International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)
Volume 3, Issue 11, November 2014

A Review on clustering techniques in Continuous Speech Recognition

Kalamani M[1], Valarmathy S[2], Mohan R[3], Anitha S [4]

Abstract— The speech recognition is the most important research area to recognize the speech signal by the compute . To develop the recognition of the continuous speech signal, a speech segmentation, feature extraction (MFCC) and clustering techniques are used as the front end approach. the continuous speech signal into identifiable and meaning full units such as syllables, phonemes words and sub words. The clustering approach ie., k means clustering and Fuzzy c means clustering is the collection of extracted features into different clusters based on similar sense. The Hidden morkov model (HMM) and Gaussian mixture model(GMM) are the most suitable acoustic models are used to train the continuous speech signal and recognize the corresponding text data.

Index Terms— Hidden Markov Model (HMM), Gaussian Mixture Model, Mel Frequency Cepstral Coefficients (MFCC), k means and Fuzzy c means (FCM) clustering.

1. INTRODUCTION

1.1. Overview of CSR

Speech Recognition (also known as continuous Speech Recognition (CSR) or computer speech recognition) is the process of converting a speech signal to a sequence of words or text which is shown in Fig.1

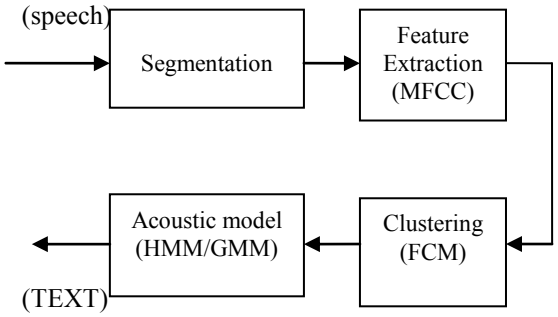

Fig 1:Block Diagram of CSR system

continuous speech recognition systems are widespread application in tasks that require human machine interface, such as voice dialing and call routing, query based information systems that provide updated travel information, simple data entry, voice dictation, access to information: travel, banking, Avionics, Automobile portal, speech transcription, Handicapped people (blind people) supermarket, railway reservations etc.

Speech segmentation is the signal processing front-end that segments continuous speech into uniquely identifiable or meaningful units as phonemes, syllables, words or sub-word and processes them to generate distinguishable features. It reduces the memory size of the input signal(speech) and minimize the computation time complexity large sequence of speech signal.

The next step is the Feature Extraction method, The segmented speech signal is parameterized by MFCC. The goal is to extract a number of parameters from the signal that has a maximum of information relevant for the classification. The number of features extracted from the waveform signal is commonly much lower than the number of signal samples, thus reducing the amount of data. The choice of suitable features varies depending on the classification technique. Typically, The Mel Frequency Cepstral Coefficients (MFCC) Extraction method is most suitable for the continuous speech signal recognition process. The extracted features of frames will be grouped using clustering techniques i.e., here we use two types of clustering methods (K means clustering and Fuzzy c means clustering) The clustered data will be handed to the classifiers i.e., acoustic models such as HMM ,GMM, KNN, ELM classifiers ,etc., That will produce the text output.

1.2 Overview of clustering techniques

Clustering is the process of grouping together similar objects. The resulting groups are called clusters. Clustering algorithms group data points according to various characteristics. Unlike most classification methods, clustering handles data that has no labels. The concept mostly utilizes geometric principles, where the samples are interpreted as points in a d-dimensional Euclidian space, and clustering is made according to the distances between points: usually, points which are close to each other will be allocated to the same cluster. This method is an unsupervised process since there are no predefined classes and no examples that would indicate grouping properties in the data set. The various clustering algorithms are based on some assumptions in order to a partitioning of a dataset. Most methods behave differently depending on the features of the dataset.

ISSN: 2278 – 909X

As compare to data classification, data clustering is considered as an unsupervised learning process which does not require any labeled dataset as training data and the performance of data clustering algorithm is generally considered as much poorer. Although data classification is better performance oriented but it requires a labeled dataset as training data and practically classification of labeled data is generally very difficult as well as expensive. As such there are many algorithms that are proposed to improve the clustering performance.

The hierarchical, partitioning and mixture model methods are the three major types of clustering processes that are applied for organizing data. There are different clustering techniques are available for the clustering of data. In this paper, two main clustering techniques are discussed which are most suitable for the continuous speech recognition process.

