

Fuzzy C-means Model and Algorithm for Data Clustering

Sadina Gagula-Palalic¹, Mehmet Can¹

¹ *International University of Sarajevo, Paromlinska 66, Sarajevo, Bosnia and Herzegovina*

sadina@ius.edu.ba (Sadina Gagula-Palalic)

Abstract

Pattern recognition has become a very important field over the last decade since automation and computerization in many systems has led to large amount of data being stored in the databases. The primary intention of pattern recognition is to automatically assist humans in analyzing the vast amount of available data and extracting useful knowledge from it. Many algorithms have been developed for many applications, especially for static pattern recognition. Since the information of these processes can be non-deterministic over the time period, fuzzy approach can be applied to deal with this. In this work, fuzzy approach for optimization techniques in the pattern recognition will be implemented. It will show a fuzzy model for data clustering and feature extraction that best suits for the process of pattern recognition when we deal with non-crisp data.

I. INTRODUCTION

Pattern recognition is a discipline whose aim is to classify objects into a number of classes or categories each characterized with some property. General classification of the pattern recognition methods differentiates two techniques: supervised and unsupervised pattern recognition. A supervised method is characterized by the property that it uses data with known classification, while unsupervised methods serve for pattern recognition in datasets with unknown classification. In this work, we will deal with unsupervised methods, with special concern on clustering methods. Clustering is one of the most essential

processes in pattern recognition, since it plays a key role in finding the structures in data. Considering that real world problems in pattern recognition require the stochastic data to be processed, fuzzy clustering method will be shown to serve best in this case.

Fuzzy clustering can be divided into two basic groups, namely fuzzy c-means clustering (FCM) based on fuzzy partitions, and fuzzy hierarchical clustering method, based on fuzzy equivalence relations. Other methods derived from the two above were also proposed by researchers [1].

The aim of this work is to apply fuzzy c-means clustering model to image segmentation process. Texture segmentation problem is one of the fundamental one in the pattern recognition and image processing, such as in remote sensing, medical imaging, textile products inspection etc. Given an image with multiple texture regions, fuzzy c-means algorithm can be used to cluster the image vectors into several classes, each corresponding to the different regions. The output is the segmented image which can be further processed for noise removal if required.

II. FUZZY C-MEANS CLUSTERING MODEL

Fuzzy c-means clustering model was first introduced by Dunn in 1974 and later in 1973 extended and generalized by Bezdek [2]. Some derivations of this model related to the mixed fuzzy- possibilistic clustering also exists in the literature [3].

The algorithm for fuzzy c-means clustering consists of an iterative clustering method whose output is optimal 'c' partition, obtained by minimizing the objective function J_{FCM} [4]. The basic FCM algorithm is formulated as follows:

Minimize

$$J_{FCM}(u, v) = \sum_{k=1}^n \sum_{i=1}^c u_{ki}^m \|x_k - v_i\|^2$$

[1]

Subject to

$$1. \sum_{i=1}^c u_{ki} = 1, \text{ for } k = 1, 2, \dots, n$$

[2]

$$2. u_{ki} \geq 0, \text{ for all } i = 1, 2, \dots, c \text{ and } k = 1, 2, \dots, n$$

[3]

Given that

$$v_i = \frac{\sum_{k=1}^n u_{ki}^m x_k}{\sum_{k=1}^n u_{ki}^m}, i = 1, \dots, c$$

[4]

$$u_{ki} = \begin{cases} \left(\frac{\|x_k - v_i\|^2}{\sum_{j=1}^c \|x_k - v_j\|^2} \right)^{\frac{2}{m-1}} & \text{if } \|x_k - v_j\| > 0, \forall j \\ 1 & \text{if } \|x_k - v_j\| = 0 \\ 0 & \text{if } \exists j \neq i, \|x_k - v_j\| = 0 \end{cases} \quad [5]$$

where,

$x = \{x_1, x_2, \dots, x_n\} \subset R^p$ - are the feature data vectors

$v = \{v_1, v_2, \dots, v_c\} \subset R^p$ - are the cluster centers

$u = (u_{ki})_{n \times c}$ - is the fuzzy partition matrix that contains the membership values of each feature vector x_k in each cluster i

n - total number of data vectors in a given data set

c - number of clusters

m - fuzzification parameter – fuzzifier, $m > 1$

III. IMPLEMENTATION OF THE IMAGE SEGMENTATION USING FUZZY CLUSTERING

In order to implement the model described above for the process of image segmentation, 256-level gray images of different size (128x128, 256x256 and 512x512) were used. The image data was represented as a vector contained of all rows in the image concatenated to each other. The FCM algorithm was implemented in Mathematica using the following steps:

