IJCSN International Journal of Computer Science and Network, Volume 5, Issue 1, February 2016
ISSN (Online) : 2277-5420 www.IJCSN.org
Impact Factor: 1.02

Improved Version of Kernelized Fuzzy C-Means using Credibility

Prabhjot Kaur

Maharaja Surajmal Institute of Technology (MSIT)
New Delhi, 110058, INDIA

Abstract - Fuzzy c-means is a clustering algorithm which performs well with noiseless data-sets. Various disadvantages of FCM are its sensitivity towards noise points and able to detect only spherical clusters due to euclidean distance metric and can work with only linear data. Kernel approaches can improve the performance of conventional clustering. It changes the behavior of algorithm from linear separability to non-linear separability. It can be achieved by using kernel function as a distance metric, which transforms the data to higher dimensional space and find the difference between points considering all the characteristics of data which are not accessible in two dimensional space. Kernel fuzzy *C*-means (KFCM) algorithm can efficiently work with non-linear data. But still it is sensitive to noisy points. This paper proposed kernel credibilistic fuzzy *C*-means (KCFCM) algorithm that uses credibility to reduce the sensitivity of noisy points. Several experimental results show that the proposed algorithm can outperform other algorithms for general data with additive noise.

Keywords - *Fuzzy C-means, Credibility, Kernel Function, Robust Image Segmentation.*

1. Introduction

Clustering process is done to divide the data into parts based upon some similarity. It can be used to solve many real time applications wherein we have to divide the data into various parts like pattern recognition, data mining, medical world to detect the deseases. Authors have suggested various ways to cluster the data into parts. Firt and formost algorithm to divide the data cased upon fuzzy sets is Fuzzy c means (FCM) proposed by J. C. Bezdek in 1991. Biggest disadvantage of FCM is that it is very much sensitive to noise. Its performance totally deteriorate when even a single outlier introduced into data. Many authors work upon this disadvantage. K. K. Chitalapudi proposed credibilistic fuzzy c means (CFCM) to remove the disadvantage of FCM. CFCM [2] is able to deal with noise points by introducing a credibility parameter, which assign lower membership to those points that are far away from the center of cluster. There is another disadvantage of FCM that it can detect only spherical clusters and only deals with linear data. Kernel approach can be used to change the behavior of algorithms from linear to non-linear separability.

This paper proposed kernel based credibilistic approach which can be used to avoid outliers as well as can deal with non-linear data. The organization of the paper is as follows: Section 2, briefly review FCM, CFCM and kernel approach. Section 3proposes the new Credibilistic fuzzy C-Means algorithm that follows RBF kernel approach. Section 4 gives the results and comparison analysis of the proposed algorithm with three already given fuzzy clustering algorithms Section 5 gives the final conclusion of this research.

2. Previous Work

In this section, FCM, CFCM and kernel function are briefly discussed.

2.1. The Fuzzy C- Means Algorithm (FCM)

Fuzzy c-means (FCM) [1] is a technique which detects clusters from the data based upon the distance of points from its centers. In FCM, membership is assigned to each point based upon its distance factor. The objective function of FCM is

$$FCM = \sum_{k=1}^{c}\sum_{i=1}^{n} u_{ki}^{m} d_{ki}^{2}, \qquad (1)$$

Where d_{ki} is the distance of any point from the center of cluster and u_{ki} is the membership of any point in the cluster. The constraint on the objective function is that the sum of all the memberships of the points within one cluster must be equal to 1. FCM is the best clustering algorithm if the data is free from noise. But in real situations, data always contain noise. So the disadvantage of FCM is that it is highly sensitive to noise and only able to detect spherical clusters because of euclidean distance norm.

2.2. Credibilistic Fuzzy C-Means (CFCM)

Krishna K. Chintalapudi[2] proposed CFCMto address the limitation of FCM i.e sensitivity of technique towards noise points. CFCM incorporates one parameter known as credibility. The impact of this factor is that it unlike FCM it assign lower membership values to noise points. CFCM defines credibility as:

$$\psi_k = 1 - \frac{(1-\theta)\alpha_k}{\max_{j=1..n}(\alpha_j)}, \quad 0 \le \beta \le 1 \tag{2}$$

Where $\alpha_k = \min_{i=1\ldots c}(d_{ik})$

α_k is the distance of vector x_k from its nearest centroid. The farther x_k is from its nearest centroid, the lower is its credibility.

