IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)
e-ISSN: 2278-2834, p-ISSN: 2278-8735
PP 63-68
www.iosrjournals.org

Perceptible Study of Some K-Means Clustering Algorithm Used for Image and Data Analysis.

Verma S.V.[1], Jadhav S. N.[2] & M.S.Wadatkar.[3]
[1](E&TC Engg. Dept., MIT COE Aurangabad, BAM Univ., Aurangabad(MS), India)
[2](E&TC Engg. Dept, KVNNIEER Nasik, SPP Univ., Pune(MS), India)
[3](E&TC Engg. Dept, BVCOR & RI Nasik, SPP Univ., Pune(MS), India)

Abstract : *K means is one of the oldest algorithms, which is widely used for clustering of images as well as data. Clustering makes the analysis much simpler and easier. K means is used in many applications and many study related programs. But in this paper we are going to compare couple of improved K means clustering algorithms. Here we are going to compare Modified K Means (MKM) with Adaptive Fuzzy K Means (AFKM) clustering algorithm for image as well as data. As we know K means was mostly used for biomedical, microscopic, satellite images only but here MKM and AFKM can be used for normal image which is taken from CCD, mobile camera, and digital cameras. Both the above mentioned algorithms are based on conventional K means technique. Here we are also using percentage scored in engineering students for data analysis.*
Keywords - *Adaptive Fuzzy K Means Clustering Algorithm, Clustering, Image and Data Analysis, Modified K Means Clustering .*

1. INTRODUCTION

Formation of cluster means the pixels having similar parameters are placed in one cluster. This is widely used in statistics [4], [5], machine learning [6]-[8], pattern recognition [9]-[11], data mining [12]-[17], and image processing [18], [19]. In digital image processing, segmentation is essential for image description and classification. The technique is commonly used by many consumer electronic products (i.e., conventional digital image) or in a specific application field such as the medical digital image [2].

Here in this paper we have compared modified K means & adaptive fuzzy K means clustering algorithm.

2. THE MODIFIED K-MEANS ALGORITHM.

Modified K means clustering technique is based on normal K means algorithm which produces empty clusters during illustration. These empty clusters are created because of importer start of iteration [1]. Therefore we use modified K means clustering algorithm which does not produce empty cluster and give better results than conventional K means algorithm. This new improved K means algorithm is called as Modified K Means (MKM) algorithm[1],[2].

As in conventional K means the iteration is begins from old center which is denoted by zk (old), the pixels are categorized taking into consideration the minimum Euclidean distance. Then a new cluster is formed which is denoted by zk (new). Now again new center is calculated by:

$$\mathbf{z}_k^{(new)} \leftarrow \frac{1}{n_k}\left\{\sum_{x_j \in C_k} (\mathbf{x}_j)\right\} \tag{1}$$

Where, nk is number of pixels in Ck cluster. Here in MKM the uses different technique to calculate center than conventional K means. As said earlier MKM algorithm is also widely used as it does not create empty cluster. Therefore the center updating in MKM is done by:

$$\mathbf{z}_k^{(new)} \leftarrow \frac{1}{n_k+1}\left\{\sum_{x_j \in C_k} (\mathbf{x}_j) + \mathbf{z}_k^{(old)}\right\} \tag{2}$$

Therefore the above equation is used for updating of new cluster and also every cluster contains at least one pixel. By using this new way of calculating cluster center will not affect the algorithm in its convergence. MKM also convergences after within few iterations without creating any empty cluster.

The MKM Algorithm
Input: a set D of d-dimensional data and an integer K.

Output: K clusters
begin
randomly pick K points €¸D to be initial means;
while measure M is not stable do
begin
compute distance $d_{kj} = ||xj\text{-}zk||^2$ for each
k, j where $1 \leq k \leq K$ and $1 \leq j \leq N$, and
determine members of new K subsets based
upon minimum distance to z_k for $1 \leq k \leq K$;
compute new center z_k for $1 \leq k \leq K$ using (2);
compute M;
end
end

3. THE ADAPTIVE FUZZY K-MEANS ALGORITHM

Adaptive fuzzy K means algorithm is combination of conventional K means & MKM clustering techniques. In AFKM each pixel belongs to more than one cluster and every pixel is assigned to a cluster depending upon minimum Euclidean distance. Consider a digital image with R×S pixels (i.e., R represents number of columns and S represents number of rows) to be clustered into nc regions or clusters. Let p(x,y) be the considered pixel and cj as the j-th centre, where x = 1, 2, ..., R, y=1, 2, ..., S and j = 1, 2, …., nc.

