SciencePG
Science Publishing Group

A Novel Approach Canberra Measure Minimal Spanning Tree Using Fuzzy C-Means Based on Gaussian Function for Image Data Mining

Senthil[1, *], Nithya[2], Bhuvaneswari[2]

[1]Department of Economics and Statistics, Government of Tamilnadu, DRDA, Dindigul, India

[2]Department of Mathematics, Mother Teresa Women's University, Kodaikanal, India

Email address:
senthilsnc83@gmail.com (Senthil)
[*]Corresponding author

Abstract: Clustering analysis has been an emerging research issue in data mining due to its variety of applications. In recently, mathematical algorithm supported automatic segmentation system plays an important role in clustering of images. The fuzzy c-means clustering is a method of cluster analysis which aims to partition n data points into k-clusters. The conventional FCM-based algorithm considers no spatial content information, which means it sensitive to noise. Unsupervised techniques need to be employed, which can be based on minimal spanning tree generated by comparing spatial neighbourhood information, the MST based clustering algorithms have been widely used due to their ability to detect clusters with irregular boundaries. We propose an automatic fuzzy c-means initialization algorithm based on Canberra distance minimal spanning tree for the purpose of segmentation of medical images, where vertices and edges are labelled with multi-dimensional vectors. A Canberra distance measure based, construct the minimal spanning tree clustering algorithm. An efficient method for calculating membership and updating prototypes by minimizing the new objective function of Gaussian based fuzzy c-means. The algorithm uses a new cluster validation criterion based on the geometric property of data partition of the dataset in order to find the proper number of cluster at each level. In this algorithm to apply medical images to reduce the inhomogeneity and allow the labelling of a pixel to be influenced by the labels in its immediate neighbourhood and reduces the time complexity and better clustering results than the existing traditional minimal spanning tree algorithm. The performance of proposed algorithm has been shown with random data set, partition coefficient and validation function are used to evaluate the validity of clustering and then new cluster separation approach to optimal number of clustering. Also this paper compares the results of proposed method with the results of existing basic fuzzy c-means.

Keywords: Fuzzy C-Means, Gaussian Function, Lagrange Multiplier, Canberra Distance, Minimal Spanning Tree, Cluster Separation, Partition Coefficient, Validation Function

1. Introduction

Clustering is one of the important tools for data analysis. It can divide an unlabeled dataset into several subsets according to some criteria to ensure similar samples to be in the same subset and dissimilar samples to be different subset. Fuzzy c-means algorithm is a widely used clustering algorithm in the field of machine learning. It was proposed by Bezdek et al in 1984 [1]. By introducing the fuzzy membership matrix, the fuzzy c-means algorithm allows data points to belong to multiple classes according to their fuzzy membership degree. We choose the class with the highest value of the current data point in the fuzzy membership matrix as the final clustering result. This method solves the problem of clustering overlap of traditional hard clustering algorithms [2]. Fuzzy clustering, as a soft segmentation method, has been widely studied and successfully applied to image segmentation [2-4, 6-7]. MR image signals have highly affected by shacking of

patients body and patients motion. So the medical MRI is seriously affected and it has improper information about the anatomic structure. Hence the segmentation of medical images is an important one before it to go for treatment planning for proper diagnosis. Initially segmentation was made manually; but manual segmentation is more difficult, time -consuming and costly. Automatic brain tumor segmentation from MR images which is not an easy task that involves various disciplines covering image analysis, mathematical algorithms [8-10] and etc.

In recent world, many of the segmentation methods are based on the unsupervised clustering algorithms. Unsupervised clustering is a process for ordering objects in such a way that samples of the identical group are more similar to one another than samples belonging to different groups. Currently fuzzy c-means of fuzzy clustering plays main role in unsupervised clustering method for segmenting medical images [11]. In Noordam, proposed a geometrically guided FCM algorithm based on a semi-supervised FCM technique for multivariate image segmentation. In their work, the geometrical condition information of each pixel is determined by taking into account the local neighbourhood of each pixel [5].

