ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Clustering Algorithms with EMD for Microarray Image Segmentation

K.Ranga Narayana
M.Tech(CST),
Department of CS&SE,
AU College of Engineering,
Andhra University, Visakhapatnam.

Prof.Kuda Nageswara Rao
Department of CS&SE,
AU College of Engineering,
Andhra University, Visakhapatnam.

Abstract:

Abstract: Microarray technology allows the simultaneous monitoring of thousands of genes. Based on the gene expression measurements, microarray technology have proven powerful in gene expression profiling for discovering new types of diseases and for predicting the type of a disease. Gridding, segmentation and intensity extraction are the three important steps in microarray image analysis. In this paper, three different clustering algorithms such as K-means, Moving K-means and Fuzzy C-means are used for segmentation of microarray images. In all the traditional clustering algorithms, number of clusters and initial centroids are randomly selected and often specified by the user. In this paper, a empirical mode decomposition algorithm for the histogram of the input image will generate the number of clusters and initial centroids required for clustering. It overcomes the shortage of random initialization and achieves high computational speed by reducing the number of iterations. The experimental results show that Fuzzy C-means has segmented the microarray image more accurately than other three algorithms.

Index terms: ***Microarray Image, Image Processing, Image segmentation***

INTRODUCTION

Microarrays widely recognized as the next revolution in molecular biology that enable scientists to monitor the expression levels of thousands of genes in parallel [1]. A microarray is a collection of blocks, each of which contains a number of rows and columns of spots. Each of the spot contains multiple copies of single DNA sequence [2]. The intensity of each spot indicates the expression level of the particular gene [3]. The processing of the microarray images [4] [5] usually consists of the following three steps: (i) gridding, which is the process of segmenting the microarray image into compartments, each compartment having only one spot and background (ii) Segmentation, which is the process of segmenting each compartment into one spot and its background area (iii) Intensity extraction, which calculates red and green foreground intensity pairs and background intensities.

In digital image segmentation applications, clustering technique is used to segment regions of interest and to detect borders of objects in an image. Clustering algorithms are based on the similarity or dissimilarity index between pairs of pixels. It is an iterative process which is terminated when all clusters contain similar data. In order to segment the image, the location of each spot must be identified through gridding process. An automatic gridding method by using the horizontal and vertical profile signal of the image presented in [5] is used to perform the image gridding. The algorithm can satisfy the requirements of microarray image segmentation.

In this paper, four different clustering algorithms are used for segmentation of microarray image. In the clustering algorithms, parameters such as cluster number and initial centroid positions are chosen randomly and often specified by the user. Instead of randomly initializing the parameters in the clustering algorithms, the EMD algorithm on the histogram of

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

input image will automatically determine the cluster centers and the number of clusters in the image. Using EMD algorithm as a preliminary stage with clustering algorithms reduces the number of iterations for segmentation and costs less execution time. The qualitative and quantitative results show that Fuzzy C-means clustering algorithm has classified the image better than other clustering algorithms. The paper is organized as follows: Section 2 presents EMD Algorithm, Section 3 presents K-means clustering algorithm, Section 4 presents Moving K-means clustering algorithm, Section 5 presents Fuzzy C-means clustering algorithm, Section 6 presents Experimental results and finally Section 7 report conclusions.

EMD ALGORITHM FOR ESTIMATION OF NUMBER OF CLUSTERS AND CENTROIDS

1. Let h(k) be the histogram for the input image I with k=0 ,…., G and G being the maximum intensity value in the image

2. Calculate the probability mass function p(k) for the input image histogram.

p(k) = h(k)/M

where M is the total number of image pixels.

3. Divide the normalized histogram p(k) into IMFs using empirical mode decomposition. The first IMF carries the histogram noise, irregularities and sharp details of the histogram, while the last IMF and residue describe the trend of the histogram. On the other hand, the intermediate IMFs describe the initial histogram with simple and uniform pulses.

4. Consider the summation of intermediate IMFs as follows:

$$\mathrm{I_{INT}} = \sum_{j=2}^{n-2} IMF_j$$

Where n is the number of IMFs.

