algorithm is similar to Bayesian Fuzzy Clustering (BFC) with few modifications in the definitions. The BHE based fuzzy clustering algorithm is the integration of both the black hole phenomenon and clustering. When compared to the BFC, the BHE based fuzzy clustering algorithms required the determination of both the parameter of Dirichlet distribution and the fuzzifier, whereas the BHE based fuzzy clustering requires only the fuzzifier w. The fuzzy clustering algorithm finds the MAP values of the parameters by clustering effectively. The optimality guarantees are leveraged using the Markov Chain Monte Carlo technique in clustering algorithms. The samples are generated from the BFC with the help of Metropolis within Gibbs sampler. The generated samples are evaluated in the posterior and the best sample is retained. Metropolis-Hastings sampling step is used for accomplishing the conditional membership distribution provided the data and cluster prototypes Given the fixed values of cluster prototypes, the joint distribution of memberships of the data, membership and prototype, $\rho(Y, X, Z)$ is proportional to the conditional membership distribution, $\rho(X|Y, Z)$. At the data point index i, for the membership vector y_i^{ε}, the other membership cluster prototypes and the vector are unchangeable. Hence, the other quantities are evaluated,

$$\rho(z_i, y_i|Z) = \rho(z_i|y_i, Z)\,\rho(y_i)$$
$$\alpha \prod_{j=1}^{U} \exp\{-\frac{1}{2} y_{ij}^{w} \|z_i - x_j\|^2\} y_{ij}^{-w} \tag{15}$$

The membership sample y_i is replaced with the new membership sample y_i^{ε} and the probability of y_i^{ε} is equal to the below ratio,

$$\tilde{a}_y = \min\left\{1, \frac{\tilde{\rho}(z_i, y_i^{+}|Z)}{\tilde{\rho}(z_i, y_i|Z)}\right\} \tag{16}$$

The new cluster prototype is sampled from the conditional distribution of the prototype of data and memberships $\rho(Z|Y, X)$, which is proportional to the joint distribution $\rho(Y, X, Z)$ for fixed values of membership and data. For the cluster prototype x_j, the terms related to membership and cluster prototypes are unchanged. Hence, the following quantities are only evaluated.

$$\rho(Y, x_J|X) = \rho(Y|X, x_j)\,\rho(x_j)$$
$$\alpha \exp\{-\frac{1}{2}\sum_{i=1}^{O} y_{ij}^{w} \|z_i - x_j\|^2\}$$
$$\times \exp\{-\frac{1}{2}(T + \beta \sum_{i=1}^{O} \ln \|z_i - x_j\|^2\} \tag{17}$$

The probability is accepted with respect to the following ratio given by,

$$\tilde{a}_x = \min\left\{1, \frac{\rho(Y, x_j^{+}|X)}{\rho(Y, x_j|X)}\right\} \tag{18}$$

The joint probability function $\rho(Y, X, Z)$ is given as

$$\rho(Y, X, Z) = \rho(Y|X, Z)\tilde{\rho}(X|Z)\rho(Z)$$
$$\alpha \exp\left\{-\frac{1}{2}\sum_{i=1}^{O}\sum_{j=1}^{U} y_{ij}^{w} \|z_i - x_j\|^2\right\}$$
$$\times \left(\prod_{i=1}^{O}\prod_{j=1}^{U} y_{ij}^{-w}\right) \tag{19}$$
$$\times \exp\left\{-\frac{1}{2}\left(T + \beta \sum_{i=1}^{O}\sum_{j=1}^{U} \ln \|z_i - x_j\|^2\right)\right\}$$

For equation (19), the β value is set to 2 and x_j can be a distribution with larger constant T. The BHEFC algorithm has the characteristics of BFC such that the fuzzy clustering is realized by the Bayesian inference thus, mutually incorporating the probability and fuzziness in clustering.

IV. RESULTS AND DISCUSSION

The results and discussion of proposed MKMFCC based BHE fuzzy algorithm is deliberated below.

A. Experimental setup

The proposed MKMFCC based BHE fuzzy clustering algorithm is performed in a PC with Windows 7 Operating system and the dataset used for the performance evaluation is the ELSDSR corpus [17] dataset.

B. Performance metrics

The performance measures used for the evaluation of the proposed method are F-measure, Diarization error rate and false alarm rate.

C. Comparative methods

The proposed MKMFCC based BHE fuzzy clustering algorithm is compared with the existing methods, such as MFCC with LION, TMFCC with LION, and MKMFCC with WLI fuzzy for the evaluation of the method.

D. Comparative analysis

The comparative analysis of the proposed MKMFCC based BHE fuzzy clustering algorithm is done with the performance metrics, like F-measure, diarization error rate, and false alarm rate. The comparative analysis is done for the speaker diarization methods for three, four, five and six different speakers. The analysis is done by changing the length of the frame and the length of the wavelength.