1. Input data – 1-dimensional, a vector containing gray levels 0-255 as values, were normalized in order to represent the membership of the

- pixel intensity to either 0 (dark) or 1 (light)
2. Initializing of the following parameters:
 - a. 'c' – number of clusters is chosen by test performer based on the graph of the gray levels of the pixels
 - b. 'm' – fuzzification parameter, $m=2$ in all examples
 - c. Maximum number of iterations – 100
 - d. Termination threshold, 'ε' – 0.0001
 - e. Initial cluster centers vector 'v'
 3. Iteration loop

Do

- Find norm distance between each point of data vector and all cluster center points for number of points: $n=1, \dots, N$; number of clusters $c=1, \dots, C$
- Calculate membership functions matrix 'u' using equation [5]
- From the obtained membership functions 'u', find new center 'v'

While

- Termination threshold is less than specified one
- The loop did not reach the maximum number of iterations

```

Do[{dis=Table[Norm[npt[[k]]-
v[[i]]],{k,1,n},{i,1,c}],
u=Table[(Sum[(dis[[k,i]]/dis[[k,j]])^(2/(m-
1))],{j,1,c})^(-1),{k,1,n},{i,1,c}],
tu=Transpose[u],
tum=Table[tu[[i,j]]^m,{i,1,c},{j,1,n}],
rep=Table[1/Sum[tum[[k,i]],{i,1,n}],{k,1,c
}],
dd=DiagonalMatrix[rep],
v1=dd.(tum.npt),
eps=Norm[v-v1],
If[eps<stop,{Print["v=",v," ","u=",
```

```
MatrixForm[N[Round[Transpose[u]*100]
```

```

/100]], " ","epsilon=",eps," ","iteration
number=",mm],
Break[{}],v=v1},{mm,1,maxiter}]
```

4. The outputs are fuzzy clustering centers (in this case normalized gray level values) and membership function vectors corresponding to each clustering center.

IV. RESULTS

The experiments were performed on three images. In the first case, a crisp image shown in figure 1 was tested. Its gray levels are plotted in the figure 2 showing that this image has 8 cluster centers. After performing the algorithm on the given image, 8 vectors of membership functions corresponding to 8 different cluster centers (gray levels) were obtained and shown in the figure 3. The first experiment showed that the fuzzy C-means clustering can perfectly cluster the image whose gray levels are well separated from each other. Clustering centers are shown in Table 1 for all images.

In the second case, Image 2 was tested, shown in the figure 4 and its graph of pixel values shown in the figure 5. The segmentation was performed for two clustering centers. The result (figure 4) shows a fairly good segmentation, differentiating well between low and high intensities.

Image 3 (figure 6), which represents mosaic of very similar textures with much similar gray levels was tested with 2 clustering centers. The result shows that the algorithm could not segment very properly the given image, especially when the textures were very similar.