Although, it is superior to FCM, PCM, and PFCM but it is observed that most of the time it assigns same outlier points to more than one cluster.

2.3. Kernel Base Approach

This approach helps to change the behavior of the algorithm from linear separability to non-linear separability. When kernel function is incorporated as a distance metric than normal function. It transform the two dimensional data to higher dimensional space and then do the operations. The major advantage of using kernel approach is that those characteristics of data which are not visible in two dimensional space are very much visible in higher dimensional space. In this way the accuracy of the algorithm become higher as compare to those which do the calculations in two dimensional space. There are many kernel functions. Based upon the application, we can choose the kernel function. As per literature, Gaussian kernel is best suitable for the images [3, 4, 8, 9].

In this paper, Radial basis kernel is used, which is the generalized form of Gaussian kernel.

$$\emptyset : R^P \rightarrow \text{H};\ x \rightarrow \emptyset(x) \tag{3}$$

Here, Ø represents the non-linear function for the higher dimensional space.

3. Proposed Technique

The paper proposed Radial basis kernel (RBF) based Credibilistic fuzzy cmeans (KCFCM). KCFCM replaces Euclidean distancemorm with RBF kernel. The objective function (OF) of KCFCM becomes:

$$KCFCM = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m}\, ||\emptyset(x_k) - \emptyset(v_i)||^2 \tag{4}$$

Subject to constraints,

$$\sum_{i=1}^{i=c} u_{ik} = \psi_k \quad ; \qquad k = 1 \ldots n \tag{5}$$

Where the membership equation is:

$$u_{ik} = u_{ik}^{FCM} * \psi_k \tag{6}$$

ψ_k is credibility providing typicality of a vector to dataset.

And, $$u_{ik}^{FCM} = \frac{1}{\sum_{i=1}^{c}(\frac{d_{ik}}{d_{jk}})^{\frac{2}{m-1}}} \tag{7}$$

In the expression for minimized OF for KCFCM, $||\emptyset(x_k - \emptyset v_i||^2$ is square of distance between $\emptyset(x_k)$ and $\emptyset(v_i)$ where distance in feature space is calculated through kernel in input space as follows:

$$||\emptyset(x_k) - \emptyset(v_i)||^2 = (\emptyset(x_k) - \emptyset(v_i))(\emptyset(x_k) - \emptyset(v_i))$$

$$||\emptyset(x_k) - \emptyset(v_i)||^2 = \emptyset(x_k)\emptyset(x_k) - 2\emptyset(x_k)\emptyset(v_i) + \emptyset(v_i)\emptyset(v_i)$$

$$||\emptyset(x_k) - \emptyset(v_i)||^2 = K(x_k, x_k) - 2K(x_k, v_i) + K(v_i, v_i)$$

$$||\emptyset(x_k) - \emptyset(v_i)||^2 = 2(1 - K(x_k, v_i))$$

Applying RBF in the above equation:

$$K(x, y) = \exp\left(-\frac{\sum |x_i^a - x_j^a|^b}{\sigma^2}\right) \tag{8}$$

Value of 'a' and'b' must be greater than one andσ is known as the kernel width, which has got different values for different data-sets.

By applying kernel function, the OF of becomes:

$$KCFCM = 2\sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m} * (1 - K(x_k, v_i))$$

$$KCFCM = 2\sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m}\left(1 - \exp\left(-\frac{\sum |x_i^a - x_j^a|^b}{\sigma^2}\right)\right) \tag{9}$$

By minimizing the objective function, following equations are obtained.

