



## Dependence of Two Different Fuzzy Clustering Techniques on Random Initialization and a Comparison

**Samarjit Das***Department of Computer Science & IT  
Cotton College, Assam, India***Hemanta K. Baruah***Vice-Chancellor  
Bodoland University, Assam, India*

**Abstract**— *In the recent past Kernelized Fuzzy C-Means clustering technique has earned popularity especially in the machine learning community. This technique has been derived from the conventional Fuzzy C-Means clustering technique of Bezdek by defining the vector norm with the Gaussian Radial Basic Function instead of a Euclidean distance. Subsequently the fuzzy cluster centroids and the partition matrix are defined using this new vector norm. In our present work we have tried to show the effect of random initialization of the membership values on the performances of both these techniques. In addition to this we have tried to show the variation of the performance of Kernelized Fuzzy C-Means clustering technique with different values of the adjustable parameter of its vector norm. Using Partition Coefficient and Clustering Entropy as validity indices we have tried to make a comparison of the performances of these two clustering techniques.*

**Keywords**— *Kernelized Fuzzy C-Means Clustering Technique, Fuzzy C-Means Clustering Technique, Gaussian Radial Basic Function, Euclidean Distance, Partition Coefficient, Clustering Entropy.*

### I. INTRODUCTION

Clustering is a technique by means of which a large dataset is partitioned into some smaller groups, called clusters, using some similarity measures, such that the similarity between any two objects in the same group is more than that between two objects in two different groups. In conventional hard clustering an object either fully belongs to a particular cluster or does not belong to it at all. With the advent of the concept of fuzzy set theory (FST), developed by Zadeh [1], which particularly deals the situations pertaining to non-probabilistic uncertainty, the conventional hard clustering technique has unlocked a new way of clustering known as fuzzy clustering, where a single object may belong exactly to one cluster or partially to more than one clusters depending on the membership value of that object. Baruah ([2], [3]) has shown that the membership value of a fuzzy number can be expressed as a difference between the membership function and a reference function, and therefore the fuzzy membership function and the fuzzy membership value for the complement of a fuzzy set are not the same. Although the Fuzzy C-Means clustering technique (FCM) of Bezdek [4] has been found to be widely studied and applied, in the recent past Kernelized Fuzzy C-Means clustering technique (KFCM) ([5], [6]) has earned popularity specially in the field of machine learning. Derrig and Ostaszewski [7] have applied the FCM of Bezdek in their research work where they have explained a method of pattern recognition for risk and claim classification. Das and Baruah [8] have shown the application of Bezdek's [4] FCM clustering technique on vehicular pollution, through which they have discussed the importance of application of a fuzzy clustering technique on a dataset of vehicular pollution instead of a hard clustering technique. Das and Baruah [9] have applied the FCM clustering technique of Bezdek [4] and the Gustafson-Kessel (GK) clustering technique of Gustafson and Kessel [10] on the same dataset to make a comparison between these two clustering techniques and found that the overall performance of FCM clustering technique is better than that of GK clustering technique. In our present work we have tried to show the effect of random initialization of the membership values on the performances of both FCM and KFCM clustering techniques. In addition to this we have tried to show the variation of the performance of KFCM clustering technique with different values of the adjustable parameter of its vector norm. Using Partition Coefficient (PC) and Clustering Entropy (CE) as validity indices ([4], [11]) we have tried to make a comparison of the performances of these two clustering techniques.

In Section-II, we shall provide the steps of FCM and KFCM algorithms. In Section-III, we explain our present work. The results and analysis of our work have been given in Section-IV. Finally we put the conclusions in Section-V.

