INT J COMPUT COMMUN, ISSN 1841-9836
9(1):56-61, February, 2014.

Improving the Efficiency of Image Clustering using Modified Non Euclidean Distance Measures in Data Mining

P. Santhi, V. Murali Bhaskaran

P.Santhi*
Computer Science Department
Paavai Engineering College, Pachal,Namakkal,India
*Corresponding author: santhipalanisamypec@paavai.edu.in

V.Murali Bhaskaran
Paavai College of Engineering
Pachal, Namakkal, India
murali66@gmail.com

Abstract: The Image is very important for the real world to transfer the messages from any source to destination. So, these images are converted in to useful information using data mining techniques. In existing all the research papers using kmeans and fuzzy k means with euclidean distance for image clustering. Here, each cluster needs its own centric for cluster calculation and the euclidean distance calculate the distance between the points. In clustering this process of distance calculation did not give efficient result. For make this in to efficient, this research paper proposes the non Euclidean distance measures for distance calculation. Here, the logical points are used to find the cluster. The result shows that image clustering based on the modified non Euclidean distance and the performance shows the efficiency of non euclidean distance measures.
Keywords: Data Mining, Image Mining, Kmeans, Fuzzy Kmeans, Euclidean Distance

1 Introduction

In real world, the image plays an important role in the entire field. Normally, all the images having some information related to any application [1] [2]. So, these images are converted in to some useful information using data mining techniques. This process of mining the image is called as image mining. Here, the clustering technique is applied to these images. The process of grouping the similar data or pixels is called as clustering [5].. The data mining having two types of learning is supervised learning and unsupervised learning. In supervised learning, the training dataset was provided to train the system and got the output based on these training data. This type of system is called as classifiers. In classification, group the same item based on the physical behavior. In unsupervised learning, the training data set is not provided to the learning system. In this paper proposes the clustering for the Image. The clustering is one of the techniques in unsupervised learning. The clustering is having many algorithms are partitioned based clustering, Hierarchical based clustering, dense based clustering and distribution based clustering. In partitioned based clustering the cluster based on the centre point [6][4]. Here, the initialization of centre is very difficult and this process takes more time to execute. In hierarchical clustering the clusters are formed by using distance connectivity [3]. In density based clustering the cluster formed using the dense value.
The distance between the points is calculated by using the distance measures. Data mining having the two types of measures called Euclidean distance and non Euclidean distance measure [7]. In euclidean distance measure, the distance is calculated based on the physical points of the cluster. But this measure is not efficient for the clustering to produce the efficient result.

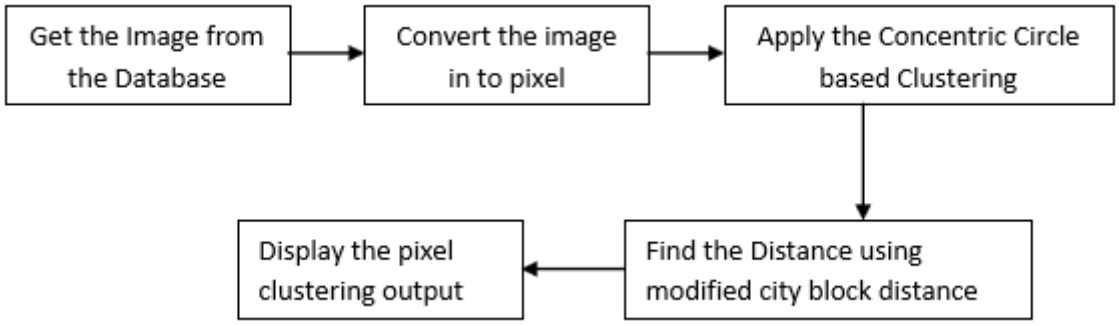

Figure 1: Steps for the proposed system [8]

In non euclidean distance, the distance is calculated using feature vectors between the clusters [4]. In the literature survey most of the researches are done by using kmeans and fuzzy kmeans algorithm for clustering and the city block distance measures for measure the distance between the clusters. In Existing formula for city block distance is [9]

$$D = \sum_{j=1}^{n} [a_j - b_j] \tag{1}$$

In this paper proposes the modified city block distance measures for calculating the distance between the clusters and using concentric circle based clustering algorithm.

2 Research Methodology

Non euclidean distance Measure is very important for the clustering to measure the distance between the points. In this research using the modified non euclidean distance measure of city block distance or manhatten distance. In Existing formula is

$$D = \sum_{j=1}^{n} [a_j - b_j]$$

In this formula is modified in to

$$D = \sum_{j=1}^{n} [a_j - b_k] \tag{2}$$

The equation 2 is called as modified city block or manhatten distance. The steps for this system will be discussed in figure 1.

