ISSN 2347 - 3983
Volume 9. No. 8, August 2021
International Journal of Emerging Trends in Engineering Research
Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter07982021.pdf
https://doi.org/10.30534/ijeter/2021/07982021

# A Novel Kernelized Fuzzy Clustering Algorithm for Data Classification

**Neha Mehra**
Department of Computer Engineering
Shri Govindram Seksaria Institute of Technology and Science Indore (M.P.),India
mehra.neha40@gmail.com

## ABSTRACT

Data are expanding day by day, clustering plays a main role in handling the data and to discover knowledge from it. Most of the clustering approaches deal with the linear separable problems. To deal with the nonlinear separable problems, we introduce the concept of kernel function in fuzzy clustering. In Kernelized fuzzy clustering approach the kernel function defines the non- linear transformation that projects the data from the original space where the data are can be more separable. The proposed approach uses kernel methods to project data from the original space to a high dimensional feature space where data can be separable linearly. We performed the test on the real world datasets which shows that our proposed kernel based clustering method gives better accuracy as compared to the fuzzy clustering method.

**Key words:** Fuzzy clustering, Fuzzy C-Means, Kernel methods

## 1. INTRODUCTION

Clustering is mostly used unsupervised technique in data mining [1]. It is a way of assigning similar data points into clusters based on some similarity measures. The main aim is to assign data point such as there should be high inter cluster distance and low intra cluster distance. Clustering is broadly divided into two parts, i.e., hierarchical and partitioning clustering. Hierarchical clustering finds the clusters by partitioning the data in either a top-down or bottom-up fashion in a recursive manner, whereas partitioning clustering creates partitions of data, by using any optimizing criteria [2]. Partitioning clustering is further divided into crisp and fuzzy clustering. In crisp clustering each data point in the sample space is assigned to only one cluster [3]. To overcome this limitation, the concept of fuzzy clustering is introduced.

In Fuzzy clustering, a data sample can belong to one than one cluster with different degrees of membership and the membership is spread among all clusters. Fuzzy C-Means become the most widely used algorithms in fuzzy clustering, wherein each data point can have membership in more than one cluster [4]. Fuzzy C-Means minimizes an objective function subject to some constraints. In most cases Fuzzy C-Means is based on dealing with linear relations among data samples. For handling non-linear data samples, the concept of kernel function is introduced [5].

## 2. LITERATURE REVIEW

Many researchers have worked on the non-linear mapping of data samples. As we know that the efficiency of fuzzy C- Means algorithm is limited to spherical clusters. Huang [6] discussed this problem by mapping nonlinear data in appropriate feature space by applying kernel tricks that extend Fuzzy C-Means method with multiple learning, where data can be linearly mapped. Girolami [7] has explored the notion of data clustering in a kernel defined feature space. He also discussed about the choice of the kernel chosen in defining the nonlinear mapping and how to choose the parameter of the kernel. Tzortzis [8] introduces K-Means groups, which minimizes the clustering error by finding the sum of the squared Euclidean distances between each dataset point and its cluster center. He discussed Kernel K-Means method that identifies nonlinearly separable cluster that minimizes the error in clustering approach in the feature space.

In the research field, Image segmentation also plays important role in identifying patterns and image analysis. Reddy [9] discussed the kernel based FCM method that was used for the segmentation of low contrast images and medical images. Baili [10] proposed the Fuzzy C-Means with multiple kernels which allow the soft linear partitioning of the feature space. In his paper, he discussed to partition the data into appropriate clusters and assigns weight for each kernel in each cluster. Liu [11] in this paper developed a kernel based fuzzy attribute c-means clustering algorithm that finds the distance in the fuzzy attribute C- Means clustering algorithm with induced kernel distance.

Lui [12] in his paper introduce an attribute multi-kernel weighted Fuzzy C- Means method for projection of data in feature space and extracting features so that efficiency can be improved. Nia [13] discussed about the limitation of segmentation method that gets fixed in local minima. Its behavior becomes random because of the outlier in the data. He proposed a meta-heuristic algorithm, differential evolution and RBF kernel based algorithm to segment image in noise data.