### II. THE ALGORITHMS

The basic task of a clustering technique is to divide  $n$  patterns, where  $n$  is a natural number, represented by vectors in a  $p$ -dimensional Euclidean space, into  $c$ ,  $2 \leq c < n$ , categorically homogeneous subsets which are called clusters. Let the data set be

$$X = \{x_1, x_2, \dots, x_n\}, \text{ where } x_k = \{x_{k1}, x_{k2}, \dots, x_{kp}\}, k = 1, 2, 3, \dots, n.$$

Each  $x_k$  is called a feature vector and  $x_{kj}$  where  $j=1, 2, \dots, p$  is the  $j^{\text{th}}$  feature of the  $k^{\text{th}}$  feature vector. A partition of the dataset  $X$  into clusters is described by the membership functions of the elements of the cluster. Let  $S_1, S_2, \dots, S_c$  denote the clusters with corresponding membership functions

$\mu_{S_1}, \mu_{S_2}, \dots, \mu_{S_c}$ . A  $c \times n$  matrix containing the membership values of the objects in the clusters

$\tilde{U} = [\mu_{S_i}(x_k)]_{i=1,2,\dots,c,k=1,2,\dots,n}$  is a fuzzy  $c$ -partition if it satisfies the following conditions

$$\sum_{i=1}^c \mu_{S_i}(x_k) = 1 \tag{1}$$

for each  $k=1, 2, \dots, n$ .

$$0 \leq \sum_{k=1}^n \mu_{S_i}(x_k) \leq n \tag{2}$$

for each  $i=1, 2, \dots, c$ .

Condition (1) says that each feature vector  $x_k$  has its total membership value 1 divided among all clusters. Condition (2) states that the sum of membership degrees of feature vectors in a given cluster does not exceed the total number of feature vectors.

In Section-A and Section-B we provide respectively the steps of FCM and KFCM algorithms.

### A. *Bezdek's FCM Algorithm*

Step 1: Choose the number of clusters,  $c, 2 \leq c < n$ , where  $n$  is the total number of feature vectors. Choose  $m, 1 \leq m < \alpha$ . Define the vector norm  $\| \cdot \|$  (generally defined by the Euclidean distance) i.e.

$$\|x_k - v_i\| = \sqrt{\sum_{j=1}^p (x_{kj} - v_{ij})^2} \tag{3}$$

where  $x_{kj}$  is the  $j^{\text{th}}$  feature of the  $k^{\text{th}}$  feature vector, for  $k=1, 2, \dots, n; j=1, 2, \dots, p$  and  $v_{ij}, j$ -dimensional centre of the  $i^{\text{th}}$  cluster for  $i=1, 2, \dots, c; j=1, 2, \dots, p; n, p$  and  $c$  denote the total number of feature vector, no. of features in each feature vector and total number of clusters respectively.

Choose the initial fuzzy partition (by putting some random values)

$$U^{(0)} = [\mu_{S_i}^{(0)}(x_k)]_{1 \leq i \leq c, 1 \leq k \leq n}$$

Choose a parameter  $\epsilon > 0$  (this will tell us when to stop the iteration). Set the iteration counting parameter  $l$  equal to 0.

Step 2: Calculate the fuzzy cluster centroids  $\{v_i^{(l)}\}_{i=1,2,\dots,c}$  given by the following formula

$$v_i^{(l)} = \frac{\sum_{k=1}^n (\mu_{S_i}^{(l)}(x_k))^m x_k}{\sum_{k=1}^n (\mu_{S_i}^{(l)}(x_k))^m} \tag{4}$$

for  $i=1, 2, \dots, c; k=1, 2, \dots, n$ .

Step 3: Calculate the new partition matrix (i.e. membership matrix)

$$U^{(l+1)} = [\mu_{S_i}^{(l+1)}(x_k)]_{1 \leq i \leq c, 1 \leq k \leq n},$$

where

$$\mu_{S_i}^{(l+1)}(x_k) = \frac{1}{\sum_{j=1}^c \left( \frac{\|x_k - v_i^{(l)}\|}{\|x_k - v_j^{(l)}\|} \right)^{\frac{2}{m-1}}} \tag{5}$$

for  $i=1, 2, \dots, c$  and  $k=1, 2, \dots, n$ .

