Computer Engineering and Intelligent Systems
ISSN 2222-1719 (Paper) ISSN 2222-2863 (Online)
Vol 3, No.7, 2012
www.iiste.org

Speech Recognition Using Vector Quantization through Modified K-meansLBG Algorithm

Balwant A. Sonkamble[1*] D. D. Doye[2]

1. Pune Institute of Computer Technology, Dhankawadi, Pune, India
2. Proefessor, Electronics and Telecommunication Department, SGGSIE&T, Nanded, India

*sonbalwant@yahoo.com

Abstract

In the Vector Quantization, the main task is to generate a good codebook. The distortion measure between the original pattern and the reconstructed pattern should be minimum. In this paper, a proposed algorithm called Modified K-meansLBG algorithm used to obtain a good codebook. The system has shown good performance on limited vocabulary tasks.

Keywords: K-means algorithm, LBG algorithm, Vector Quantization, Speech Recognition

1. Introduction

The natural way of communication among human beings is through speech. Many human beings are exchanging the information through mobile phones as well as other communication tools in a real manner [L. R. Rabiner et al., 1993]. The Vector Quantization (VQ) is the fundamental and most successful technique used in speech coding, image coding, speech recognition, and speech synthesis and speaker recognition [S. Furui, 1986]. These techniques are applied firstly in the analysis of speech where the mapping of large vector space into a finite number of regions in that space. The VQ techniques are commonly applied to develop discrete or semi-continuous HMM based speech recognition system.

In VQ, an ordered set of signal samples or parameters can be efficiently coded by matching the input vector to a similar pattern or codevector (codeword) in a predefined codebook [[Tzu-Chuen Lu et al., 2010].

The VQ techniques are also known as data clustering methods in various disciplines. It is an unsupervised learning procedure widely used in many applications. The data clustering methods are classified as hard and soft clustering methods. These are centroid-based parametric clustering techniques based on a large class of distortion functions known as Bregman divergences [Arindam Banerjee et al., 2005].

In the hard clustering, each data point belongs to exactly one of the partitions in obtaining the disjoint partitioning of the data whereas each data point has a certain probability of belonging to each of the partitions in soft clustering. The parametric clustering algorithms are very popular due to its simplicity and scalability. The hard clustering algorithms are based on the iterative relocation schemes. The classical K-means algorithm is based on Euclidean distance and the Linde-Buzo-Gray (LBG) algorithm is based on the Itakura-Saito distance. The performance of vector quantization techniques depends on the existence of a good codebook of representative vectors.

In this paper, an efficient VQ codebook design algorithm is proposed known as Modified K-meansLBG algorithm. This algorithm provides superior performance as compared to classical K-means algorithm and the LBG algorithm. Section-2 describes the theoretical details of VQ. Section-3 elaborates LBG algorithm. Section-4 explains classical K-means algorithm. Section -5 emphasizes proposed modified K-meansLBG algorithm. The experimental work and results are discussed in Section-6 and the concluding remarks made at the end of the paper.

2. Vector Quantization

The main objective of data compression is to reduce the bit rate for transmission or data storage while maintaining the necessary fidelity of the data. The feature vector may represent a number of different possible speech coding parameters including linear predictive coding (LPC) coefficients, cepstrum coefficients. The VQ can be considered as a generalization of scalar quantization to the quantization of a vector. The VQ encoder encodes a given set of k-dimensional data vectors with a much smaller subset. The subset C is called a codebook and its elements C_i are called codewords, codevectors, reproducing vectors, prototypes or design samples. Only the index i is transmitted to the decoder. The decoder has the same codebook as the encoder, and decoding is operated by table look-up procedure. The commonly used vector quantizers are based on nearest neighbor called Voronoi or nearest neighbour vector quantizer. Both the classical K-means algorithm and the LBG algorithm belong to the class of nearest neighbor quantizers.

A key component of pattern matching is the measurement of dissimilarity between two feature vectors. The measurement of dissimilarity satisfies three metric properties such as Positive definiteness property, Symmetry property and Triangular inequality property. Each metric has three main characteristics such as computational complexity, analytical tractability and feature evaluation reliability. The metrics used in speech processing are derived from the Minkowski metric [J. S. Pan et al. 1996]. The Minkowski metric can be expressed as

$$D_p(X, Y) = \sqrt[p]{\sum_{i=1}^{k} \left| x^i - y^i \right|^p},$$

Where $X = \{x^1, x^2, \ldots, x^k\}$ and $Y = \{y^1, y^2, \ldots, y^k\}$ are vectors and p is the order of the metric.