5. Determine all local minima in I_{INT}.

$$I^* = \left\{ \min_{0 \le T \le G} I_{INT}(T) \right\}$$

I* is the vector carrying all local minima. All those local minima could express image clusters, but most of them are very close to each other and some of them lie too high to be a cluster. So, truncate the local minima to the important ones that could express an image cluster.

6. The truncation process is carried out in two steps. In the first step, the algorithm truncates all local minima that have a value larger than the threshold, where threshold is equal to average of the values of local minima. The truncation step is expressed as follows:

$$thr = \frac{1}{2N_{I^*}} \sum_{I_i^* \in I^*} I_i^*$$

where N_{I^*} is the number of local minima belonging to I^* and I_i^* is the local minima belonging to vector I^*. $I^t = \{I_i^*\}$, if $I_i^* < \text{thr}$ and $I_i^* \in I^*$.

Where I^t consists of all local minima which are less than the estimated threshold thr.

6. In the second truncation step, the algorithm applies an iterative procedure that calculates the number of image pixels belonging to each candidate image cluster and prunes the cluster with smallest number of image pixels (less than 2 percent of total number of image pixels). The pruned candidate clusters are merged with their closest image clusters.

7. The number of elements in final vector I^t represents the number of clusters denoted by NC.

K-MEANS CLUSTERING ALGORITHM

K-means is one of the basic clustering methods introduced by Hartigan [6]. This method is applied to segment the microarray image in recent years. The K-means clustering algorithm for segmentation of microarray image is summarized as follows [14]:

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Algorithm K-means(x, N, c)

Input:

N: number of pixels to be clustered;

$x=\{x_1, x_2, x_3,\ldots\ldots, x_N\}$: pixels of microarray image

$c = \{c_1, c_2, c_3,\ldots., c_j\}$: clusters respectively.

Output:

cl: cluster of pixels

Begin

Step 1: cluster centroids and number of clusters are determined by EMD algorithm.

Step 2: compute the closest cluster for each pixel and classify it to that cluster, ie: the objective is to minimize the sum of squares of the distances given by the following:

$$\Delta_{ij} = \| x_i - c_j \|. \quad \arg\min \sum_{i=1}^{N} \sum_{j=1}^{C} \Delta_{ij}^2 \qquad (1)$$

Step 3: Compute new centroids after all the pixels are clustered. The new centroids of a cluster is calculated by the following

$$c_j = \frac{1}{N_j} \sum x_i \text{ where } x_i \text{ belongs to } c_j . \qquad (2)$$

Step 4: Repeat steps 2-3 till the sum of squares given in equation is minimized.

End.

MOVING K-MEANS CLUSTERING ALGORITHM

The Moving K-means clustering algorithm is the modified version of K-means proposed in [7]. It introduces the concept of fitness to ensure that each cluster should have a significant number of members and final fitness values before the new position of cluster is calculated. The Moving K-means clustering algorithm for segmentation of microarray image is summarized as follows:

Algorithm Moving K-means(x, N, c)

Input:

N: number of pixels to be clustered;

$x=\{x_1,x_2,x_3,\ldots\ldots,x_N\}$: pixels of microarray image

$c =\{c_1,c_2,c_3,\ldots,c_j\}$: clusters respectively.

Output:

cl: cluster of pixels

Begin

Step 1: cluster centroids and number of clusters are determined by EMD algorithm.