2. Clustering techniques in CSR

Clustering is basically considered as classification of similar objects or in other words, it is precisely partitioning of datasets into clusters so that data in each cluster shares some common trait. There are two main clustering techniques are implemented in the continuous speech recognition syetem.ie.,

- k-means clustering algorithm.
- Fuzzy c-means clustering algorithm.

2.1 k-means clustering Algorithm

K-means clustering or Hard C-Means clustering is an elementary but very popular approximate method that can be used to simplify and accelerate convergence. Its goal is to find K mean vectors which will be the K cluster centroids.

This partitioning method applied to analyze data and treats observations of the data as objects based on locations and distance between various input data points. Partitioning the objects into mutually exclusive clusters (K) is done by it in such a fashion that objects within each cluster remain as close as possible to each other but as far as possible from objects in other clusters. Each cluster is characterized by its centre point i.e. centroid.

The distances used in clustering in most of the times do not actually represent the spatial distances. In general, the only solution to the problem of finding global minimum is exhaustive choice of starting points. But use of several replicates with random starting point leads to a solution i.e. a global solution. In a dataset, a desired number of clusters K and a set of k initial starting points, the K-Means clustering algorithm finds the desired number of distinct clusters and their centroids. A centroid is the point whose co-ordinates are obtained by means of computing the average of each of the co-ordinates of the points of samples assigned to the clusters. Finally, this algorithm aims at minimizing an *objective function*, in this case a squared error function. The objective function

$$J(U,V) = \sum_{i=1}^{k}\sum_{j=1}^{n} \left\| x^{(j)} - c_j \right\|^2 \qquad (1)$$

2.1.1 Algorithm

The process of the K-means clustering algorithm is as follows:

Step 1: *Set K* - Randomly select k objects,

Step 2: *Initialization*- each object expresses an initial cluster center.

step 3: *Classification* – To examine each point in the dataset and assign it to the cluster whose centroid is nearest to it.

step 4 : *Centroid calculation* – When each point in the data set is assigned to a cluster, it is needed to recalculate the new k centroids using:

$$v_i = (1/c_i)\sum_{j=1}^{c_i} x_i \qquad (2)$$

step 5: *Convergence criteria* – The steps of (iii) and (iv) require to be repeated until no point changes its cluster assignment or until the centroids no longer move.

The input parameters of the clustering algorithm are the number of clusters that are to be found along with the initial starting point values. When the initial starting values are given, the distance from each sample data point to each initial starting value is found using equation. Then each data point is placed in the cluster associated with the nearest starting point. After all the data points are assigned to a cluster, the new cluster centroids are calculated. For each factor in each cluster, the new centroid value is then calculated. The new centroids are then considered as the new initial starting values and steps (iii) and (iv) of the algorithm are repeated. This process continues until no more data point changes or until the centroids no longer move.

2.2 Fuzzy c-means clustering Algorithm

Fuzzy c-means (FCM) clustering is one of unsupervised clustering techniques and is based on the minimization of an objective function called c-means functional. The FCM employs fuzzy partitioning such that a data point can belong to all groups with different membership grades between 0 and 1. FCM is an iterative algorithm to find cluster centers (centroids) that minimize a dissimilarity function, which is different from HKM(hard k means) that employs hard partitioning.

Fuzzy clustering is more natural than hard clustering in such cases since it allows pattern(s) to belong to multiple clusters simultaneously with different degrees of membership. This clustering is achieved by iteratively minimizing an objective function that is depend on the distance of the pattern to the cluster prototypes in the input space.

2.2.1Algorithm:

In FCM the Clustering is thus achieved by iteratively minimizing an objective function of Equation (3), which is dependent on the distance of the pattern to the cluster prototypes in the input space.

ISSN: 2278 – 909X

$$J(U,V) = \sum_{i=1}^{n}\sum_{j=1}^{c}(\mu_{ij})^{m}\left\|x_i - v_j\right\|^2 \qquad (3)$$

step 1: Randomly select *'c'* cluster centers.

step 2: Calculate the fuzzy membership $'\mu_{ij}'$ using:

$$\mu_{ij} = 1/\sum_{k=1}^{c}(d_{ij}/d_{ik})^{(2/m-1)} \qquad (4)$$

step 3: Compute the fuzzy centers $'v_j'$ using:

$$v_j = (\sum_{i=1}^{n}(\mu_{ij})^{m}x_i)/(\sum_{i=1}^{n}(\mu_{ij})^{m}), \forall_j = 1,2,....c \qquad (5)$$

step 4: Repeat step 2 and 3 until the minimum *'J'* value is achieved or $||U^{(k+1)} - U^{(k)}|| < \beta$.

where,

'k' is the iteration step

'β' is the termination criterion between [0,1].