The proposed kernel based Fuzzy C- Means algorithm deals with the limitation of not separating the data sample linearly by mapping data into a high dimensional feature space, where the patterns can be linear among data samples. Section 2 briefly discussed about the preliminaries needed to develop our algorithm. Our proposed method is described in detail in section 3. The experimental results are discussed and compared with other clustering methods in section 4, followed by concluding remarks in section 5.

## 3. PRELIMINARIES

The mostly used clustering algorithm is Fuzzy C-Means, where each data point belongs to a cluster with some degree of membership. Fuzzy C-Means were originally developed by Jim Bezdek in 1981 [14]. This technique takes a dataset $x_i=\{x_1, x_2,\ldots,x_N\}$ as input and produces the membership matrix U, which denotes the possibility of belongingness of a data sample to the clusters.

The objective function J can be minimized by:

$$\boldsymbol{J(U,V) = \sum_{j=1}^{c}\sum_{i=1}^{N} u_{ij}^{m}\left\|x_i - v_j\right\|^2} \quad (1)$$

Where, each data sample $x_i$ is an object, N is the total number of training samples, c is the number of clusters and has a set of membership values $u_{ij}$ associated with it, the set of cluster centers is represented by $v = \{v_1, v_2,\ldots v_c\}$. The fuzzy matrix $U = (u_{ij})_{cxN}$ is the fuzzy memberships of each training sample $x_i$ to each cluster $v_j$. The membership degree of all the samples on fuzzy partitions must be one. The value of fuzzification parameter m affects the result of clustering algorithms and also used to reduce sensitivity to noise in the clustering algorithm. The higher values lead to more clusters; all elements tend to belong to all clusters and lower values lead to hard clusters. Bezdek has proved that the value of fuzzification parameter should be set between 1.5 and 2.5 for better clustering results.

The cluster membership matrix $u_{ij}$ corresponding to data point $x_i$ to each cluster $v_j$ can be calculated as

$$u_{ij} = \frac{\|x_i - v_j\|^{\frac{-1}{m-1}}}{\sum_{k=1}^{c}\|x_i - v_k\|^{\frac{-1}{m-1}}}, \ \forall\, i, j \quad (2)$$

Cluster Center can be computed as follows

$$v_j' = \frac{\sum_{i=1}^{n}[u_{ij}]^m x_i}{\sum_{i=1}^{n}[u_{ij}]^m}, \ \forall\, j \quad (3)$$

To minimize the objective function given in equation (1) the basic FCM algorithm is explained in Algorithm 1.

---

Algorithm 1: Fuzzy C-Means

---

Input: X, V, c, m
Output: U, V'
Step 1: Initialize cluster centers at random $V= \{v_1, v_2, \ldots, v_c\}$
Step 2: Compute cluster membership by using equation (2)
Step 3: Compute cluster centers by using equation (3)
Step 4: If $\|V' - V\| < \epsilon$ then stop, otherwise continue with step 2.

---

The equation (2) and (3) are updated until the predefined convergence threshold is met and then the algorithm terminates.

Fuzzy C-Means can handle linear relations; to overcome this limitation the concept of kernel method is introduced [15]. The kernel method maps data space that is highly non linear to a high dimensional feature space, using a set of mathematical functions ($\phi: X \rightarrow F$).

The kernel space can be of any number of dimensions. The main aim of going to higher dimensions is that, while remaining in feature space where the data samples are highly non-linear and not linearly separable, it is possible to apply a linear classifier in that space. So, that data sample can be linearly separated.

A main concept behind the kernel based algorithms is the kernel trick. A kernel trick mathematically applied to any algorithm which mainly depends on the dot product between two vectors. The main aim of the kernel trick is to convert the nonlinear problem to linear problem in the low dimensional input space [16].

A kernel is a function K that for all x, k from the original input space x satisfies

$$K(x,k) = \langle \phi(x), \phi(k) \rangle \quad (4)$$

Where $\phi(x), \phi(k)$ denotes the inner dot product and $\phi$ is a nonlinear map function from the input space x to a rather high dimensional feature space F.

$$\phi : x \rightarrow \phi(x) \in F \quad (5)$$

With the consideration of the beneficial result of the kernel based method that is more robust to outliers and performs better on non-linearity of data. We will expand the equation (1) using the kernel in our proposed method.