1. Get the Image.
2. Convert this image into pixel.
3. The concentric circle based clustering algorithm is applied to pixel.
4. Find the distance between the clusters using modified city block distance.
5. Step 4 will be repeated until all the pixels come under any one of the clustering.

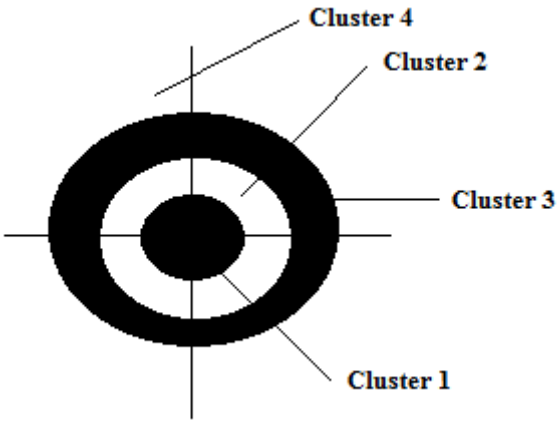

Figure 2: Sample for Concentric Circle Based Clustering

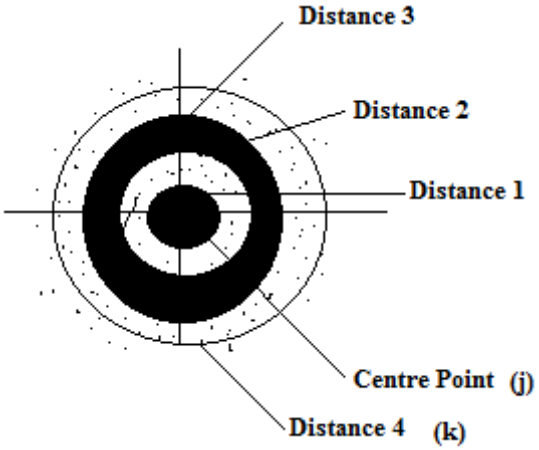

Figure 3: Sample for Pixel Based Clustering

6. Display the output of pixel clustering

In concentric circle based clustering; only one centre is used for all the clusters. This process will reduce the time for image clustering. In existing research, the reduction in the running time is already discussed. Now, this paper discusses the improvement in the efficiency of image clustering. The figure 2 shows the sample of concentric circle based clustering. Here it contains the four groups or clusters using only one centre. The figure shows the pixel clustering using modified city block distance or manhatten distance. Consider the centre assumption is 0.

The formula is

$$D = \sum_{j=1}^{n} [a_j - b_k]$$

- D is the Modified city block or Modified Manhatten distance Measure,
- J is the dimension from 1 to k dimension
- a_j is the centre pixel assumption point