Now let us go through the steps used in Adaptive Fuzzy K-Means Clustering Algorithm:

1.Select the value of K randomly. Where, K indicates the number of clusters to be formed.
2) Assign every pixel of the image to a particular cluster.
3) Find the distance between each pixel and cluster center. Then minimize the distance between pixel and cluster center.
4) Again calculate new centers for every cluster by averaging all of the pixels in that particular cluster.
5) Follow second & third step till the algorithm convergence i.e not a single pixel changes its position from one cluster to other and also the cluster centers does not change its position.

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} || x_i^{(j)} - c_j ||^2 \tag{3}$$

where $X_i^{(j)}$ - no of pixels
C_j - cluster center.
$||X_i^{(j)} - C_j||^2$ - distance between pixels and cluster center.
K - number of clusters.

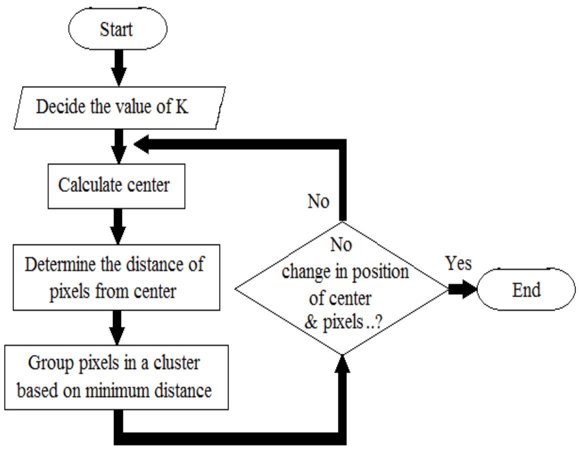


Fig.1 Flowchart of AFKM algorithm.

4. STIMULATION RESULTS

For stimulation purpose we have considered two types that is data and image. The processing time analysis is also performed to measure the algorithm efficiency. The analysis is performed using *Intel® Core™ i3 CPU 3.07GHz Processor, 2.00GB* of *RAM*, and *500GB* of disk drive space. The coding is done using MATLAB. For data we have collected students percentage from one of the engineering college affiliated to Pune University. We considered percentage of 100 students and we divided them into five cluster (K) or groups on the basis of the percentage. And considering K=5 we have calculated the cluster accuracy rate (CAR) for each cluster as well as the time required for convergence of algorithm and for stimulation of image we have randomly considered two images which are taken using normal digital camera. Then we decided the number of iterations after which the algorithm should converge. And also the number of clusters to be formed. As we all know that there is a disadvantage of K means algorithm that we have to specify the number of clusters in advance. Then we again calculated CAR and the time required for convergence of algorithm.

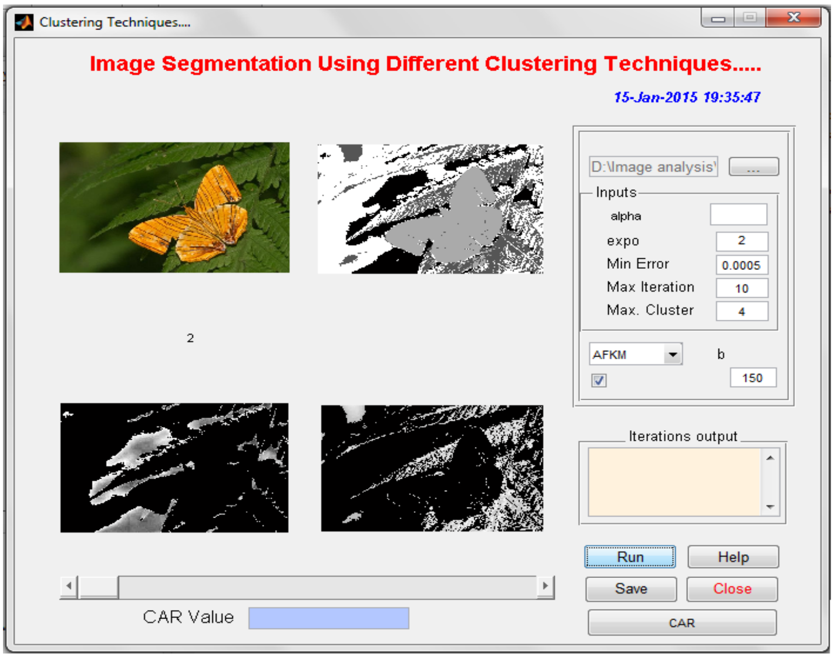


Fig. 2 GUI for Image analysis.