Step 2: compute the closest cluster for each pixel and classify it to that cluster, ie: the objective is to minimize the sum of squares of the distances given by the following:

$$\Delta_{ij} = \| x_i - c_j \|. \quad \arg\min \sum_{i=1}^{N} \sum_{j=1}^{C} \Delta_{ij}^2 \qquad (3)$$

Step 3: The fitness for each cluster is calculated using

$$f(c_k) = \sum_{t \in c_k} (\| x_t - c_k \|)^2 \qquad (4)$$

All centers must satisfy the following condition:

$$f(c_s) \geq \alpha_a f(c_l) \qquad (5)$$

where α_a is small constant value initially with value in range $0< \alpha_a <1/3$, c_s and c_l are the centers that have the smallest and the largest fitness values. If (5) is not fulfilled, the members of c_l are assigned as members of c_s, while the rest are maintained as the members of c_l. The positions of c_s and c_l are recalculated according to:

$$C_s = 1/n_{cs} (\sum_{t \in c_s} x_t) \qquad (6)$$

$$C_l = 1/n_{cl} (\sum_{t \in c_l} x_t) \qquad (7)$$

The value of α_a is then updated according to:

$$\alpha_a = \alpha_a - \alpha_a/n_c \qquad (8)$$

The above process are repeated until (12) is fulfilled. Next all data are reassigned to their nearest center and the new center positions are recalculated using (9).

Step 4: The iteration process is repeated until the following condition is satisfied.

$$f(c_s) \geq \alpha_a f(c_l) \qquad (9)$$

End

EXPERIMENTAL RESULTS

Qualitative Analysis:

The proposed three clustering algorithms are performed on a two microarray image drawn from the standard microarray database corresponds to breast category aCGH tumor tissue. Image 1 consists of a

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

total of 38808 pixels and Image 2 consists of 64880 pixels. Clustering algorithms with and without EMD are executed on the two microarray images. The segmentation result of Fuzzy C-means clustering algorithm with EMD on two images is shown in figure 1. The EMD algorithm is executed on the histogram of input images for identification of number of clusters and initial centroids which are required for clustering algorithms. The centroids for the first image are 2 and 127, and for the second image the centroids are 17 and 181.

Quantitative Analysis:

Quantitative analysis is a numerically oriented procedure to figure out the performance of algorithms without any human error. The Mean Square Error (MSE) [10] is significant metric to validate the quality of image. It measures the square error between pixels of the original and the resultant images. The MSE is mathematically defined as

$$MSE = \frac{1}{N}\sum_{j=1}^{k}\sum_{i \in c_j} ||v_i - c_j||^2 \qquad (13)$$

Where N is the total number of pixels in an image and xi is the pixel which belongs to the jth cluster. The lower difference between the resultant and the original image reflects that all the data in the region are located near to its centre. Table 1 shows the quantitative evaluations of three clustering algorithms. The results confirm that Fuzzy C-means algorithm produces the lowest MSE value for segmenting the microarray image. As the initial centroids required for clustering algorithms are determined by EMD algorithm, the number of iterative steps required for classifying the objects is reduced. While the initial centroids obtained by EMD are unique, the segmented result is more stable compared with traditional algorithms. Table 2 shows the comparison of iterative steps numbers for clustering algorithms with and without EMD.

Method	MSE Values (IMAGE 1)	MSE Values (IMAGE 2)
K-means	282.781	346.47
Moving K-means	216.392	298.69
Fuzzy C-means	138.327	198.76

Table 1: MSE values

	Clustering algorithm	Iterative steps (without EMD)	Iterative steps (with EMD)
IMAGE 1	K-means	10	4
	Moving K-means	14	6
	Fuzzy C-means	17	9
IMAGE 2	K-means	20	12
	Moving K-means	27	16
	Fuzzy C-means	31	21