## 4. KERNEL BASED FUZZY CLUSTERING

Consider some non-linear mapping function $\phi : x \rightarrow \phi(x) \in F^d$ with d dimensions of the feature vector. The kernel function can be of many types Polynomial, radial basis function (RBF), etc.

RBF kernel $K(x_1, x_2) = \exp\left(\frac{-1}{\sigma^2}\|x_i - v_j\|^2\right)$ is the most popular kernel function, where $\sigma$ is the kernel parameter. The proposed method uses the RBF kernel.

Kernel based FCM (KFCM) function using the mapping $\phi$ can be generally defined as the constrained minimization of the objective as described below:

$$J(U, V) = \sum_{j=1}^{c}\sum_{i=1}^{N} u_{ij}^{m}\|\phi(x_i) - \phi(v_j)\|^2 \quad (6)$$

Where $\|\phi(x_i) - \phi(v_j)\|^2$ is the Euclidean distance between $\phi(x_i)$ and $\phi(v_j)$, and $\phi(x_i)$ and $\phi(v_j)$, are the kernel spaces of $x_i$ and $v_j$, respectively. N is the total number of training samples and c is the number of clusters. $u_{ij}$ is the fuzzy membership on the each training sample $x_i$ to each cluster $v_j$.

$$\|\phi(x_i) - \phi(v_j)\|^2 = \left(\phi(x_i) - \phi(v_j)\right)^T\left(\phi(x_i) - \phi(v_j)\right)$$

$$= \phi(x_i)^T\phi(x_i) - \phi(v_j)^T\phi(x_i) - \phi(x_i)^T\phi(v_j) + \phi(v_j)^T\phi(v_j)$$

$$= K(x_i, x_i) + K(v_j, v_j) - 2K(x_i, v_j) \quad (7)$$

In the case of RBF kernel $K(x_i, x_i) = 1$ and $K(v_j, v_j) = 1$

So the equation (7) can be rewritten as

$$J(U, V) = 2\sum_{j=1}^{c}\sum_{i=1}^{N} u_{ij}^{m}\left(1 - K(x_i, v_j)\right) \quad (8)$$

To minimize the objective function the membership matrix $u_{ij}$ and cluster center $v_j$ need to be computed.

The membership matrix can be calculated as

$$u_{ij} = \frac{\left(1 - K(x_i, v_j)\right)^{-1/(m-1)}}{\sum_{j=1}^{c}\left(1 - K(x_i, v_j)\right)^{-1/(m-1)}} \quad (9)$$

The cluster center can be computed as

$$v_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} K(x_i, v_j)\, x_i}{\sum_{i=1}^{N} u_{ij}^{m} K(x_i, v_j)} \quad (10)$$

The choice of kernel parameter is the most critical task. In our proposed work with the kernel parameter can be selected as below:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(d_i - \bar{d})^2}{N-1}} \quad (11)$$

Where, $d_i = \|x_i - \bar{x}\|$ and $\bar{d}$ Is the average of all distances $d_i$.

To minimize the objective function given in equation (9) the proposed kernel based fuzzy clustering algorithm is explained in Algorithm 2.

---

Algorithm 2: Kernel based Fuzzy Clustering Algorithm

---

Input: X, V, c, m, σ
Output: U, V'
Step 1: Randomly Initialize cluster centers V= {$v_1$, $v_2$, ..,$v_c$}
Step 2: Compute cluster membership by using equation (9)
Step 3: Compute cluster centers by using equation (10)
Step 4: If $\|V' - V\| < \epsilon$ then stop, otherwise continue with step 2.

---

The equation (9) and (10) are updated until the predefined convergence threshold is met and then the algorithm terminates.