If  $x_k = v_i^{(l)}$ , formula (5) cannot be used. In this case the membership function is

$$\mu_{S_i}^{(l+1)}(x_k) = \begin{cases} 1 & \text{if } k=i \\ 0 & \text{if } k \neq i, i=1, 2, \dots, c \end{cases}$$

Step 4: Calculate

$$\Delta = \|U^{(l+1)} - U^{(l)}\| \tag{6}$$

If  $\Delta > \epsilon$ , repeat steps 2, 3 and 4. Otherwise, stop at some iteration count  $l^*$ .

### B. *The KFCM Algorithm*

Kernelized Fuzzy C-Means (KFCM) algorithm ([5], [6]) has been derived from traditional FCM of Bezdek [4] by changing the vector norms, fuzzy cluster centroids and the partition matrix as described in the following.

In step1 of KFCM the vector norm is defined by Gaussian Radial Basic Function (GRBF) (see equation (7)) instead of Euclidean distance (see equation (3)) of FCM.

$$k(x_r, v_i) = \exp\left(\frac{-\|x_r - v_i\|^2}{\sigma^2}\right) \quad (7)$$

In step2 of KFCM the fuzzy cluster centroids are calculated using equation (8) instead of equation (4) of FCM.

$$v_i^{(l)} = \frac{\sum_{r=1}^n (\mu_{s_i}^{(l)}(x_r))^m k(x_r, v_i) x_r}{\sum_{r=1}^n (\mu_{s_i}^{(l)}(x_r))^m k(x_r, v_i)} \quad (8)$$

for  $i = 1, 2, \dots, c$ ;  $r = 1, 2, \dots, n$ .

In step 3 of KFCM the new partition matrix is calculated by using equation (9) instead of equation (5) of FCM.

$$\mu_{s_i}^{(l+1)}(x_r) = \frac{(1 - k(x_r, v_i^{(l)}))^{-\frac{1}{m-1}}}{\sum_{j=1}^c (1 - k(x_r, v_j^{(l)}))^{-\frac{1}{m-1}}} \quad (9)$$

for  $i=1,2,\dots,c$  and  $r=1,2,\dots,n$ .

Remaining steps and calculations of KFCM are kept same as those of FCM.

### III. OUR PRESENT WORK

In our present work we have executed FCM three (03) different times on the same dataset (see Table-I) by initializing the membership values randomly in each execution. With this we have tried to show the effect of random initialization of the membership values on the performances of FCM (see Fig. 3 & Fig. 4). Next, we have executed KFCM several times by changing the value of  $\sigma$  of its vector norm (keeping the initialization of the membership values fixed) on the same dataset (see Table-I). The same experiment has been repeated using the same set of values of  $\sigma$  on the same dataset thrice with the three different sets of initializations of membership values which had been used in FCM. With this we have tried to show two aspects – first, there is significant variation of the performance of KFCM when the value of  $\sigma$  changes (see Fig. 1 & Fig. 2), secondly the effect of random initialization of the membership values on the performance of KFCM (see Fig. 1 & Fig. 2).

For the purpose of measuring the performances of FCM and KFCM we have used two validity measures namely Partition Coefficient (PC) and Clustering Entropy (CE) ([4], [11]). The mathematical formulae of these two validity measures have been given in the following.

- (a) Partition Coefficient (PC): measures the overlapping between clusters.

$$PC(c) = \frac{1}{n} \sum_{i=1}^c \sum_{j=1}^n (\mu_{ij})^2$$

- (b) Clustering Entropy (CE): measures the fuzziness of the cluster partition.

$$CE(c) = \frac{1}{n} \sum_{i=1}^c \sum_{j=1}^n \mu_{ij} \log(\mu_{ij})$$

TABLE I  
DATA SET OF INDIVIDUAL DIFFERENCES OF FIFTY (50) FEATURE VECTORS WITH DIMENSION  
(FEATURE) THREE (03).