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}{}^{m} * k(x_k, v_i) * x_k}{\sum_{k=1}^{n} u_{ik}{}^{m} * k(x_k, v_i)} \tag{10}$$

$$u_{ik} = \frac{\frac{-\lambda_i}{2m(1-k(x_k,v_i))}^{1/m-1}}{\sum \frac{-\lambda_i}{2m(1-k(x_k,v_i))}^{1/m-1}} * \psi \quad (11)$$

The equations are obtained through following derivation:

$$KCFCM = 2\sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m} ||\emptyset(x_k) - \emptyset(v_i)||^2 + \sum_{i=1}^{c} \lambda_i \left(\sum_{k=1}^{n} u_{ik} - 1\right) \quad (12)$$

For v_i , Partial derivative of J_{KCFCM} with respect to v_i

$$\frac{\partial J}{\partial v_i} = \sum_{k=1}^{n} {u_{ik}}^{m} * 2\left(e^{-\frac{(x_k - v_i)^2}{\sigma^2}}\right) * 2\,\frac{(x_k - v_i)}{\sigma^2}$$

Equating it to zero, $\frac{\partial J}{\partial v_i} = 0$

$$\sum_{k=1}^{n} {u_{ik}}^{m} * 2\left(e^{-\frac{(x_k - v_i)^2}{\sigma^2}}\right) * 2\frac{(x_k - v_i)}{\sigma^2} = 0 \quad (13)$$

$$v_i = \frac{\sum_{k=1}^{n} {u_{ik}}^{m} * k(x_k, v_i) * x_k}{\sum_{k=1}^{n} {u_{ik}}^{m} * k(x_k, v_i)} \quad (14)$$

For u_{ik}, Partial derivative of J_{KCFCM} with respect to u_{ik}

$$KCFCM = 2\sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m} ||\emptyset(x_k) - \emptyset(v_i)||^2 + \sum_{i=1}^{c} \lambda_i \left(\sum_{k=1}^{n} u_{ik} - 1\right)$$

$$\frac{\partial J}{\partial u_{ik}} = m\,{u_{ik}}^{m-1} * 2\big(1 - k(x_k, v_i)\big) + \lambda_i$$

On equating it to zero, we get, $\frac{\partial J}{\partial u_{ik}} = 0$

$$m\,{u_{ik}}^{m-1} * 2\big(1 - k(x_k, v_i)\big) + \lambda_i = 0$$

$${u_{ik}}^{m-1} = \frac{-\lambda_i}{2m(1 - k(x_k, v_i))}$$

$$u_{ik} = \frac{-\lambda_i}{2m(1 - k(x_k, v_i))}^{1/m-1} \quad (15)$$

Keeping the constraint condition for KCFCM in mind,

$$\sum_{i=1}^{i=c} u_{ik} = \psi_k$$

$$u_{ik} = \frac{u_{ik}}{\sum u_{ik}} * \psi$$

$$u_{ik} = \frac{\frac{-\lambda_i}{2m(1-k(x_k,v_i))}^{1/m-1}}{\sum \frac{-\lambda_i}{2m(1-k(x_k,v_i))}^{1/m-1}} * \psi \quad (16)$$

Since the values of lagrangian multipliers vary negligibly, therefore they could be taken out of equation

$$.u_{ik} = \frac{1}{1-k(x_k,v_i)}^{1/m-1} * \psi \quad (17)$$

4. Results and Simulations

This section compares the performance of KCFCM with FCM, CFCM and KFCM. Empirical calculation are done using MATLAB tool with version 7.0. The following input values are used to conduct the experiment: m = 2, ε = 0.03,α= 0.85, which is the preferred value of α for images. Various data sets used MR and CT Scan images loaded from Medical image gallery.

4.1 Qualitative Analysis

4.1.1 MR Image of Brain

The segmentation of medical image is done with different values of kernel width 'σ'as shown in the Fig. 1, From the figure it is clear that the best results are detected with 'σ' value of 1. For the comparison of techniques in case of images, ground truth of images are used. In case of the testing image , its ground truth image was also available on medical image gallery. In the Fig. 2, the comparison of various algorithms are made with 'σ' value equal to 1.