## 5. EXPERIMENTS AND RESULTS

In this section, experimental studies of the proposed approach are conducted on five real world dataset. Firstly, c objects are chosen at random from datasets as initial clusters initialize the membership matrix U. We set fuzzification parameter m as 1.75 and convergence threshold as $10^{-3}$. In the experiments, we compare the performance of FCM and KFCM to show the clustering results and robustness of KFCM over FCM on various datasets.

### 5.1 Datasets

All the datasets are taken from the UCI Machine learning Repository. The dataset which has different number of instances and attributes are chosen for comparison.

*1) Seed:* The dataset consists of 210 instances, with each instance has 7 attributes. All data points were divided into 3 classes.

*2) Wine*: The dataset consists of 178 instances, with each instance has 13 attributes. All data points were divided into 3 classes.

*3) Breast Cancer Wisconsin:* The dataset consists of 699 instances, with each instance has 10 attributes. All data points were divided into 2 classes. The dataset contains 16 missing values. For our experimentation, we have ignored the instances which have missing values and considered 683 instances.

*4) Balance Scale:* The dataset consists of 625 instances, with each instance has 4 attributes. All data points were divided into 3 classes.

*5) Ecoli:* The dataset consists of 336 instances, with each instance has 7 attributes. All data points were divided into 8 classes.

Table 1 shows the different dataset name, number of instances in each dataset and the number of classes which each data point is divided.

Table 1. Description of Datasets

| Datasets | Parameters | | |
|---|---|---|---|
| | No. of Instances | No. of Features | No. of Classes |
| Seed | 210 | 7 | 3 |
| Wine | 178 | 13 | 3 |
| Breast Cancer Wisconsin | 683 | 10 | 2 |
| Balance Scale | 625 | 4 | 3 |
| Ecoli | 336 | 7 | 8 |

### 5.2 Performance Parameters

We evaluate the performance of the clustering method using four parameters described below:

#### *5.2.1 F-Measure*

Originally, the F-measure was proposed by Rijsbergen [17] in the terms of information retrieval. The F-measure is often used as a standard balance between precision and recall for measuring goodness or accuracy of clustering methods. The value near to one of F-measure denotes the better clustering results.

$$\text{Recall}_{ij} = \frac{N_{ij}}{N_i} \tag{12}$$

$$\text{Precision}_{ij} = \frac{N_{ij}}{N_j} \tag{13}$$

$$F-\text{Measure} = \frac{2*\text{Recall}_{ij}*\text{Precision}_{ij}}{\text{Recall}_{ij}+\text{Precision}_{ij}} \tag{14}$$

Where $N_j$ is elements of class $j$ and $N_i$ is the number of elements of cluster $i$, $N_{ij}$ is the numbers of elements of class $i$ in cluster $j$.

#### *5.2.2 Normalized Mutual Information (NMI)*

NMI is a good measure for determining the quality of clustering [18].

$$NMI = \frac{\sum_{c=1}^{k}\sum_{p=1}^{m} n_c^p \log\left(\frac{n \cdot n_c^p}{n_c \cdot n_p}\right)}{\sqrt{\left(\sum_{c=1}^{k} n_c \log\left(\frac{n_c}{n}\right)\right)}\left(\sqrt{\left(\sum_{c=1}^{k} n_p \log\left(\frac{n_p}{n}\right)\right)}\right)} \tag{15}$$

Where, $n$ is the total number of data points, $n_c$ and $n_p$ are the numbers of data points in the $c^{th}$ cluster and the $p^{th}$ class, respectively, and $n_c^p$ is the number of common data points in class $p$ and cluster $c$.

#### *5.2.3 Adjusted Rand Index (ARI)*

The Rand index measures the similarity between two data clustering. An extended form of the Rand index is the adjusted Rand index. The adjusted Rand index has the maximum value 1, and the value 0 is in the case of random clusters [19].

$$ARI = \frac{\sum_{i,j}\binom{n_{ij}}{2} - \left[\sum_i \binom{n_{i.}}{2}\sum_j \binom{n_{.j}}{2}\right] / \binom{n}{2}}{\frac{1}{2}\left[\sum_i \binom{n_{i.}}{2} + \sum_j \binom{n_{.j}}{2}\right] - \left[\sum_i \binom{n_{i.}}{2}\sum_j \binom{n_{.j}}{2}\right] / \binom{n}{2}} \tag{16}$$

#### *5.2.4 Objective Function*

The main aim of the clustering approach is to minimize the objective function. The proposed kernel based Fuzzy clustering approach is also evaluated on the basis of the objective function.