FV	IQ	AM	SA	FV	IQ	AM	SA
1	91	18	55	26	110	18	55
2	85	16	40	27	100	16	40
3	120	19	74	28	100	18	75
4	90	18	75	29	70	14	30
5	92	17	74	30	105	17	55
6	82	17	55	31	79	14	35
7	95	19	75	32	80	15	34
8	89	18	74	33	125	20	75
9	96	19	75	34	100	19	75
10	90	17	55	35	125	19	85
11	97	16	54	36	80	18	60
12	125	21	74	37	85	18	70

<b>13</b>	100	19	75	<b>38</b>	145	25	90
<b>14</b>	90	17	54	<b>39</b>	80	18	74
<b>15</b>	100	18	84	<b>40</b>	92	17	55
<b>16</b>	95	19	75	<b>41</b>	120	18	70
<b>17</b>	130	23	85	<b>42</b>	145	30	80
<b>18</b>	130	19	75	<b>43</b>	95	18	50
<b>19</b>	90	17	55	<b>44</b>	80	16	36
<b>20</b>	91	17	56	<b>45</b>	90	17	55
<b>21</b>	140	22	82	<b>46</b>	115	23	84
<b>22</b>	92	18	75	<b>47</b>	100	18	80
<b>23</b>	101	18	55	<b>48</b>	80	14	35
<b>24</b>	85	16	54	<b>49</b>	105	19	75
<b>25</b>	97	19	54	<b>50</b>	120	21	74

FV: feature Vector, IQ: Intelligent Quotient, AM: Achievement motivation, SA: Social adjustment.

We have obtained the best performance and average performance of KFCM out of different choices of  $\sigma$  for three different sets of random initializations of the membership values and compared these with the performance of FCM (see Fig. 3 & Fig. 4). We have obtained the mean performance of FCM out of the three different sets of initializations of membership values. We have also calculated the mean of best performances and the mean of average performances of KFCM out of the three different sets of initializations of membership values. Finally we have compared these mean performances thus obtained (see Fig. 5 & Fig. 6).

#### IV. RESULTS AND ANALYSIS

In this section we place the results obtained in our work along with its analysis. In Fig. 1 we see the variation of the validity index PC obtained through KFCM against different values of  $\sigma$  of Gaussian Radial Basic Function. Again with three (03) different random initializations of the membership values Fig. 1 shows three (03) different lines of variations of PC of KFCM against the same set of values of  $\sigma$ . It shows that the performance of KFCM is maximum (i.e. the value of PC is maximum) when  $\sigma = 76$ ,  $\sigma = 71$  and  $\sigma = 19$  in first, second and third random initialization of membership values respectively.