E. Analysis based on Diarization error rate (DER)

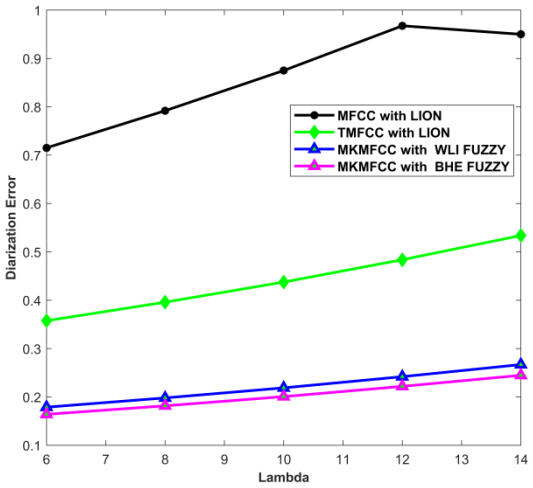

Fig. 2. DER for a change in wavelength

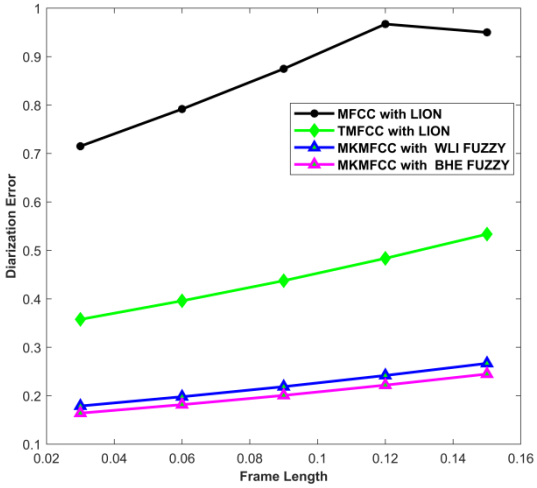

Fig. 3. DER for a Change in length of the frame

Figure 2 and Figure 3 depicts the evaluation of the Diarization Error rate by changing the wavelength and the frame length. Figure 2 shows the diarization error rate by changing the wavelength. For $\lambda=8$ the existing MFCC with LION, TMFCC with LION, MKMFCC with WLI fuzzy method and the proposed MKMFCC based BHE fuzzy algorithm obtained a diarization error rate of 0.7919, 0.3959, 0.1980 and 0.1816, respectively.

Figure 3 shows the diarization error rate by changing the frame length. For frame length = 0.15, the existing MFCC with LION, TMFCC with LION, MKMFCC with WLI fuzzy method and the proposed MKMFCC based BHE fuzzy algorithm obtained a diarization error rate of 0.9500, 0.5335, 0.2668 and 0.2447, respectively. The proposed MKMFCC based BHE fuzzy algorithm had minimum diarization error rate when compared to the existing methods.

F. Analysis based on F-Measure

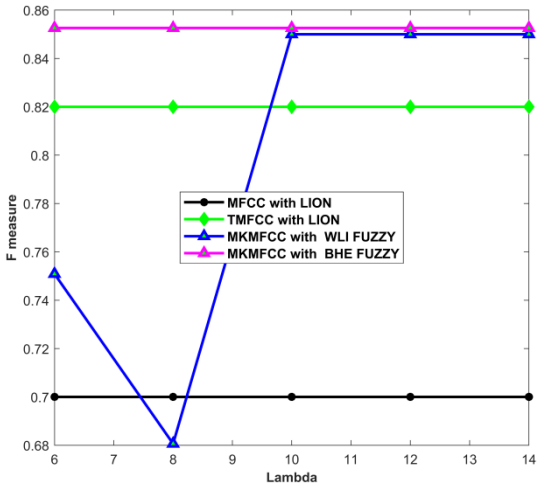

Fig. 4. F-measure for a change in wavelength

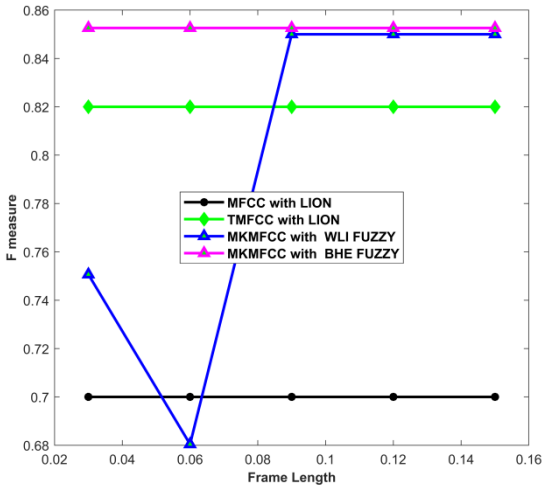

Fig. 5. F-measure for a change in length of the frame

Figure 4 and Figure 5 depicts the evaluation of the F-measure by changing the wavelength and the frame length. Figure 4 shows the F-measure by changing the wavelength. For $\lambda=8$, the existing MFCC with LION, TMFCC with LION, MKMFCC with WLI fuzzy method and the proposed MKMFCC based BHE fuzzy algorithm obtained a F-measure of 0.7000, 0.8200, 0.6806 and 0.8526, respectively.