Figure 1. Image 1 with crisp gray levels

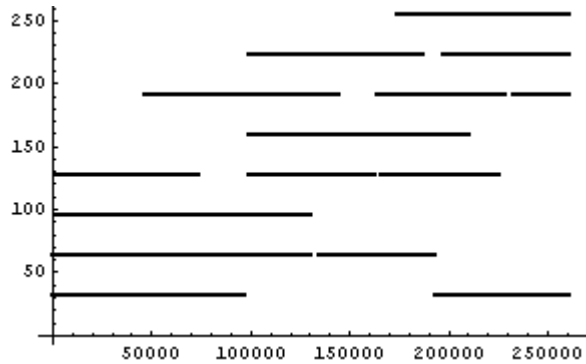


Figure 2. Gray levels of the pixels corresponding to Image 1

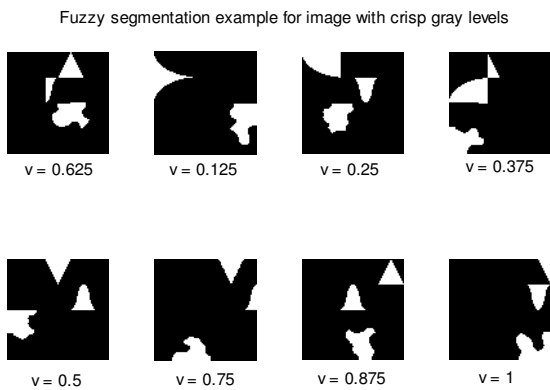


Figure 3. Segments of the Image 1 according to the clustering centers

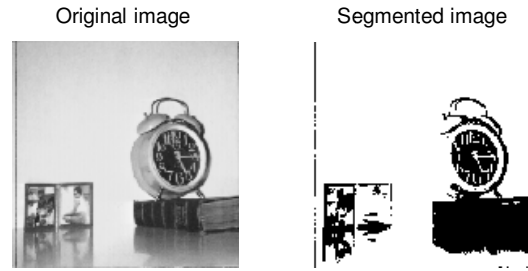


Figure 4. Image 2 – segmentation for $c=2$

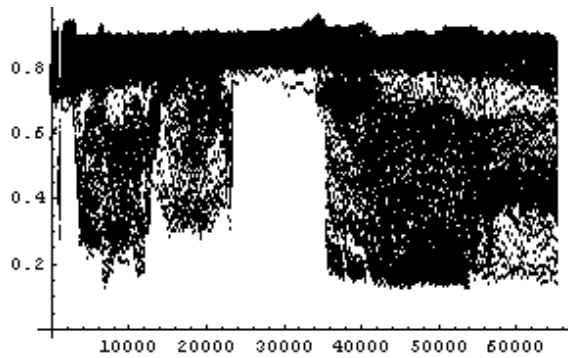


Figure 5. Gray levels of pixels corresponding to Image 2

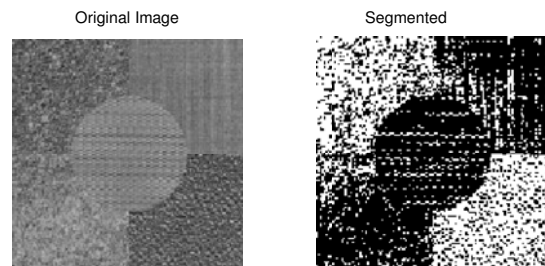


Figure 6. Image 3 - texture segmentation for $c=2$

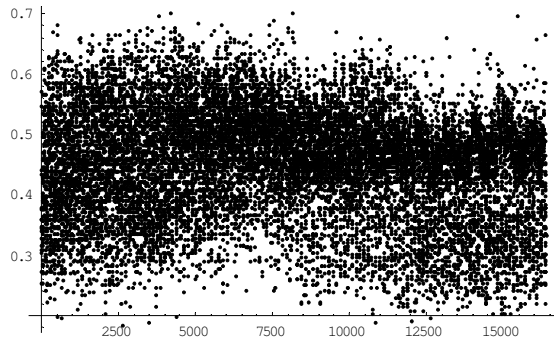


Figure 7. Image 3 - texture segmentation for $c=2$

TABLE 1: CLUSTER CENTERS FOR THREE IMAGES

Image	No of clusters	Cluster centers
Image 1	8	$v = \{0.625, 0.125, 0.25, 0.375, 0.5, 0.75, 0.875, 0.996094\}$
Image 2	2	$v = \{0.367103, 0.852043\}$
Image 3	2	$v = \{0.355303, 0.508327\}$

V. CONCLUSIONS

In this work, fuzzy c -means clustering model was applied in image segmentation. The algorithm was tested on the three kinds of images. It has been shown that this model can work very well on the images with high difference in gray level intensity between the background and front, as it has been shown in the case of clock image. On the other hand, the model fails to segment properly the image with very low difference in the intensity of the gray levels, like it shows the segmentation of the mosaic image. Therefore this method should be applied to images where segmentation is needed for the image

background. In addition, the described model and algorithm is sensitive to noise since it does not cover the spatial characteristics of an image. To overcome this problem, some filtering should be applied. The future work can be based on applying another fuzzy clustering model, which is fuzzy equivalence relations and improving signal to noise ratio of an image.

REFERENCES

- [1] G. Klir, Bo Yuan, "Fuzzy Sets and Fuzzy Logic: Theory and Applications", Prentice Hall, 1995.
- [2] J. C. Bezdek, J. Keller, R. Krishnapuram, N. R. Pal, "Fuzzy Models and Algorithms for Pattern Recognition and Image Processing", Springer, 2005.
- [3] N. R. Pal, K. Pal, J. C. Bezdek, "A Mixed C-Means Clustering Model", Fuzz IEEE, 1997.
- [4] Z. Chi, H. Yan, T. Pham, "Fuzzy Algorithms: with Applications to Image Processing and Pattern Recognition", Advances in Fuzzy Systems-Applications and Theory, Vol 10, 1996.
- [5] S. Theodoridis, K. Koutroumbas, Pattern Recognition, Elsevier, USA, 2003.