The main disadvantages of fuzzy clustering technique are its need for a large amount of time to converge and it is more sensitive to the noise and outliers in the data, because of squared-norm to measure similarity between prototypes and data points. To cluster more general dataset, lots of algorithms have been proposed by replacing the squared-norm with other similarity measures. A recent development is to use kernel method to construct the kernel versions of FCM algorithm, KFCM for clustering the incomplete data and medical image segmentation was proposed [12-13]. However a disadvantage of KFCM in segmentation of medical images it is not considered about any spatial information in image context; which makes it compute the neighbourhood term in each step, which is very time consuming. Although the above methods are claimed to be robust to noise, they are confronted with the problem of selecting the parameters that control the role of the spatial constraints. However, FCM has two serious shortcomings, Firstly, it easily falls into local minima, Secondly, it is necessary to specify the number of clusters and the algorithm is very sensitive to the initial center [14-15]. The graph data structure is being considered as a suitable mathematical tool to model the inherent relationship among data. Simple linked data are generally modeled using simple graph $G = (V, E)$, where V is the set of vertices representing key concepts or entities and $E \subseteq V \times V$ is the set of links between the vertices representing the relationships between the concepts or entities. There are many complex data link online social networks, where is characterized by a set of features and multiple relationships exist between an entity pair, such data, the concept of graph can be used where in each vertex is represented by an n-dimensional vector. One of the important tasks related to graph data analysis is to decompose a given graph into multiple cohesive subgraphs, called clusters, based on some common properties. The clustering is an unsupervised learn process to identify the underlying structure of data, which is generally based on some similarity/ distance measures between data elements.

Graph clustering is special case of clustering process which divides an input graph into a number of connected components such that intra-component edges are maximum and inter- components edges are minimum. Each connected components is called a cluster [16]. Graph clustering techniques on the minimum spanning tree (MST) of a weighted graph is minimum weight spanning tree of that graph with the classical MST algorithms [17, 18, 19] and the cost of constructing a minimal spanning tree is $O(m \log n)$, where m is the number of edges in the graph and n is the number of vertices. Our study on these methods have led us to believe that all these applications have used the MSTs in some heuristic way; eg cutting long edges to separate clusters without fully exploring their power and understanding their rich properties related to clustering. Geometric notion of centrality are closely linked to facility location problem. Since similarity/ distance measure is the key requirement for any clustering algorithm, In this paper, we have proposed a new weighted distance measure based on weighted Euclidean norm to calculate the distance between the vertices and the algorithm is initialized by a given kernel function using minimal spanning tree based FCM algorithm, which helps to speed up the convergence of the algorithm.

The rest of the paper is organized as follows section 2 presents a Canberra distance measures and Cluster Separation using CMST clustering algorithm. Section 3 presents the kernel based FCM and Validation function. Section 4 presents the experimental results using random data show that the proposed method can achieve comparable results to those from many derivatives of efficient initialized by a given kernel function using minimal spanning tree based FCM algorithm and effective and more robust to reduce the noise and outliers.

2. Canberra Measure Based minimal Spanning Tree Algorithm

In this section, we proposed Canberra distance measures for construct the minimal spanning tree.

2.1. Canberra Distance Measure MST

Given the grayscale point set D, the hierarchical methods starts by constructing a minimal spanning tree (MST) from the points in D. In $x = (x_1, x_2, \ldots \ldots x_n)^T$ and $y = (y_1, y_2, \ldots \ldots y_n)^T$ are two points of a MST and $e(x, y)$ is an edge between x and y then the Canberra distance between x and y is denoted by $d(x, y)$ and calculated using equation (1) [20],

$$d(x, y) = \frac{1}{K} \sum_{i=1}^{n} \frac{|x_i - y_i|}{x_i + y_i} \qquad (1)$$

where K be the number of non-zero pairs.