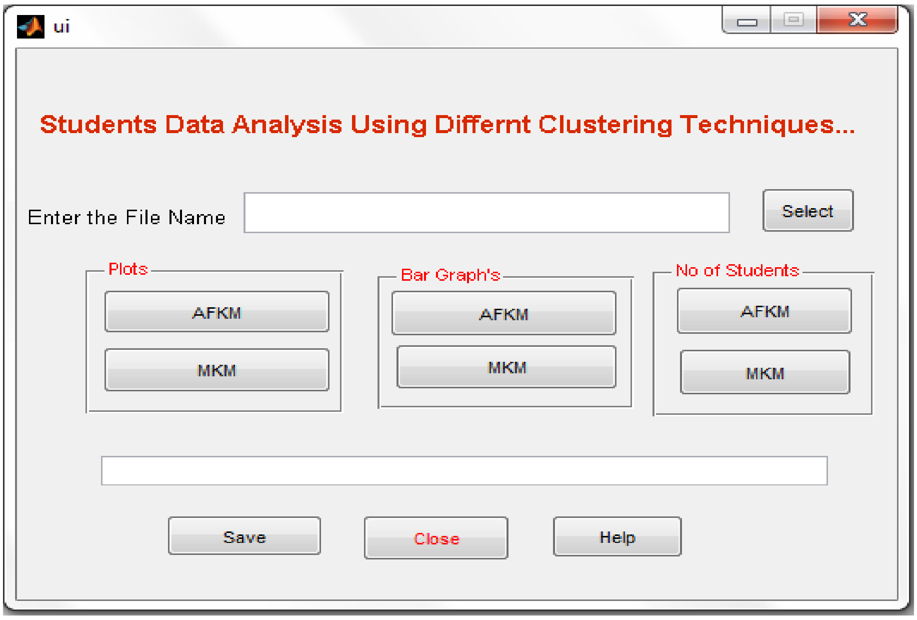


Fig. 3 GUI for data analysis.

Table. 1 Stimulation result of students data.

Input Data		CAR	
Range	**No of Students**	**AFKM**	**MKM**
0 to 20	5	0.97	0.96
20 to 40	21	0.95	0.84
40 to 60	53	0.9	0.89
60 to 80	15	0.98	0.92
80 to 100	6	1	0.97
Total	**100**		

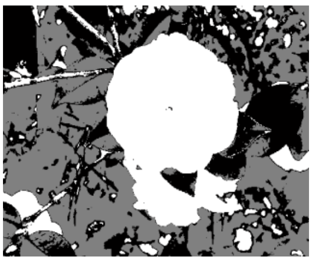
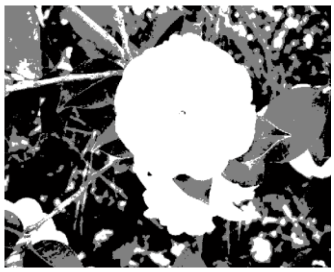
<table>
<tr><th colspan="2" rowspan="2">Input Image</th><th colspan="2">Output of clustering</th></tr>
<tr><th>MKM</th><th>AFKM</th></tr>
<tr><td colspan="2"></td><td></td><td></td></tr>
<tr><td>No of Iterations = 15</td><td>CAR →</td><td>82.7342</td><td>97.0077</td></tr>
<tr><td>No of Clusters = 3</td><td>Time req→</td><td>12.39</td><td>12.70</td></tr>
</table>

Fig. 4 Stimulation result of flower image.

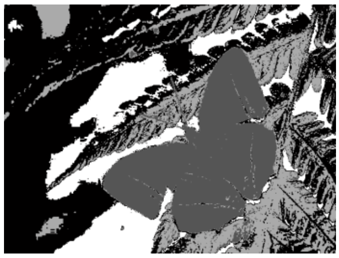
Input Image		Output of clustering	
		MKM	**AFKM**
			
No of Iterations = 10	CAR →	63.9063	82.1868
No of Clusters = 4	Time req→	6.785	6.738

Fig. 5 Stimulation result of butterfly image.