The City block metric, Euclidean metric and Manhattan metric are the special cases of Minkowski metric. These metrics are very essential in the distortion measure computation functions.

The distortion measure is one which satisfies only the positive definiteness property of the measurement of dissimilarity. There were many kinds of distortion measures including Euclidean distance, the Itakura distortion measure and the likelihood distortion measure, and so on.

The Euclidean metric [Tzu-Chuen Lu et al., 2010] is commonly used because it fits the physical meaning of distance or distortion. In some applications division calculations are not required. To avoid calculating the divisions, the squared Euclidean metric is employed instead of the Euclidean metric in pattern matching.

The quadratic metric [Marcel R. Ackermann et al., 2010] is an important generalization of the Euclidean metric. The weighted cepstral distortion measure is a kind of quadratec metric. The weighted cepstral distortion key feature is that it equalizes the importance in each dimension of cepstrum coefficients. In the speech recognition, the weighted cepstral distortion can be used to equalize the performance of the recognizer across different talkers. The Itakura-Saito distortion [Arindam Banerjee et al., 2005] measure computes a distortion between two input vectors by using their spectral densities.

The performance of the vector quantizer can be evaluated by a distortion measure D which is a non-negative cost $D(X_j, \hat{X}_j)$ associated with quantizing any input vector X_j with a reproduction vector $\hat{X}_j$. Usually, the Euclidean distortion measure is used. The performance of a quantizer is always qualified by an average distortion $D_v = E[D(X_j, \hat{X}_j)]$ between the input vectors and the final reproduction vectors, where E represents the expectation operator. Normally, the performance of the quantizer will be good if the average distortion is small.

Another important factor in VQ is the codeword search problem. As the vector dimension increases accordingly the search complexity increases exponentially, this is a major limitation of VQ codeword search. It limits the fidelity of coding for real time transmission.

A full search algorithm is applied in VQ encoding and recognition. It is a time consuming process when the codebook size is large.

In the codeword search problem, assigning one codeword to the test vector means the smallest distortion between the codeword and the test vector among all codewords. Given one codeword C_t and the test vector X in the k-dimensional space, the distortion of the squared Euclidean metric can be expressed as follows:

$D(X, C_t) = \sum_{i=1}^{k} (x^i - c_t^i)^2$, where $C_t = \{c_t^1, c_t^2, ..., c_t^k\}$ and $X = \{x^1, x^2, ..., x^k\}$.

There are three ways of generating and designing a good codebook namely the random method, the pair-wise nearest neighbor clustering and the splitting method. A wide variety of distortion functions, such as squared Euclidean distance, Mahalanobis distance, Itakura-Saito distance and relative entropy have been used for clustering. There are three major procedures in VQ, namely codebook generation, encoding procedure and decoding procedure. The LBG algorithm is an efficient VQ clustering algorithm. This algorithm is based either on a known probabilistic model or on a long training sequence of data.

3. Linde–Buzo–Gray (LBG) algorithm

The LBG algorithm is also known as the Generalised Lloyd algorithm (GLA). It is an easy and rapid algorithm used as an iterative nonvariational technique for designing the scalar quantizer. It is a vector quantization algorithm to derive a good codebook by finding the centroids of partitioned sets and the minimum distortion partitions. In LBG, the initial centroids are generated from all of the training data by applying the splitting procedure. All the training vectors are incorporated to the training procedure at each iteration. The GLA algorithm is applied to generate the centroids and the centroids cannot change with time. The GLA algorithm starts from one cluster and then separates this cluster to two clusters, four clusters, and so on until N clusters are generated, where N is the desired number of clusters or codebook size. Therefore, the GLA algorithm is a divisive clustering approach. The classification at each stage uses the full-search algorithm to find the nearest centroid to each vector. The LBG is a local optimization procedure and solved through various approaches such as directed search binary-splitting, mean-distance-ordered partial codebook search [Linde et al., 1980, Modha et al., 2003], enhance LBG, GA-based algorithm [Tzu-Chuen Lu et al., 2010, Chin-Chen Chang et al. 2006], evolution-based tabu search approach [Shih-Ming Pan et al., 2007], and codebook generation algorithm [Buzo et al., 1980].