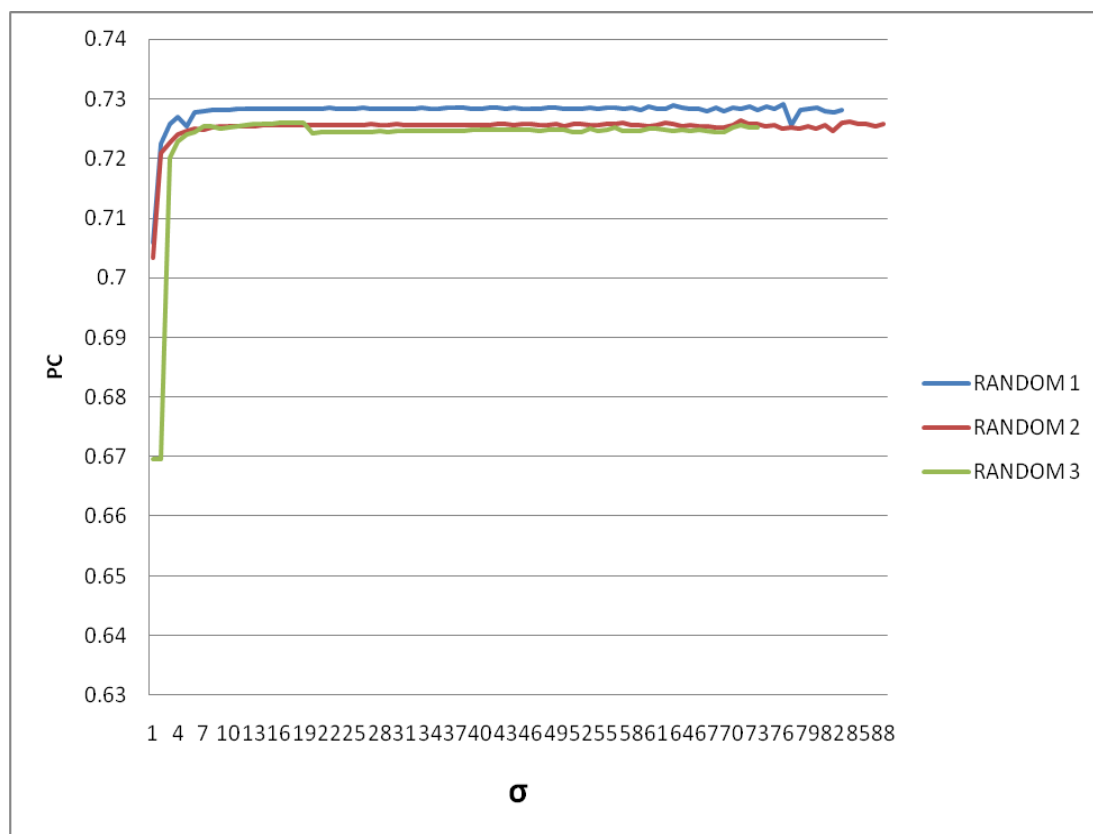


Fig. 1 Variation of PC of KFCM against different values of  $\sigma$  as well as variation of PC for three (03) different random initializations.

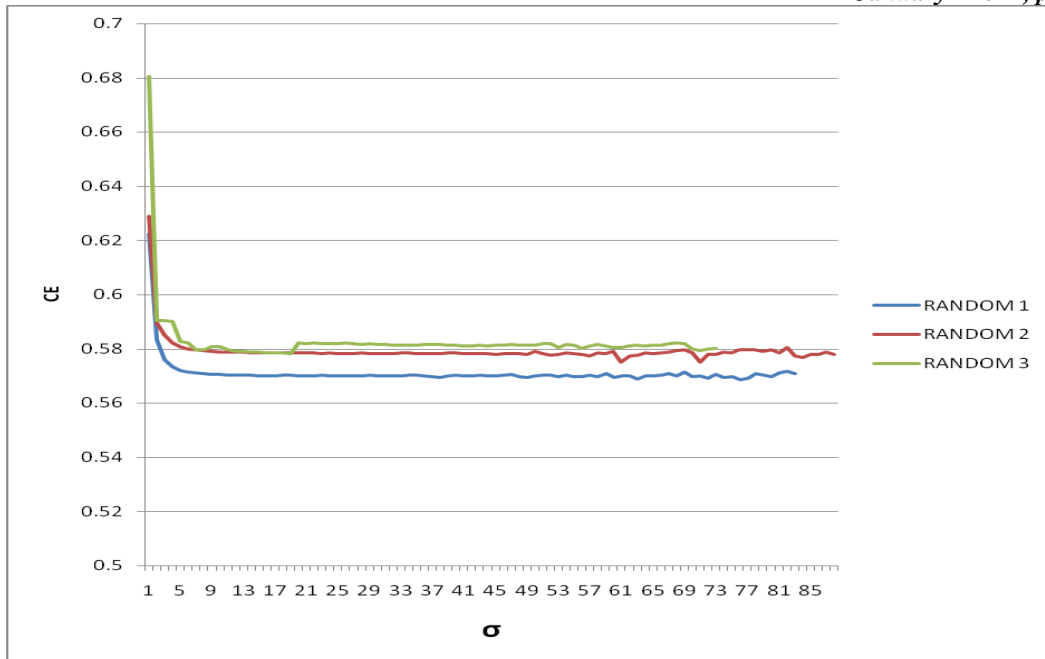


Fig. 2 Variation of CE of KFCM against different values of  $\sigma$  as well as variation of CE for three (03) different random initializations.