'U = $(\mu_{ij})_{n*c}$' is the fuzzy membership matrix.

'J' is the objective function.

3. CONCLUSION

The k means clustering algorithm is only suited for the distinct or well separated data set and it requires a priori specification of the number of cluster centers. The k means is not suitable for the nonlinear data. But Fuzzy c means clustering gives the best result for overlapped data set and comparatively better than k-means algorithm. Unlike k-means where data point must exclusively belong to one cluster center here data point is assigned membership to each cluster center as a result of which data point may belong to more than one cluster center. Future work is to implement the Fuzzy c means clustering for continuous speech recognition system.

4. REFERENCES

[1]. Wang Xianbao1, Chen Yong, Tang Liping, "Speech recognition research based on MFCC analysis and biomimetic pattern recognition",Computer Engineering and Applications, vol.47 No.12, 2011, pp.20-22. reconstruction," IEEE Trans. Audio, Speech, Lang. Process., vol. 15, no. 1, pp. 24–33, Jan. 2007.

[2]. L. R. Rabiner and B. H. Juang. Fundamentals of Speech Recognition. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[3]. A. Rakhlin and A. Caponnetto, "Stability of K-Means clustering",Advances in Neural Information Processing Systems, MIT Press, Cambridge, MA, 2007, pp. 216–222.

[4]. A. Rakhlin and A. Caponnetto, "Stability of K-Means clustering",Advances in Neural Information Processing Systems, MIT Press, Cambridge, MA, 2007,pp.216–222.

[5]. Li K, Chang S L, "Study of Ensemble Method of Classifiers for Neural Networks Based on K-Means Clustering", Granular computing, 2008.Grc 2008.IEEE International Conference on Digital Object Identifier pp.375-378, 2008.

[6]. OkkoRasanen, "Speech Segmentation and Clustering Methods for a New Speech Recognition Architecture", M.Sc Thesis, Department of Electrical and Communications Engineering, Laboratory ofAcoustics and Audio Signal Processing, Helsinki University of Technology,Espoo, November 2007.

[7]. Knill K and and Young S, "Hidden Markov Models in Speech and Language Processing", Kluwer Academic Publishers, pp. 27-68, 1997.

[8]. Juang B H and Rabiner L R, "Automatic Speech Recognition – A Brief History of The Technology Development", Elsevier Encyclopedia of Languageand Linguistics, Second Edition, 2005.

[9]. Liu Z, Wang Y (1998) Audio feature extraction and analysis for scene segmentation and classification. J VLSI Sign Process 20:61–79.

[10]. Luong HV and Kim J-M. (2009) A Generalized spatial fuzzy C-means algorithm for medical image segmentation. IEEE Int. Conf. on Fuzzy Systems, 409–414 .

[11]. Park DC, Nguyen DH, Beack SH, Park S (2005) Classification of audio signals using Gradient-based fuzzy c-means algorithm with divergence measure. Adv Multimedia Inf Process PCM 2005:698–708.

[12]. Jiang, H., Li, X., Liu, C.: 'Large margin hidden Markov models for speech recognition', IEEE Trans.Audio, Speech Lang. Process., 2006, 14,(5), pp. 1584–1595.

[13]. X. H. Wu, J. J. Zhou. Alternative Possibilistic Fuzzy c-Means Clustering Algorithm [J]. Journal of Computational Information Systems, 2006, 2(3):925-931.

[14]. Andrew Skabar and KhaledAbdalgader "Clustering Sentence Level Text Using a Novel Fuzzy Relational Clustering Algorithm," IEEE Transactions on Knowledge and Data Engineering, Vol. 25, No. 1, pp. 62-75, Jan. 2013.

[15]. Yogeshjain, AmitkumarNandanwar" A Fuzzy C Means Clustering Based Approach for TeamAssignment" IJRIT International Journal of Research in Information Technology, Volume 2, Issue 5, May 2014, Pg: 639-644

Author Details:

[1]Assistant Professor(Sr.G), Dept.of ECE, Bannari Amman Institute Of Technology.

[2] Professor & Head, Dept.of ECE, Bannari Amman Institute Of Technology.

[3]M.E. Applied Electronics, Bannari Amman Institute Of Technology.

[4]M.E. Communication Systems, Bannari Amman Institute Of Technology.

ISSN: 2278 – 909X