In speech processing, vector quantization is used for instance of bit stream reduction in coding or in the tasks based on HMM. Initialization is an important step in the codebook estimation. Two approaches used for initialization are Random initialization, where L vectors are randomly chosen from the training vector set and Initialization from a smaller coding book by splitting the chosen vectors.

The detailed LBG algorithm using unknown distribution is described as given below:

Step 1: Design a 1-vector codebook.

Set $m=1$. Calculate centroid

$$C_1 = \frac{1}{T}\sum_{j=1}^{T} X_j \quad .$$

Where T is the total number of data vectors.

Step 2: Double the size of the codebook by splitting.

Divide each centroid C_i into two close vectors $C_{2i-1} = C_i \times (1+\delta)$ and $C_{2i} = C_i \times (1-\delta), \ 1 \le i \le m$. Here δ is a small fixed perturbation scalar.

Let $m = 2m$. Set $n = 0$, here n is the iterative time.

Step 3: Nearest-Neighbor Search.

Find the nearest neighbor to each data vector. Put X_j in the partitioned set P_i if C_i is the nearest neighbor to X_j.

Step 4: Find Average Distortion.

After obtaining the partitioned sets

$P = (P_i\ ,\ 1 \le i \le m)$, Set $n = n + 1$.

Calculate the overall average distortion

$$D_n = \frac{1}{T}\sum_{i=1}^{m} \sum_{j=1}^{T_i} (D_j^{(i)}, C_i),$$

Where $P_i = \{X_1^{(i)}, X_2^{(i)}, \ldots, X_{T_i}^{(i)}\}$.

Step 5: Centroid Update.

Find centroids of all disjoint partitioned sets P_i by

$$C_i = \frac{1}{T_i}\sum_{j=1}^{T_i} X_j^{(i)}.$$

Step 6: Iteration 1.

If $(D_{n-1} - D_n)/D_n > \varepsilon$, go to step 3;

otherwise go to step 7 and ε is a threshold.

Step 7: Iteration 2.

If $m = N$, then take the codebook C_i as the final codebook; otherwise, go to step 2.

Here N is the codebook size.

The LBG algorithm has limitations like the quantized space is not optimized at each iteration and the algorithm is very sensitive to initial conditions.

4. Classical K-means Algorithm

The K-means algorithm is proposed by MacQueen in 1967. It is a well known iterative procedure for solving the clustering problems. It is also known as the C-means algorithm or basic ISODATA clustering algorithm. It is an unsupervised learning procedure which classifies the objects automatically based on the criteria that minimum distance to the centroid. In the K-means algorithm, the initial centroids are selected randomly from the training vectors and the training vectors are added to the training procedure one at a time. The training procedure terminates when the last vector is incorporated. The K-means algorithm is used to group data and the groups can change with time. The algorithm can be applied to VQ codebook design. The K-means algorithm can be described as follows:

Step 1: Randomly select N training data vectors as the initial codevectors $C_i, i = 1,2,\ldots,N$ from T

training data vectors.

Step 2: For each training data vector $X_j, j = 1,2,\ldots,T$, assign X_j to the partitioned set S_i

if $i = \arg\min_l D(X_j, C_l)$.

Step 3: Compute the centroid of the partitioned set that is codevector using

$$C_i = \frac{1}{|S_i|}\sum_{X_j \in S_i} X_j$$

Where $|S_i|$ denotes the number of training data vectors in the partitioned set S_i. If there is no change in the clustering centroids, then terminate the program; otherwise, go to step 2.

There are various limitations of K-means algorithm. Firstly, it requires large data to determine the cluster. Secondly, the number of cluster, K, must be determined beforehand. Thirdly, if the number of data is a small it difficult to find real cluster and lastly, as per assumption each attribute has the same weight and it quite difficult to knows which attribute contributes more to the grouping process.

It is an algorithm to classify or to group objects based on attributes/features into K number of group. K is positive integer number. The grouping is done by minimizing the sum of squares of distances between data and the corresponding cluster centroid. The main aim of K-mean clustering is to classify the data. In practice, the number of iterations is generally much less than the number of points.

5. Proposed Modified K-meansLBG Algorithm

The proposed algorithms objective is to overcome the limitations of LBG algorithm and K-means algorithm. The proposed modified KmeansLBG algorithm is the combination of advantages of LBG algorithm and K-means algorithms. The KmeansLBG algorithm is described as given below:

Step 1: Randomly select N training data vectors as the initial codevectors.

Step 2: Calculate the no. of centroids.

Step 3: Double the size of the codebook by splitting.