2.2. Cluster Separation (CS)

The definition of CS between cluster centers is given by the following

$$CS = \frac{E_{\min}}{E_{\max}} \qquad (2)$$

where $E_{\min}$ is the maximum length edge in the MST, which represents two centroids that are at maximum separation and $E_{\min}$ is the length edge in the MST, which represents two centroids that are nearest to each other. Then the CS represents the relative separation of the centroids. The value of CS ranges from 0 to 1. A low value of CS means that the two centroids are too close to each other and the corresponding MST Separation not valid. A high CS value means the MST separation of the data is even and valid. If the CS is greater than the threshold, the MST partition of the dataset is valid. Then, we increase the number of cluster by and test the CS again. This process continuous until the CS is smaller than the threshold. The value setting of the threshold for the CS will be practical and is dependent on the dataset. The higher the value of the threshold the smaller the number of clusters would be, generally the value of the threshold will be $\succ 0.8$ [21].

2.3. Canberra Distance Based Minimal Spanning Tree Algorithm

Algorithm: CMST
Input: Data points
Output: optimal number of cluster centers
Let e1 be an edge in the CMST1 constructed from data points
Let e2 be an edge in the CMST2 constructed from C.
Let S_T be the set of disjoint subtrees of CMST1.
1. Create a node v, for each data points.
2. Compute the edge weight using equation (1)
3. Construct an CMST1 from 2
4. $S_T = \varphi$, $n_c = 1$, $C = \varphi$.
5. Repeat.
6. For each $e1 \in CMST1$.
7. Current longest edge e remove e1 from CMST1.
8. $S_T = S_T \bigcup \{T'\}$ // T' is new disjoint subtrees (regions).
9. $n_c = n_c + 1$.
10. Compute the center c_i of T_i using average of points.
11. $C = \bigcup_{T_i} \in S_T \{c_i\}$.
12. Compute the edge weight using equation (1)
13. Construct an CMST2 T from C.
14. $E_{\min}$ =get-min length edge.
15. $E_{\max}$ =get-max length edge.
16. $CS = \frac{E_{\min}}{E_{\max}}$
17. Until CS < 0.8
18. Merge the closest neighbour from CMST2.
19. Update the clusters points, repeat step 12 to step 18.
20. Finally we obtain the cluster centers.

3. Formulation of Proposed Kernel Function Induced FCM Based on Gaussian Function

3.1. Fuzzy C-Means Algorithm

This method was first introduced by Dunn in [22] and improved by Bezdek in [23]. It is based on minimization of the following objective function:

$$J_m(U,V) = \sum_{i-1}^{N} \sum_{j=1}^{C} u_{ij}{}^m \left\| x_i - c_j \right\|^2 \quad 1 < m < \infty \qquad (3)$$

where m is any real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster j, x_i is the ith of p - dimensional measured data, c_j is the p -dimension center of the cluster and $\| * \|$ is any norm expressing the similarity between any measured data and center.

Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership u_{ij} and the cluster centers c_j by:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left[\frac{\left\| x_i - c_j \right\|}{\left\| x_i - c_k \right\|} \right]^{\frac{2}{m-1}}} \qquad (4)$$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}{}^m x_i}{\sum_{i=1}^{N} u_{ij}{}^m} \qquad (5)$$

This iteration will stop when

$$\max_{ij} \left\{ \left| u_{ij}{}^{k+1} - u_{ij}{}^k \right| \right\} < \in ,$$

where $\in$ is a termination criterion between 0 and 1, whereas k is the iteration steps. This procedure converges to a local minimum or a saddle point of J_m.

3.2. Kernel Function Induced FCM Algorithm

Recently, the powerful technique of kernel function using learning machines were proposed and found to have successful applications such as signal processing pattern recognition and image processing etc. Kernel method often studies and employs a high-dimensional feature space S for having nonlinear classification boundaries. For this a mapping is given below;