Table 2. F-Measure for FCM and KFCM on various Datasets

| Datasets | FCM | KFCM |
|---|---|---|
| Seed | 0.43517 | 0.60976 |
| Wine | 0.34721 | 0.64912 |
| Breast Cancer Wisconsin | 0.47674 | 0.78793 |
| Balance Scale | 0.21968 | 0.63089 |
| Ecoli | 0.37596 | 0.68654 |

Table 3. NMI for FCM and KFCM on various Datasets

| Datasets | FCM | KFCM |
|---|---|---|
| Seed | 0.69483 | 0.70976 |
| Wine | 0.46952 | 0.87422 |
| Breast Cancer Wisconsin | 0.73001 | 0.89603 |
| Balance Scale | 0.24539 | 0.79285 |
| Ecoli | 0.55712 | 0.8274 |

Table 4. ARI for FCM and KFCM on various Datasets

| Datasets | FCM | KFCM |
|---|---|---|
| Seed | 0.3381 | 0.6234 |
| Wine | 0.48052 | 0.92468 |
| Breast Cancer Wisconsin | 0.95608 | 0.97281 |
| Balance Scale | 0.3024 | 0.6752 |
| Ecoli | 0.28571 | 0.71905 |

Table 5. Objective Function for FCM and KFCM on various Datasets

| Datasets | FCM | KFCM |
|---|---|---|
| Seed | 1182.73 | 838.9414 |
| Wine | 2788.39 | 777.5576 |
| Breast Cancer Wisconsin | 30147.6 | 832.4609 |
| Balance Scale | 1250.44 | 891.3369 |
| Ecoli | 959.84 | 853.4041 |

The above result shows that in all the datasets F-Measure is increased in our proposed algorithm. Maximum improvement is shown in Balance Scale dataset. Nearly 10 % to 20% improvement is observed in F-Measure. The NMI measure also shows the improvement of our proposed algorithm. Maximum improvement is shown in Balance Scale dataset. Ecoli dataset shows the maximum improvement in ARI measure. Nearly 10 % to 20% improvement is observed in Ari Measure. The main aim of the fuzzy clustering approach is to minimize the objective function. The proposed algorithm shows maximum minimization of the objective function in Wine and Breast Cancer Wisconsin dataset. The above discussion shows that the proposed kernel based clustering algorithm is comparable efficient to the Fuzzy C-Means algorithm.

## 6. CONCLUSION

Clustering is most efficient technique used in the field of research. The main advantage of the research is to include the concept of kernel methods. In this paper a Kernel based Fuzzy clustering algorithm is proposed to handle the non-linear data which deals with the limitation of Fuzzy C-Means algorithm. It uses RBF kernel as the kernel function for fuzzy clustering. The effectiveness of the proposed method is compared with the Fuzzy C means, which shows that the proposed approach performs better clustering results. The proposed algorithm shows 10 % to 20% improvement in all the measures that we have discussed. The work can be extended in future to process big data.

## REFERENCES

[1] Garima, Hina Gulati, P.K Singh, "Clustering Techniques in Data Mining: A Comparison", 2nd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, pp. 410 -415, 2015.
[2] Malika Bendechache, Nhien-An Le-Khac and M-Tahar Kechadi, "Efficient Large Scale Clustering based on Data Partitioning", IEEE International Conference on Data Science and Advanced Analytics, Montreal, QC, Canada, pp. 612 - 621, 2016.
[3] Raju G, Binu Thomas, Sonam Tobgay and Th. Shanta Kumar, "Fuzzy Clustering Methods in Data Mining: A comparative Case Analysis", International Conference on Advanced Computer Theory and Engineering, Phuket, Thailand, pp. 489 - 493, 2008.
[4] Dae-Won Kim, Kwang H. Lee, Doheon Lee, “On cluster validity index for estimation of the optimal number of fuzzy clusters”, Pattern Recognition Society, Published by Elsevier, pp. 2009- 2025, April 2004.
[5] Xiao-Hong Wu, Jian-Jiang Zhou, “Kernel-based Fuzzy K-nearest-neighbor Algorithm”, International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies

and Internet Commerce (CIMCA-IAWTIC'06), Vienna, Austria, Vol. 2, pp. 159 - 162, November 2005.
[6] Hsin-Chien Huang, Yung-Yu Chuang, and Chu-Song Chen, "Multiple Kernel Fuzzy Clustering", IEEE Transactions on Fuzzy Systems, Vol. 20, No. 1, pp. 120 - 134, February 2012.
[7] Mark Girolami, "Mercer Kernel – Based Clustering in Feature Space", IEEE Transactions on Neural Networks, Vol. 12, No. 3, pp, 780 – 784, May 2002.
[8] Grigorios F. Tzortzis and Aristidis C. Likas, "The Global Kernel- Means Algorithm for Clustering in Feature Space", IEEE Transactions on Neural Networks, Vol. 20, No. 7, pp, 1181 – 1194, July 2009.
[9] G. R. Reddy, K. Ramudu, S. Zaheeruddin, and R. R. Rao, "Image segmentation using kernel fuzzy c-means clustering on level set method on noisy images," in Proceedings of the International Conference on Communications and Signal Processing (ICCSP '11), Calicut, India, pp. 522–526, February 2011.
[10] Naouel Baili and Hichem Frigui, "Fuzzy Clustering with Multiple Kernels in Feature Space", WCCI IEEE World Congress on Computational Intelligence, Brisbane, Australia, pp. 1- 8, 10-15, June, 2012.
[11] J. Liu and M. Xu, "Kernelized fuzzy attribute C-means clustering algorithm," Fuzzy Sets Syst., vol. 159, pp. 2428–2445, 2008.
[12] Zhulin Liu, C. L. Philip Chen, Long Chen, and Jin Zhou, 'Multi-Attribute Based Fuzzy C-means in Approximated Feature Space", International Conference on Fuzzy Theory and Its Applications, Pingtung, Taiwan, pp. 1 – 6, 2017.
[13]Alireza Farrokhi Nia and Mousa Shamsi, "Brain MR Image Segmentation Using Differential Evolution Guided by RBF-Kernel based FCM", 10th Iranian Conference on Machine Vision and Image Processing (MVIP), Isfahan, Iran, pp. 221 – 227, 2017.
[14] James C. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithms", Plenum Press, New York, 1981.
[15]Tomer Hertz Tomboy, Aharon Bar Hillel and Aharonbh Daphna Weinshall, "Learning a Kernel Function for Classification with small Training Samples", In proceeding of Machine Learning, proceedings of the twenty-third International Conference (ICML 2006), Pittsburgh, Pennsylvania, USA, pp. 401 - 408, 25-29 June 2006.
[16] Tomas M. Cover, "Geometrical and statistical properties of systems of linear inequalities in pattern recognition", IEEE transactions on Electronic Computers. 14, Vol. EC-14, No. 3, pp, 326–334, June 1965.
[17] C. J. Van Rijsbergen, Information Retrieval, Butterworth, 1979.
[18] Neha Bharill, Aruna Tiwari and Aayushi Malviya, "Fuzzy Based Scalable Clustering Algorithms for handling Big data Using Apache Spark", IEEE Transactions on Big Data, Vol. 2, No. 4, pp, 339–352, December 2016.
[19] Liang Bai, Jiye Liang and Yike Guo, "An ensemble clusterer of multiple fuzzy k-means clusterings to recognize arbitrarily shaped clusters", IEEE Transactions on Fuzzy Systems, IEEE Early Access Articles**,** pp, 1-1, May 2018.
[20 Himanshi Agrawal, Neha Mehra, Vandan Tewari, "Implementation of Protein Sequence Classification for Globin family using Ensemble Learning", International Journal of Emerging Trends in Engineering Research, Volume 9. No. 4, April 2021.