Fig. 2 shows the variation of the validity index CE obtained through KFCM against different values of  $\sigma$ . It also shows three (03) different lines of variations of CE of KFCM against the same set of values of  $\sigma$  when the membership values are randomly initialized for three different times. We see here that the performance of KFCM is maximum (i.e. the value of CE is minimum) when  $\sigma = 76$ ,  $\sigma = 71$  and  $\sigma = 19$  in first, second and third random initialization of membership values respectively.

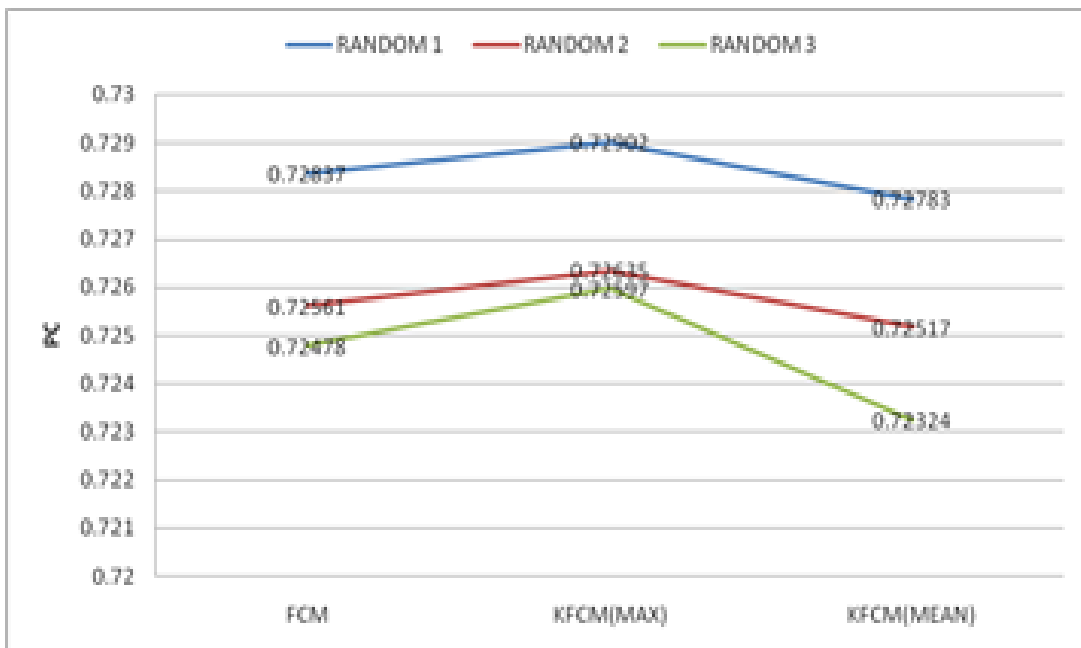


Fig. 3 Comparison of FCM, KFCM with PC of FCM, maximum PC of KFCM and mean PC of KFCM for three (03) different random initializations.

In Fig. 3 by KFCM(MAX) and KFCM(MEAN) we mean respectively the maximum and mean of the values of PC against different values of  $\sigma$  obtained through KFCM. It is evident in Fig. 3 that in every random initialization the performance is maximum in KFCM (MAX) followed by FCM and KFCM (MEAN). Similarly in Fig. 4 we see that in every random initialization the performance is maximum (i.e. CE is minimum) in KFCM (MIN) followed by FCM and KFCM (MEAN). In Fig. 5 we have visualized the respective mean values of PC of FCM, KFCM(MAX) and KFCM(MEAN) of three (03) different random initializations. The best performance is reflected in KFCM(MAX). Similar to it Fig. 6 shows the best performance in KFCM (MIN) (i.e. the validity measure CE is minimum).