$\varphi : R^p \rightarrow S$ is used where by an object x is mapped into S:

$$\varphi (x) = (\varphi_1(x), \varphi_2(x), ...)$$

As x is the p-dimensional vector, $\varphi(x)$ may have the infinite dimension. In the nonlinear classification method, an explicit form of $\varphi(x)$ is unavailable, but the inner product is defined by:

$$K(x,y) = \langle \varphi(x), \varphi(y) \rangle$$

The function $K(x,y)$ is called a kernel function and we assume this known function, as Gaussian radial basis function:

$$K(x,y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}}$$

This paper proposes an efficient weighted MST based FCM by introducing kernel function that allows the clustering of objects to be more reasonable. The modified proposed objective function is given by

$$J_m(U,V) = \sum_{k=1}^{N} \sum_{i=1}^{C} u_{ik}{}^m \| \varphi(x_k) - \varphi(v_i) \|^2 \qquad (6)$$

where φ stands as map and the distance function can be expressed using in product space as

$$\| \varphi(x_k) - \varphi(v_i) \|^2 = \langle \varphi(x_k), \varphi(x_k) \rangle + \langle \varphi(v_i), \varphi(v_i) \rangle - 2\langle \varphi(x_k), \varphi(v_i) \rangle$$

To obtain kernel induced FCM based Gaussian function the distance function can be modified as

$\| \varphi(x_k) - \varphi(v_i) \|^2 = G(x_k, x_k) + G(v_i, v_i) - 2G(x_k, v_i)$ where $k = 1,2,3..........N$ and $i = 1,2,3...........C$.

Let us express $G(x_k, v_i)$, between pixel x_k and v_i as the product of a feature similarity term and spatial proximity term:

$$G(x_k, v_i) = \exp\left(\frac{-\|x_k - v_i\|^2}{2\sigma_X{}^2} \right) * \exp\left(\frac{-\|I(x_k) - I(v_i)\|^2}{2\sigma_I{}^2} \right) \quad (7)$$

where σ is a parameter which can be adjusted by users.

Using the above expression, we obtain $G(x_k, x_k) = 1$ and $G(v_i, v_i) = 1$, so the distance function can be rewritten as

$$\| \varphi(x_k) - \varphi(v_i) \|^2 = 2(1 - G(x_k, v_i)) \qquad (8)$$

Substituting kernel induced Gaussian function based FCM is given by

$$J_m(U,V) = 2\sum_{k=1}^{N} \sum_{i=1}^{C} u_{ik}{}^m (1 - G(x_k, v_i)) \qquad (9)$$

3.3. Obtaining Membership

To obtain equation for calculating membership we minimizing the objective function

$$J_m(U,V) = 2\sum_{k=1}^{N} \sum_{i=1}^{C} u_{ik}{}^m (1 - G(x_k, v_i)) \qquad (10)$$

subject to the constraints $\sum_{i=1}^{C} u_{ik} = 1$

Therefore, the above objective function can be minimized using one Lagrangian multiplier:

$$J_m(U,V,\lambda) = 2\sum_{k=1}^{N} \sum_{i=1}^{C} u_{ik}{}^m (1 - G(x_k, v_i))) - \lambda\left(\sum_{i=1}^{C} u_{ik} - 1 \right) \quad (11)$$

where λ is a Lagrange multiplier.

To adjust $u_{ik} \& v_i$ for minimum J_m, we set to zero the derivative of $J_m(U,V,\lambda)$ with respect to u_{ik} *for* $m \succ 1$.

$$\frac{\partial J_m}{\partial u_{ik}} = 2m u_{ik}{}^{m-1} [1 - G(x_k - v_i)] - \lambda = 0$$

$$u_{ik} = \left(\frac{\lambda}{m} \right)^{1/m-1} \Big/ (1 - G(x_k, v_i))^{1/m-1}$$

To calculate λ, Substitute the above u_{ik} in the identity constraint for all values of k, we get following relation,

$$\frac{\partial J_m}{\partial v_i} = -2\sum_{k=1}^{N} u_{ik}{}^m (x_k - v_i) = 0$$

The minimum of J_m with respect to v_i was computed by taking the partial derivative of J_m equal to zero.

$$\left(\frac{\lambda}{m} \right)^{\frac{1}{m-1}} = \frac{1}{\sum_{i=1}^{C} \left[\frac{1}{(1 - G(x_k, v_i))} \right]^{\frac{1}{m-1}}}$$