Step 4: Nearest-Neighbor Search.

Step 5: Find Average Distortion.

Step 6: Update the centroid till there is no change in the clustering centroids, terminate the program otherwise go to step 1.

6. Experimentation and Results

The TI46 database [NIST, 1991] is used for experimentation. There are 16 speakers from them 8 male speakers and 8 female speakers. The numbers of replications are 26 for utterance by each person. The total database size is 4160 utterances of which 1600 samples were used for training and remaining samples are used for testing of 10 words that are numbers in English 1 to 9 and 0 are sampled at a rate of 8000 Hz. A feature vector of 12-dimensional Linear Predicting Coding Cepstrum coefficients was obtained and provided as an input to vector quantization to find codewords for each class.

There are five figures shows comparative graphs of the distortion measure obtained using LBG algorithm and K-means algorithm and proposed K-meansLBG algorithm. The distortion measure obtained by the proposed algorithm is smallest as compared to the K-means algorithm and the LBG algorithm.

The proposed modified KmeanLBG algorithm gives minimum distortion measure as compared to K-means algorithm and LBG algorithm to increase the performance of the system. The smallest measure gives superior performance as compared to both the algorithms as is increased by about 1% to 4 % for every digit.

7. Conclusion

The Vector Quantization techniques are efficiently applied in the development of speech recognition systems. In this paper, the proposed a novel vector quantization algorithm called K-meansLBG algorithm. It is used efficiently to increase the performance of the speech recognition system. The recognition accuracy obtained using K-meansLBG algorithm is better as compared to K-means and LBG algorithm. The average recognition accuracy of K-meansLBG algorithm is more than 2.55% using K-means algorithm while the average recognition accuracy of K-meansLBG algorithm is more than 1.41% using LBG algorithm.

References

L.R. Rabiner , B. H. Juang [1993], "Fundamentals of Speech Recognition", Prentice-Hall, Englewood Cliffs, N.J.

Tzu-Chuen Lu, Ching-Yun Chang [2010], "A Survey of VQ Codebook Generation", Journal of Information Hiding and Multimedia Signal Processing, Ubiquitous International, Vol. 1, No. 3, pp. 190-203.

S. Furui [1986], "Speaker-independent isolated word recognition using dynamic features of speech spectrum", IEEE Transactions on Acoustic, Speech, Signal Processing, Vol. 34, No. 1, pp. 52-59.

S. Furui [1994], "An overview of speaker recognition technology", ESCA Workshop on Automatic Speaker Recognition, Identification and Verification, pp. 1-9.

F. K. Soong, A. E. Rosenberg, B. H. Juang [1987], "A vector quantization approach to speaker recognition", AT&T Technical Journal, Vol. 66, No. 2, pp. 14-26.

Wiploa J. G. , Rabiner L. R. [1985], "A Modified K-Means Clustering Algorithm for Use in Isolated Word Recognition", IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 33 No.3, pp. 587-594.

Arindam Banerjee, Srujana Merugu, Inderjit S. Dhillon, Joydeep Ghosh [2005], "Clustering with Bregman Divergences", Journal of Machine Learning Research, Vol. 6, pp. 1705–1749.

J. S. Pan, F. R. McInnes and M. A. Jack [1996], "Bound for Minkowski metric or quadratic metric applied to codeword search", IEE-Proceedings on Vision Image and Signal Processing, Vol. 143, No. 1, pp. 67–71.

Marcel R. Ackermann, Johannes Blomer, Christian Sohler [2010], "Clustering for Metric and Nonmetric Distance Measures", ACM Transactions on Algorithms, Vol. 6, No. 4, Article 59, pp. 1-26.

Shih-Ming Pan, Kuo-Sheng Cheng [2007], "An evolution-based tabu search approach to codebook design", Pattern Recognition, Vol. 40, No. 2, pp. 476-491.

Chin-Chen Chang, Yu-Chiang Li, Jun-Bin Yeh [2006], "Fast codebook search algorithms based on tree-structured vector quantization", Pattern Recognition Letters, Vol. 27, No. 10, pp. 1077-1086.

A. Buzo, A. H. Gray, R. M. Gray, J. D. Markel [1980], "Speech coding based upon vector quantization", IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 28, No. 5, pp. 562–574.

Y. Linde, A. Buzo, and R. M. Gray [1980], "An algorithm for vector quantizer design", IEEE Transactions on Communications, Vol. 28, No.1, pp. 84–95.