5. CONCLUSION

As we have implemented MKM and AFKM algorithm in a very effective way. Form the results we concluded by using AFKM algorithm for image as well as data we get good CAR and also the time required for convergence of algorithm is less[1]. That's why AFKM is more efficient than Modified K-means clustering algorithm.

So the aim of developing an exact and more consistent image which can be used in finding cracks, face recognition, finger print recognition and in locating an object clearly from a satellite image etc [6]can be successfully done using AFKM. It works effectively when pixels are not properly separated from each other. AFKM also reduces cluster to cluster variance, but it does not make sure that the final result has a global minimum of variance.

REFERENCES

[1] Malay K. Pakhira, A Modified k-means Algorithm to Avoid Empty Clusters, International Journal of Recent Trends in Engineering, vol 1, No. 1, May 2009.
[2] Siti Noraini Sulaiman and Nor Ashidi Mat Isa, Adaptive fuzzy-K-means clustering algorithm for image segmentation, IEEE Transactions on Consumer Electronics, Vol. 56, No. 4, November 2010.
[3] M. Y. Mashor, Hybrid Training Algorithm for RBF Network, International Journal of Computer, Internet and Management, vol. 8, no. 2, pp. 50-65, 2000.
[4] D. Auber, and M. Delest, A clustering algorithm for huge trees, Advances in Applied Mathematics, vol. 31, no. 1 pp. 46-60 2003.
[5] A. S. Mahani, A. E. Carlsson, and R. Wessel, Motion repulsion arises from stimulus statistics when analyzed with a clustering algorithm, Biological Cybernetics, vol. 92, no. 4, pp. 288-291, 2005.
[6] T. Abeel, Y. V. d. Peer, and Y. Saeys, Java-ML: A machine learning library, Journal of Machine Learning Research, vol. 10, pp. 931-934, 2009.
[7] A. Y. Al-Omary, and M. S. Jamil, A new approach of clustering based machine-learning algorithm, Knowledge-Based Systems, vol.19, no.4, pp. 248-258, 2006.
[8] M. J. Rattigan, M. Maier, and D. Jensen. Graph clustering with network structure indices, in Proceedings of the 24th International Conference on Machine Learning. 2007. Corvallis, OR.
[9] P. C. Boutros, and A. B. Okey, Unsupervised pattern recognition: An introduction to the whys and wherefores of clustering microarray data, Briefings In Bioinformatics,vol. 6, no. 4, pp. 331-343, 2005.
[10] J. P. Jesan, The neural approach to pattern recognition, Ubiquity, vol.5, no.7, 2004.
[11] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, An efficient k-means clustering algorithm: analysis and implementation, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 881-892, 2002.
[12] H. Wang, W. Wang, J. Yang, and P. S. Yu, Clustering by pattern similarity in large data sets, in Proceedings of the 2002 ACM SIGMOD International Conference on Management of data. 2002, Madison, Wisconsin.
[13] M. Halkidi, and M. Vazirgiannis, Clustering validity assessment: finding the optimal partitioning of a data set, Proceedings IEEE International Conference on Data Mining. 2001.
[14] Z.-H. Zhou, Three perspectives of data mining, Artificial Intelligence, vol. 143, no. 1, pp. 139-146, 2003.
[15] T. Zhang, R. Ramakrishnan, and M. Livny, BIRCH: an efficient data clustering method for very large databases, in Proceedings of the 1996 ACM SIGMOD international conference on Management of data, 1996, Montreal, Quebec, Canada.
[16] C. Ordonez, Clustering binary data streams with K-means, in Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery. 2003, San Diego, California.

[17] J. Handl, J. Knowles, and D. B. Kell, Computational cluster validation in post-genomic data analysis, Bioinformatics, 2005
[18] S. K. Singh, K. Shishir, G. S. Tomar, K. Ravi, and G. K. A. Santhalia, Modified framework of a clustering algorithm for image processing applications, in First Asia International Conference on Modelling & Simulation, AMS '07, 2007.
[19] F. Hoeppner, Fuzzy shell clustering algorithms in image processing: fuzzy C-rectangular and 2-rectangular shells, IEEE Transactions on Fuzzy Systems,vol. 5, no. 4, pp. 599-613, 1997.