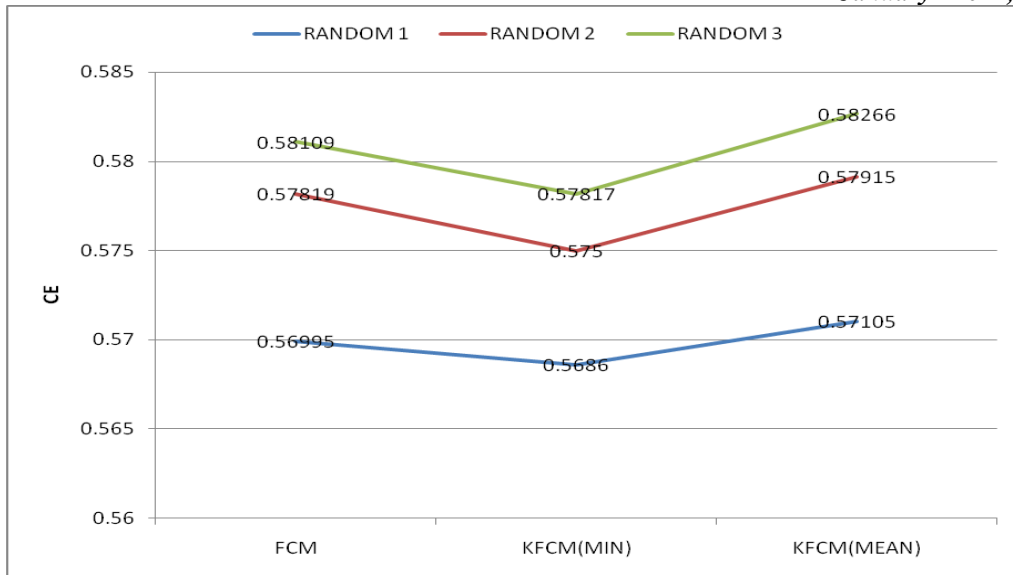


Fig. 4 Comparison of FCM, KFCM with CE of FCM, minimum CE of KFCM and mean CE of KFCM for three (03) different random initializations.

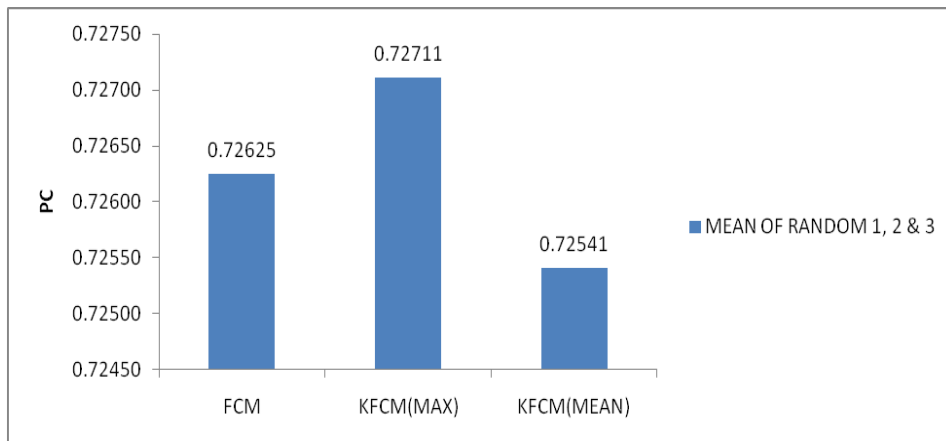


Fig. 5 Comparison of FCM, KFCM with mean PC of FCM, mean of maximum values of PC of KFCM and mean of mean values of PC of KFCM of three (03) different random initializations.