Table 2: Comparison of iterative step numbers

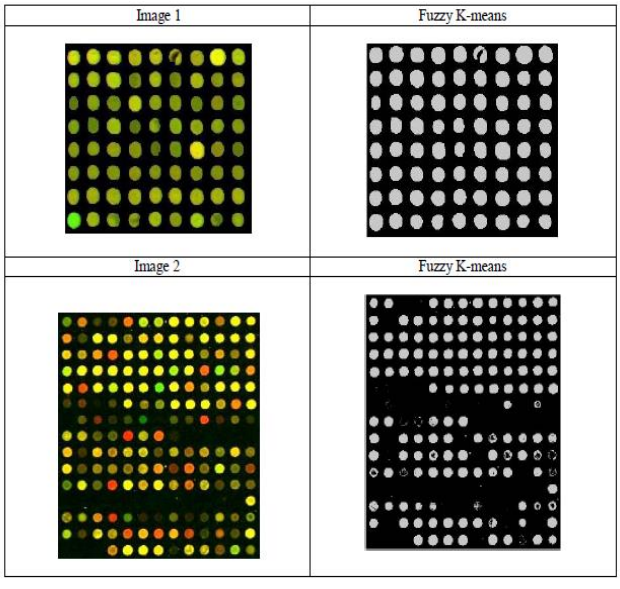

Fig 1: Segmentation result

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

CONCLUSION

This paper has presented three clustering algorithms namely K-means, Moving K-means and Fuzzy C-means combined with EMD for the segmentation of microarray image. The qualitative and quantitative analysis done proved that Fuzzy C-means has higher segmentation quality than other clustering algorithms. Clustering algorithm combined with EMD overcomes the problem of random selection of number of clusters and initialization of centroids. The proposed method reduces the number of iterations for segmentation of microarray image and costs less execution time.

REFERENCES

1. M.Schena, D.Shalon, Ronald W.davis and Patrick O.Brown,"Quantitative Monitoring of gene expression patterns with a complementary DNA microarray", Science, Oct 20;270(5235):467-70.

2. Wei-Bang Chen, Chengcui Zhang and Wen-Lin Liu, "An Automated Gridding and Segmentation method for cDNA Microarray Image Analysis", 19th IEEE Symposium on Computer-Based Medical Systems.

3. Tsung-Han Tsai Chein-Po Yang, Wei-ChiTsai, Pin-Hua Chen, "Error Reduction on Automatic Segmentation in Microarray Image", IEEE 2007.

4. Eleni Zacharia and Dimitirs Maroulis, "Microarray Image Analysis based on an Evolutionary Approach" IEEE 2008.

5. J.Harikiran, B.Avinash, Dr.P.V.Lakshmi, Dr.R.Kiran Kumar,"Automatic Gridding Method for Microarray Images", Journal of Applied Theoritical and Information Technology",Vol 65, No 1, pp 235-241, 2014.

6. Volkan Uslan, Omur Bucak, " clustering based spot segmentation of microarray cDNA Microarray Images ", International Conference of the IEE EMBS , 2010.

7. Siti Naraini Sulaiman, Nor Ashidi Mat Isa, "Denoising based Clutering Algorithms for Segmentation of Low level of Salt and Pepper Noise Corrupted Images", IEEE Transactions on Consumer Electronics, Vol. 56, No.4, November 2010.

8. LJun-Hao Zhang, Ming Hu HA , Jing Wu," Implementation of Rough Fuzzy K-means Clustering Algorithm in Matlab", Proceedings of Ninth International Conference on Machine Learning and Cybernetics", July 2010.

9. Nor Ashidi Mat Isa, Samy A.Salamah, Umi Kalthum Ngah.," Adaptive Fuzzy Moving K-means Clustering Algorithm for Image Segmentation", IEEE Transaction on Consumer Electronics, 12/2009; DOI: 10.1109/TCE.2009.5373781.

10. B.Saichandana, Dr.K.Srinivas, Dr.R.KiranKumar," De-noising based clustering Algorithm for Classification of Remote Sensing Image", Journal of Computing, Volume 4, Issue 11, November 2012.

11. Zhengjian DING, Juanjuan JIA, DIA LI , "Fast Clustering Segmentation Method Combining EMD for Color Image", Journal of Information and Computational Sciences, Vol 8, pp. 2949-2957.

12. Takumi OHASHI, Zaher AGHBARI, Akifumi MAKINOUCHI," EMD algorithm for Efficient Color bases Image Segmentation", Signal Processing, Pattern Recognition and Applications, 01/2003.

13. Zhengjian DING, Juanjuan JIA, DIA LI , "Fast Clustering Segmentation Method Combining EMD for Color Image", Journal of Information and Computational Sciences, Vol 8, pp. 2949-2957.