4.2 Quantitative Analysis

To verify the performance of proposed technique and four previous algorithms based upon misclassifications, we have used three simulated/real brain images and calculated their scores defined by the following Quantitative index [6, 7].

$$r_{ij} = \frac{A_{ij} \cap A_{refj}}{A_{ij} \cup A_{refj}} \tag{18}$$

Here A_{ij} represents the set of pixels belonging to the j[th] cluster. A_{refj} represents the set of pixels belonging to the jth class in the reference segmented image. In, r_{ij} is defined as a fuzzy similarity measure which indicates the degree of equality between A_{ij} and A_{refj}. Large values of r_{ij} represents better segmentation. Table 1 lists the comparison scores for the four methods. From Fig. 2 and Table 1, we observe that KCFCM outperformed the other three algorithms.

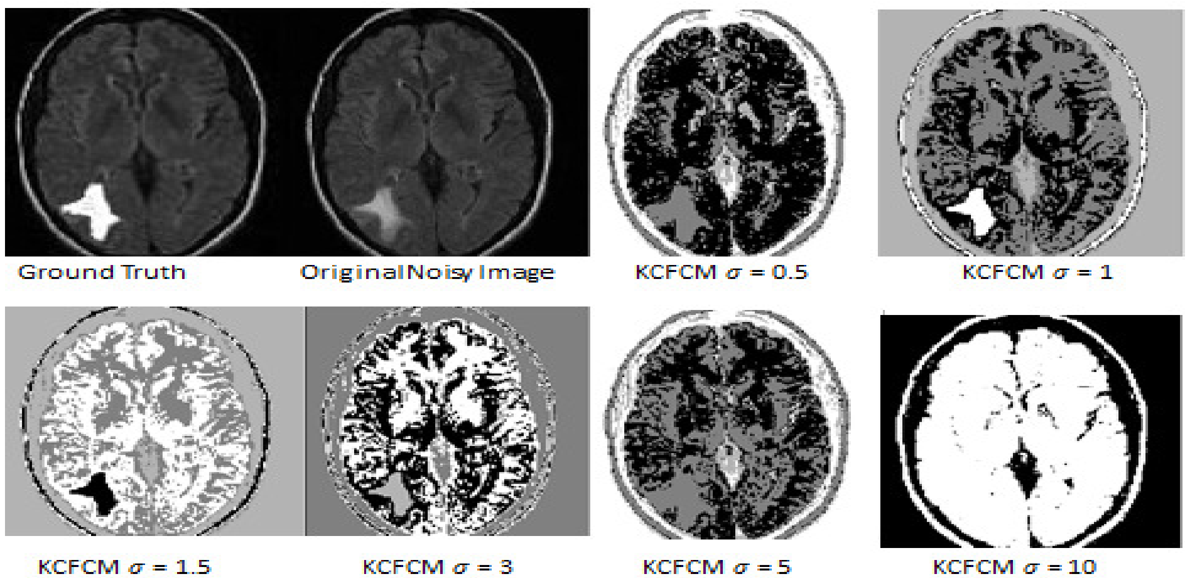

Fig.1: Comparison of MR image of brain (a) Ground truth(b) original noisy image (c) KCFCM for σ=0.5 (d) σ=1 (e) σ=1.5 (f) σ=3 (g) σ=5 (h) σ=10