So that v_i and u_{ik} can be calculated by the relation, we obtain

$$u_{ik} = \frac{1}{\sum_{j=1}^{C} \left[\frac{1 - G(x_k, v_i)}{1 - G(x_k, v_j)} \right]^{\frac{1}{m-1}}} \qquad (12)$$

$$v_i = \frac{\sum_{k=1}^{N} (u_{ik})^m G(x_k, v_i) x_k}{\sum_{k=1}^{N} (u_{ik})^m G(x_k, v_i)}. \qquad (13)$$

The kernel function based FCM algorithm iteratively optimizes J_m by continuous updating u_{ik} and v_i until the difference in successive u_{ik} values is very small $\leq \in$, where $\in$ is a small positive value between 0 and 1.

4. Efficient Kernel Induced FCM Based on Gaussian Function

The objective function (I) of the standard FCM algorithm does not take into account any spatial information, which means the clustering process is related to gray levels independently of the pixels of image segmentation. Therefore, the limitation makes FCM to be very sensitive to noise. The general principle of the technique presented in this paper is to incorporate the neighbourhood information into the FCM algorithm during classification.

4.1. Efficient KFCM Algorithm

Stage 1: Set the cluster centroids $\{v_i\}_{i=1}^c$ by using Canberra MST initialization method.
Stage 2: Compute the membership function using (12)
Stage 3: Update the cluster centroids using (13)
Stage 4: Estimate objective function using (10)
Stage 5: Go to stage (2)-(3), repeat until convergence. The termination criterion is as follows $|J_m - J_{m-1}| \prec \in$, where m is the iteration count, $\in$ is a small number that can be set by the user.

4.2. Validation Function Based on Feature Structures

Two representative functions for the fuzzy partition namely; Partition coefficient V_{pc} and Validation function V_p are used to evaluate the validity of clustering [3, 24].

$$V_{pc} = \frac{1}{N} \sum_{k=1}^{N} \sum_{i=1}^{C} u_{ik}^2 \qquad (14)$$

$$V_p = \frac{\sum_{k=1}^{N} \sum_{i=1}^{C} u_{ik}^2 \|x_k - v_i\|^2}{N \times \min\{\|v_i - v_j\|^2\}} \qquad (15)$$

The proposed efficient weighted MST obtained cluster centers; the EKFCM algorithm continues iteratively updates, membership and centroids with these values. When this improved, Efficient KFCM algorithm has converged, another defuzzification process takes place in order to convert the fuzzy partition matrix to a crisp partition matrix that is segmented.

5. Results and Discussion

This section describes some experimental results on random data, corrupted with noise to show the segmentation performance of the proposed method.

***Table 1.** Random data.*

Data				Data			
S. No	X	Y	Intensity	S. No	X	Y	Intensity
1	1.80	2.00	0.50	11	12.00	4.00	0.80
2	2.00	2.20	0.91	12	11.50	3.50	0.45
3	2.00	1.80	0.12	13	12.50	3.50	0.55
4	2.00	3.50	0.40	14	21.00	10.00	0.65
5	8.80	3.00	0.50	15	21.00	11.00	0.25
6	9.00	3.20	0.38	16	20.50	10.50	0.35
7	9.00	2.80	0.60	17	21.50	10.50	0.75
8	9.20	3.00	0.12	18	2.00	4.00	0.70
9	7.00	2.80	0.80	19	19.00	20.00	0.60
10	12.00	3.00	0.90	20	11.00	12.00	0.40