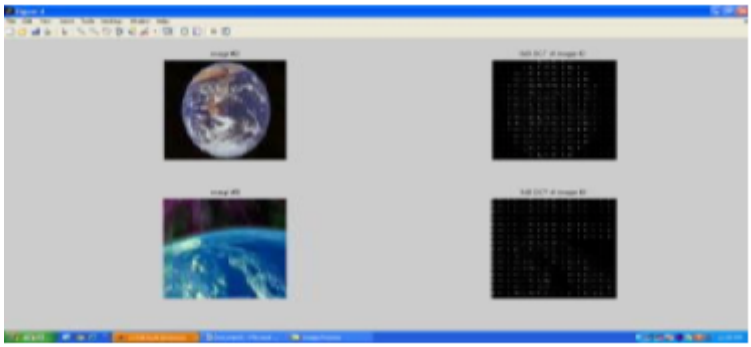

Figure 4: Input Images for Pixel Clustering

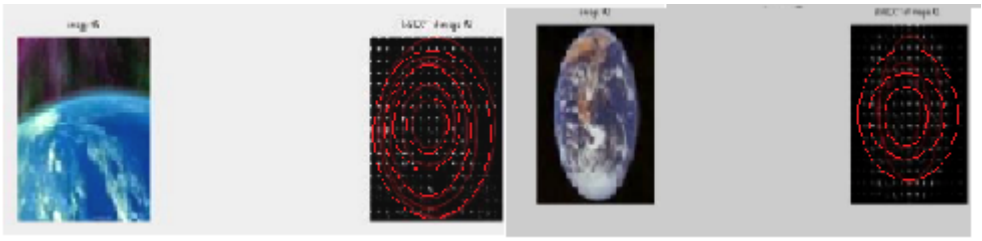

Figure 5: Output images with Pixel Clustering

- b_k is the Top most pixel assumption point

Based on this formula, a_j=0 and the value of b_k=1.Here, the j is the centre value of 0.This value is same for all the clusters. The k value is min distance from that centre. The same manner all the k values are calculated. For the cluster 1, D=0-1=[-1]=1 The cluster 1, the min distance from the centre is 1.So, the pixels within the 1 from X and Y axis are considered as the first cluster. For the cluster 2, D=0-2=[-2]=2, Here a_j=0 and b_k= 2. The cluster 2, the next to the min distance from the centre is 2.So, the pixels within the 2 from X and Y axis are considered as the second cluster. For the cluster 3, D=0-3=[-3]=3, Here a_j=0 and b_k= 3. The cluster 3, the next to the min distance from the centre is 3.So, the pixels within the 3 from X and Y axis are considered as the second cluster. The same way all the clusters are formed.

3 Results and Discussion

The result shows the original image and the equivalent pixel to that image.The figure 5 shows the clustering of pixel using concentric circle based clustering. The image 1 having the 3 clusters and the image 2 shows the 5 clusters.

In this result all the clusters in the image 1 and the image 2 using the single centre point for all the clusters and the distance between the centers is calculated using modified city block distance or manhatten distance.

4 Performance Analysis

The graph shows the performance of clustering algorithm using modified non euclidean distance. In the performance analysis the concentric circle based clustering with modified city block

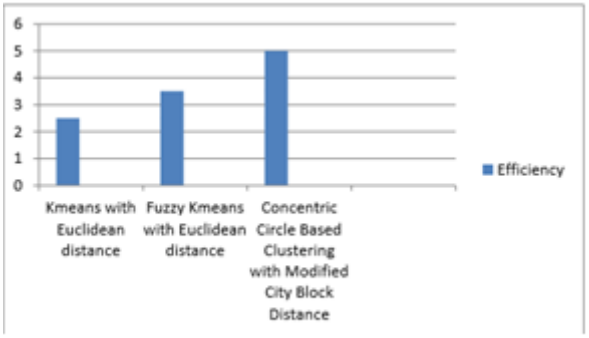

Figure 6: Performance Graph

distance algorithm having more efficiency comparing to existing system.

5 Conclusion and Future Enhancement

The images are very important to the real world in entire field. The images are converted in to useful knowledge using data mining techniques. In existing the researches are did by using k means and Fuzzy k means for clustering with euclidean distance. But it is having more difficulties in assigning centre for each cluster and the efficiency is not high. So, improving the efficiency this research paper gives the image clustering using concentric circle based clustering with non euclidean distance measure of modified city block distance with single centre point. This Result gives the better performance in the accuracy comparing to existing system. In Future, the concentric circle based clustering is combined with soft computing techniques for the image clustering.

Bibliography

[1] Nor Ashidi Mat Isa; Samy Salamah, A, Umi Kalthum Ngah; Adaptive Fuzzy Moving K-means Clustering Algorithm for Image Segmentation, *IEEE Trans on Knowledge and Data Engineering*, 2145-2153, 2009.

[2] Siti Noraini Sulaiman; Nor Ashidi Mat Isa; Adaptive Fuzzy-K-means Clustering Algorithm for Image Segmentation, *IEEE Trans on Knowledge and Data Engineering*, 2661-2668, 2010.

[3] Maduri, A.Tayal;Raghuwanesh,M.M.; Review on various Clustering Methods for Image Data, *J. of Emerging Trends in Computing and Information Sciences*, 34-38, 2010.

[4] Keh-Shih Chuang; Hong-Long Tzeng; Sharon Chen; Jay wu; Tzong-Jer Chen; Fuzzy c-Means Clustering with spatial information for image segmentation,*Elseiver*, 9-15, 2006.

[5] Sneha Silvia, A.; Vamsidhar, Y.; Sudhakar,G.; Color Image Clustering using K-Means,*IJCST*, 11-13, 2011.

[6] Vasuda, P.; Satheesh, S.; Improved Fuzzy C-Means Algorithm for MR Brain Image Segmentation, *IJCSE*, 1713-1715, 2010.

[7] Fahim, A.M.; Saake, G.; Salem, A.M.; Torkey, F.A.; K-Means for Spherical Clusters with Large Variance in Sizes, *World Academy of Science, Engineering and Technology*, 177-182, 2008.

[8] John Peter, S.; Minimum Spanning Tree-based Structural Similarity Clustering for Image Mining with Local Region Outliers,*Int. J. of Computer Applications*, 33-40, 2010.

[9] Chawan, P.M.; Saurabh Bhonde,R.; Shirish Patil; Concentric Circle-Based Image Signature for Near-Duplicate Detection in Large Databases,*Electronics and telecommunication Research Institute*, 2010.