Figure 5 shows the F-measure by changing the frame length. For frame length = 0.15, the existing MFCC with LION, TMFCC with LION, MKMFCC with WLI fuzzy method and the proposed MKMFCC based BHE fuzzy algorithm obtained a F-measure of 0.7, 0.82, 0.85 and 0.8526, respectively. The proposed MKMFCC based BHE fuzzy algorithm had minimum F-measure when compared to the existing methods.

G. Analysis based on false alarm rate

Figure 6 and Figure 7 shows the evaluation of the false alarm rate by changing the wavelength and the frame length.

Figure 6 shows the false alarm rate by changing the wavelength.

For $\lambda = 8$, the existing MFCC with LION, TMFCC with LION, MKMFCC with WLI fuzzy method and the proposed MKMFCC based BHE fuzzy algorithm obtained a false alarm rate of 0.9500, 0.6945, 0.3472 and 0.3190, respectively.

Figure 7 shows the false alarm rate by changing the frame length. For frame length = 0.15, the existing MFCC with LION, TMFCC with LION, MKMFCC with WLI fuzzy method and the proposed MKMFCC based BHE fuzzy algorithm obtained a false alarm rate of 0.9500, 0.9359, 0.4680 and 0.4298, respectively.

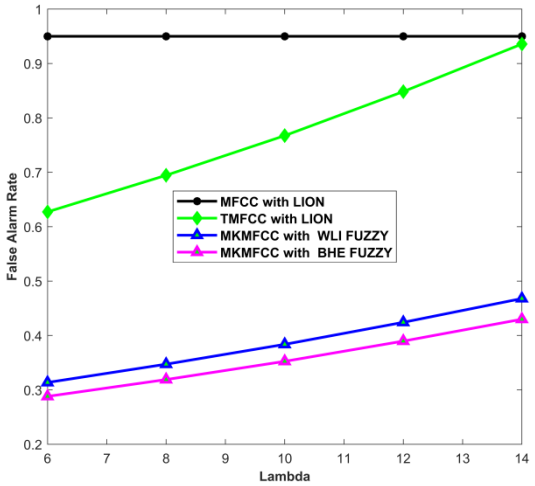

Fig. 6. False alarm rate for a change in wavelength

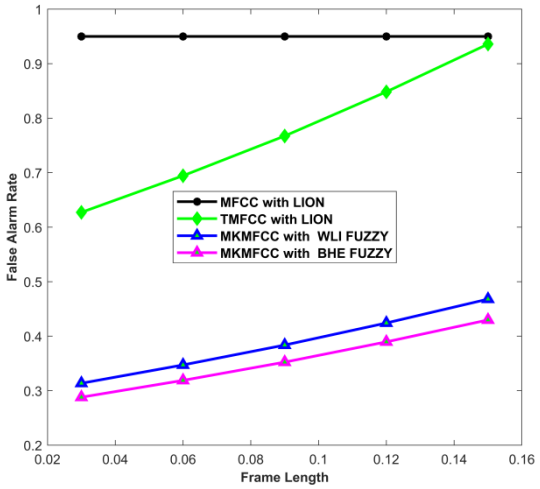

Fig. 7. False alarm rate for a change in length of the frame

V. CONCLUSION

In this research, a speaker diarization method using MKMFCC descriptor and Black-hole entropy fuzzy clustering was proposed. Initially, the features, such as cepstral feature is extracted from the input audio signal using the MKMFCC descriptor. The Black-hole entropy fuzzy clustering clusters the speaker as groups using the extracted features. The speaker clustering and feature parameterization is enhanced using speaker diarization. The performance of the proposed MKMFCC with BHE fuzzy is evaluated using the metrics, like F-measure, diarization error rate, and false alarm rate for six different speaker signals with ELSDSR corpus data sets audio signals. The proposed MKMFCC with BHE fuzzy obtained a minimum diarization error rate of 0.1816, maximum F-measure of 0.8526 and minimum false alarm rate of 0.3190, respectively for the change in wavelength and obtained a minimum diarization error rate of 0.2447, maximum F-measure of 0.8526 and minimum false alarm rate of 0.4298 when compared to the existing methods for the change in the frame length. This method identified the individual speaker from the multi-speaker effectively. The future enhancement can be done using different clustering algorithms for speaker diarization.