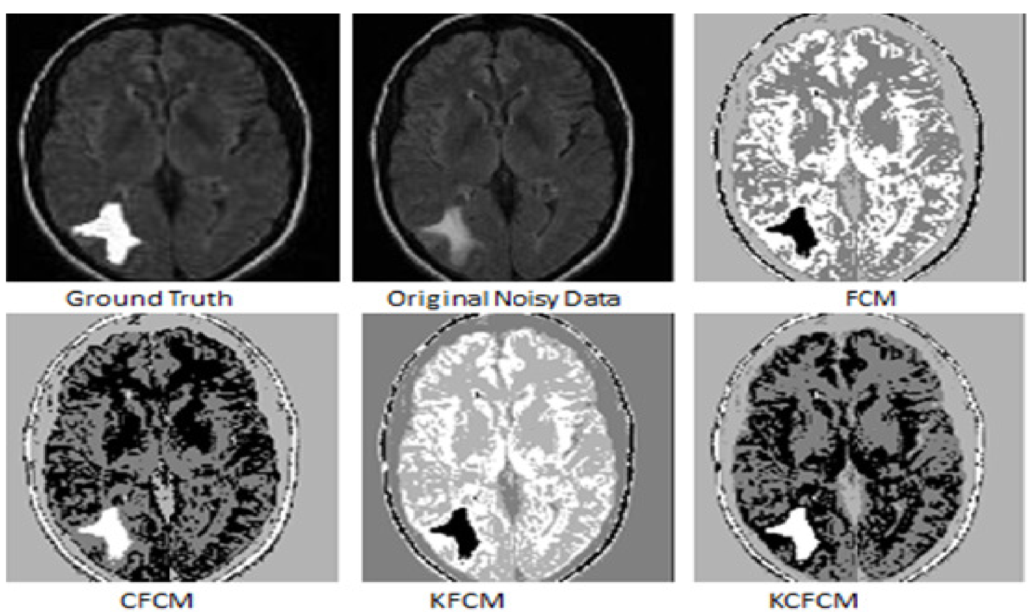

Fig.2 Comparison of MR image of brain (a) Ground Truth (b) Original Noisy Data (c) FCM (d) CFCM (e) KFCM (f) KCFCM σ=1

Table 1: Comparison Scores of four methods.

Name of the Algorithm	r_{ij} (Comparison Scores)
	Brain Image
	Tumor Region
FCM	**0.56**
CFCM	**0.81**
KFCM	**0.82**
KCFCM	**0.84**

5. Conclusion

This paper proposed a kernel based credibilistic fuzzy C-means (KCFCM) algorithm that applies the credibility parameter to the kernel fuzzy C-means (KFCM) algorithm to reduce the noise sensitivity of FCM.We observed empirically that the proposed Kernelized Credibilistic fuzzy C-Means(KCFCM) algorithm gives best results compared to the FCM, CFCM and KFCM algorithms when subjected to image data for general data with additive noise. KCFCM gives appropriate results for brain data it was subjected to and thus we conclude by saying that it is the most appropriate algorithm for image processing and segmentation amongst the algorithms studied. We also would bring to the readers' notice that the performance of the algorithm is strongly dependent on the value of sigma and the best value of sigma may vary for different datasets.

References

[1] Bezdek J. C., Pattern Recognition with Fuzzy Pointive Function Algorithm, Plenum, NY, 1981.

[2] Chintalapudi K. K. and M. kam, "A noise resistant fuzzy c-means algorithm for clustering," IEEE conference on Fuzzy Systems Proceedings, vol. 2, May 1998, pp. 1458-1463.

[3] Zhang Daoqiang, Chen Songcan 2002, "Fuzzy Clustering Using Kernel Method" , proceedings of the 2002 Interntional Conference on Control and Automation, Xiamen, China

[4] Du-Ming Tsai, Chung-ChanLin (2011), "Fuzzy C-means based clustering for linearly and nonlinearly separable data", in Elsevier Pattern Recognition 44 (2011) 1750–1760

[5] BrainWeb [online] available from the link as given:<http://www.bic.mni.mcgill.ca/brainweb>.

[6] Masulli, F., Schenone, A., 1999. A fuzzy clustering based segmentation system as support to diagnosis in medical imaging. Artif. Intell. Med. 16 (2), 129–147.

[7] W.A. Yasnoff, et al. (1977),"Error measures for scene segmentation", Pattern Recognition 9, 217–231.

[8] D. Q. Zhang and S. C. Chen, "Clustering incomplete data using kernel based fuzzy c-means algorithm", Neural Processing Letters 18(2003) 155-162.

[9] D. Q. Zhang and S. C. Chen, "A novel kernelized fuzzy c-means algorithm with application in medical image segmentation", Artificial Intelligence in Medicine 32(2004) 37-50.

Prabhjot Kaur is working as a Reader in Maharaja Surajmal Institute of Technology. She has done her Ph.D. in Computer Science. She is the member of IEEE, CSI and ISTE.