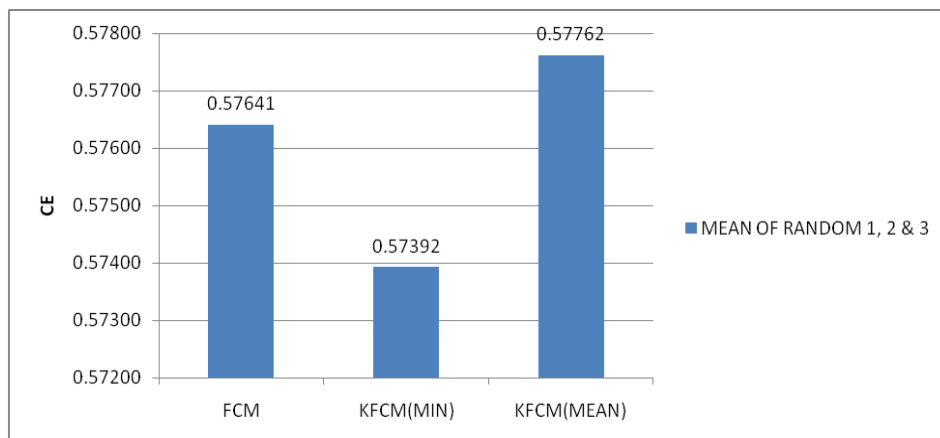


Fig. 6 Comparison of FCM, KFCM with mean CE of FCM, mean of minimum values of CE of KFCM and mean of mean values of CE of KFCM of three (03) different random initializations.

## V. CONCLUSIONS

The performance of FCM varies significantly with the randomly initialized membership values. KFCM is an attempt to improve the performance of FCM. Like FCM its performance also varies with the randomly initialized membership values. In addition to this the performance of it also depends on the value of  $\sigma$  of Gaussian Radial Basic Function used in

it as a distance function instead of Euclidian distance of FCM. The best choice of  $\sigma$  in KFCM provides better performance than FCM. However with random value of  $\sigma$  the performance of KFCM may not always be better than that of FCM.

## REFERENCES

- [1] L. A. Zadeh, "Fuzzy Sets", *Information and Control*, Vol.8, Issue 3, pp. 338-353, 1965.
- [2] H. K. Baruah, "Towards Forming a Field of Fuzzy Sets", *International Journal of Energy, Information and Communications*, Vol. 2, Issue 1, pp. 16-20, 2011.
- [3] H. K. Baruah, "The Theory of Fuzzy Sets: Beliefs and Realities", *International Journal of Energy, Information and Communications*, Vol. 2, Issue 2, pp. 1-22, 2011.
- [4] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, 1981.
- [5] G. A. Carl, "Fuzzy clustering and fuzzy merging algorithm", CS-UNR-101 Tech. Rep. , 1999.
- [6] K. R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf, "An introduction to kernel-based learning algorithms", *IEEE Trans Neural Networks*, Vol. 12, Issue 2, pp. 181—202, 2001.
- [7] R. A. Derrig, and K. M. Ostaszewski, "Fuzzy Techniques of Pattern Recognition in Risk and Claim Classification", *Journal of Risk and Insurance*, Vol. 62, Issue 3, pp. 447-482, 1995.
- [8] S. Das, and H. K. Baruah, "Application of Fuzzy C-Means Clustering Technique in Vehicular Pollution", *Journal of Process Management – New Technologies*, Vol. 1, Issue 3, pp. 96-107, 2013
- [9] S. Das, and H. K. Baruah, "A Comparison of Two Fuzzy Clustering Techniques", *Journal of Process Management – New Technologies*, Vol. 1, Issue 4, pp. 1-15, 2013.
- [10] D. E. Gustafson, and W. C. Kessel, "Fuzzy clustering with a fuzzy covariance matrix", in *Proc. IEEE CDC*, 1979, San Diego, CA, USA, p. 761–766.
- [11] A. M. Bensaid, L. O. Hall and J. C. Bezdek, "Validity-guided (re) clustering with applications to image segmentation", *IEEE Trans. On Fuzzy object*, Vol. 2, Issue 2, pp. 112-123, 1996.