D. Modha , S. Spangler [2003], "Feature weighting in k-means clustering. Machine Learning", Vol. 52, No.3, pp. 217–237.

NIST, [1991], "TI46 CD".

Balwant A. Sonkamble received his BE (Computer science and Engineering in 1994 and M. E. (Computer Engineering) in 2004. Currently he is research scholar at SGGS College of Engineering and Technology, Vishnupuri, Nanded (MS)–INDIA. He is working as Associate Professor in Computer Engineering at Pune Institute of Computer Technology, Pune, India. His research areas are Speech Recognition and Artificial Intelligence.

D. D. Doye received his BE (Electronics) degree in 1988, ME (Electronics) degree in 1993 and Ph. D. in 2003 from SGGS College of Engineering and Technology, Vishnupuri, Nanded (MS) – INDIA. Presently, he is working as Professor in department of Electronics and Telecommunication Engineering, SGGS Institute of Engineering and Technology, Vishnupuri, Nanded. His research fields are speech processing, fuzzy neural networks and image processing.

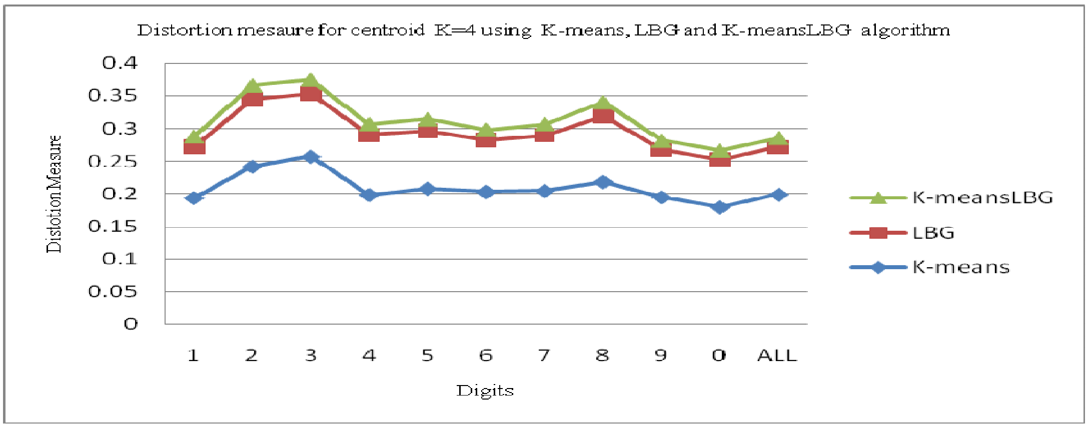

Figure 1. Comparative graph for centroid K=4

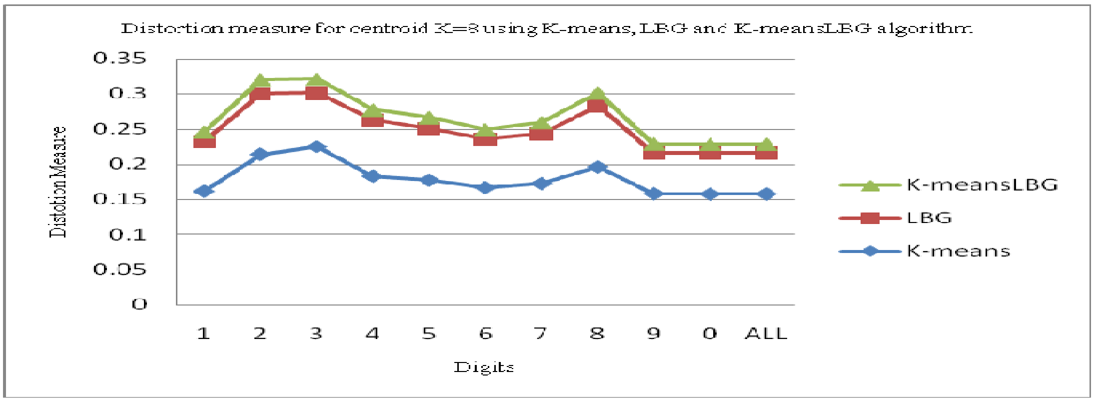

Figure 2. Comparative graph for centroid K=8

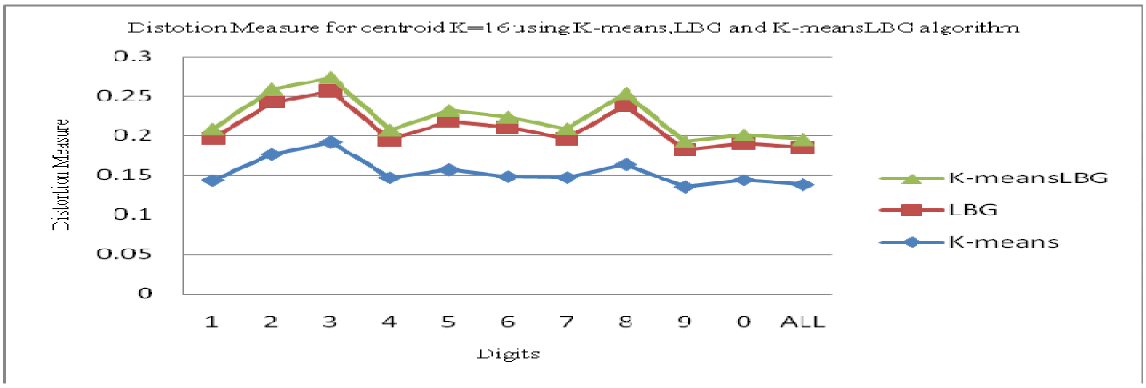

Figure 3. Comparative graph for centroid K=16

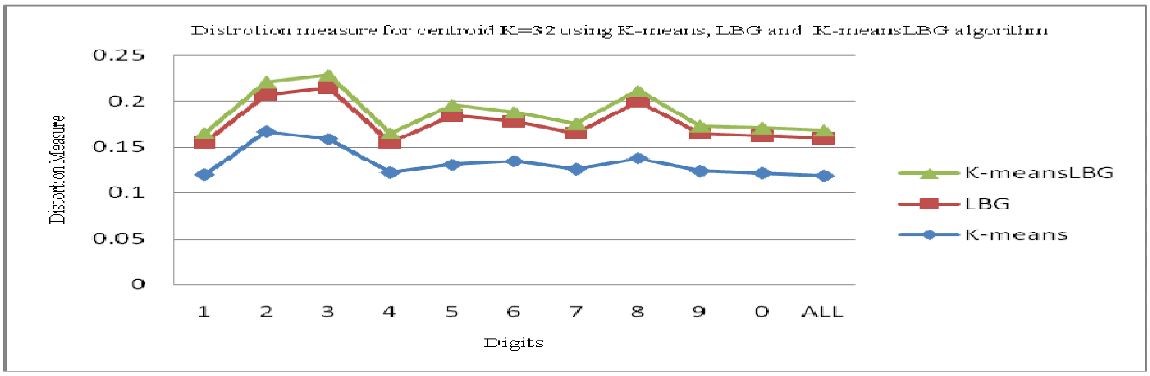

Figure 4. Comparative graph for centroid K=32

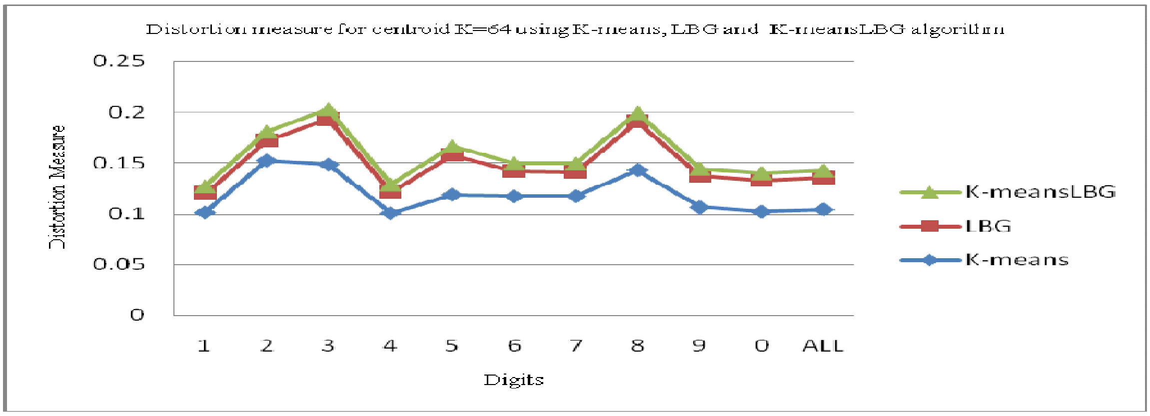

Figure 5. Comparative graph for centroid K=64