REFERENCES

1. Yu C., and Hansen J. H. L., Active Learning Based Constrained Clustering For Speaker Diarization, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol.25, no.11, pp.2188–2198, 2017.
2. V. Subba Ramaiah, and R. Rajeswara Rao, A novel approach for speaker diarization system using TMFCC parameterization and Lion optimization, Journal of Central South University, *springer-verilog*, vol.24, 2017, pp.2649–2663.
3. Karim D., Salah H., and Adnen C., Hybridization DE with K-means for speaker clustering in speaker diarization of broadcasts news, International Journal of Speech Technology, 2019.
4. Le Lan G., Charlet D., Larcher A., and Meignier S., An Adaptive Method for Cross-Recording Speaker Diarization, *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol.26, 2018, pp.1821–1832.
5. Liu J., Chung F.-L., and Wang S., Black Hole Entropic Fuzzy Clustering, *IEEE Transactions on Systems Man and Cybernetics: Systems*, vol. 48, 2018, pp. 1622–1636,.
6. Madikeri S, Himawan I, Motlicek P, Ferras M., Integrating Online I-vector extractor with Information Bottleneck based Speaker Diarization system, Idiap, 2015.
7. S. Davis P. Mermelstein, Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences, *IEEE Transaction on Acoustic Speech Signal Processing*, vol. 28, no. 4, 1980, pp. 357–366.
8. A. Solomonoff A. Mielke M. Schmidt and Gish, Clustering speakers by their voices, in Proceedings of the *IEEE International Conference on Acoustics Speech and Signal Processing*, vol. 2, 1998, pp. 757–760.
9. B. Zhou and J. H. L. Hansen, Efficient audio stream segmentation via the combined t/sup 2/statistic and Bayesian information criterion, *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 4, 2005, pp. 467–474.
10. M.A. Siegler, U. Jain, B. Raj, and R. M. Stern, Automatic segmentation, classification and clustering of broadcast news audio, in Proc. of DARPA speech recognition workshop, 1997.
11. S. J. Prince and J. H. Elder, Probabilistic linear discriminant analysis for inferences about identity, in *Proceedings of the IEEE 11th International Conference on Computer Vision*, 2007, pp. 1–8.
12. N. Dehak, P. J. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, Front-end factor analysis for speaker verification, *IEEE Transactions on Audio Speech and Language Processing*, vol. 19, 2011, pp. 788–798.
13. G. Sell, D. Garcia-Romero, and A. McCree, Speaker diarization with i-vectors from DNN senone posteriors," in *Proc. of Sixteenth Annual Conference of the International Speech Communication Association*, 2015.
14. M. Rouvier, P. Bousquet, and B. Favre, Speaker diarization through speaker embeddings, in *Proc. of Signal Processing* (EUSIPCO), 2015, pp. 2082–2086.
15. G. Dupuy, S. Meignier, P. Del´eglise, and Y. Esteve, Recent improvements on ILP based clustering for broadcast news speaker diarization, in Proceedings of Odyssey, 2014.
16. J.S.Sohal, Sukhvinder Kaur, Optimization of Speaker Diarization by Reducing Diarization Error Rate: A Review, International Journal of Electronics and Communication Engineering, 2015, pp. 84-87.
17. ELSDSR Dataset, http://cogsys.compute.dtu.dk/soundshare/elsdsr.zip, accessed on January 2020.

AUTHORS PROFILE

Dr. V. Subba Ramaiah is working as Assistant Professor, Department of CSE, Mahatma Gandhi Institute of Technology, JNTUH, Hyderabad, India. He received his Ph.D. (Computer Science & Engineering) degree from JNTUH, India. He has 17 years teaching experience and has published 15 research papers in International journals, conferences. His research interest areas include Speaker Diarization, Image Processing, Wireless Networks, Machine Learning, Deep Learning and IoT.

Dr. S. Srinivasa Rao is working as Associate Professor, Department of ECE, Mahatma Gandhi Institute of Technology, JNTUH, Hyderabad, India. He received his Ph.D. (Electronics and Communication Engineering) degree from ANU, Guntur, Andhra Pradesh, India. He has 18 years teaching experience and has published 8 research papers in International journals, conferences. His research interest areas include Embedded Systems, Signal Processing, Wireless Networks, and IoT.

Mr. V.S.N.Kumar Devaraju is working as Assistant Professor, Department of ECE, Mahatma Gandhi Institute of Technology, JNTUH, Hyderabad, India. He received his M.Tech (Systems & Signal Processing) & M.Tech (Remote Sensing & GIS), degrees from JNTUH, Hyderabad India. He has 17 years of teaching experience and has published 7 research papers in International Journals & Conferences. His research interest areas include Deep learning using Digital Image Processing, Video processing, VLSI, and IoT.