***Table 2.** Dissimilarity matrix-I.*

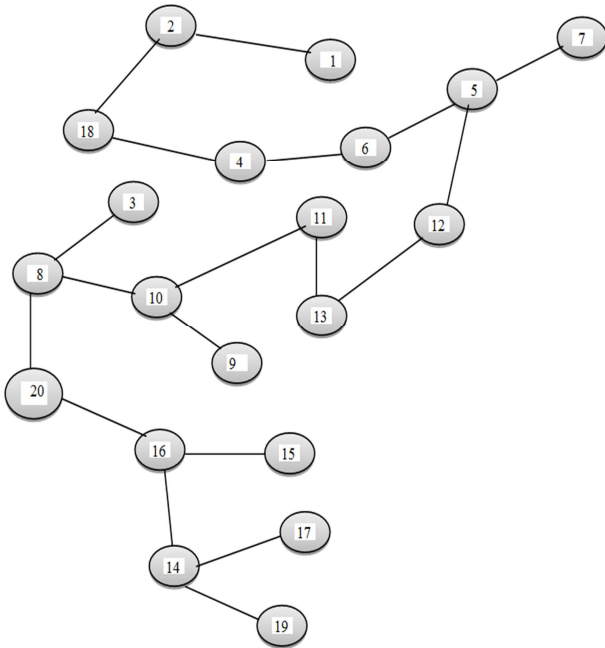| Co-ordinate			intensity												
S. No	x	y	I(v)	S. No	1	2	3	4	5	6	7	8	9	10	11
1	1.8	2.0	0.50	1	0.000	0.130	0.239	0.145	0.287	0.345	0.308	0.495	0.329	0.408	0.434
2	2.0	2.2	0.91	2		0.000	0.289	0.206	0.358	0.411	0.321	0.521	0.247	0.291	0.356
3	2.0	1.8	0.12	3			0.000	0.286	0.498	0.479	0.507	0.298	0.504	0.576	0.611
4	2.0	3.5	0.40	4				0.000	0.273	0.236	0.316	0.419	0.333	0.392	0.371
5	8.8	3.0	0.50	5					0.000	0.060	0.046	0.212	0.126	0.147	0.176
6	9.0	3.2	0.38	6						0.000	0.097	0.188	0.183	0.194	0.203
7	9.0	2.8	0.60	7							0.000	0.237	0.089	0.126	0.154
8	9.2	3.0	0.12	8								0.000	0.303	0.299	0.338
9	7.0	2.8	0.80	9									0.000	0.119	0.147
10	12.0	3.0	0.90	10										0.000	0.067
11	12.0	4.0	0.80	11											0.000
12	11.5	3.5	0.45	12											
13	12.5	3.5	0.55	13											
14	21.0	10.0	0.65	14											
15	21.0	11.0	0.25	15											
16	20.5	10.5	0.35	16											
17	21.5	10.5	0.75	17											
18	2.0	4.0	0.70	18											
19	19.0	20.0	0.60	19											
20	11.0	12.0	0.40	20											

Figure 1 shows a typical example of CMST1 constructed from point set (from Dissimilarity matrix), in which inconsistent edges are removed to create subtree (clusters/regions). Our algorithm finds the center of each

clusters, which will be useful in many applications.

Table 3. Canberra distance based minimal spanning tree edges.

S. No	Edges	Canberra measure	S. No	Edges	Canberra measure
1	(1,2)	0.130	11	(10,9)	0.119
2	(2,18)	0.140	12	(10,8)	0.299
3	(18,4)	0.113	13	(8,3)	0.298
4	(4,6)	0.236	14	(8,20)	0.409
5	(6,5)	0.060	15	(20,16)	0.145
6	(5,7)	0.046	16	(16,15)	0.067
7	(5,12)	0.088	17	(16,14)	0.112
8	(12,13)	0.047	18	(14,17)	0.036
9	(13,11)	0.091	19	(14,19)	0.141
10	(11,10)	0.067			

Figure 1. Canberra distance based Minimal spanning tree connected through points.

Generally in most of the clustering algorithm data points can be represented as dissimilarity matrix representation. It contains the distance values between the data points represented as lower or upper triangular matrix. Our Canberra distance based minimal spanning tree algorithm constructs CMST1 from the dissimilarity matrix is shown figure 1. First to identify the longest edge in the CMST1 to generate subtree (clusters). Table 3, the longest edge weight 0.409 connecting the data points 8 and 20 is find to be inconsistent one. By removing the inconsistent edge from the CMST1, data points in the CMST1 partitioned into two subtrees or clusters T_1 and T_2 namely $T_1 = \{1,2,3,4,5,6,7,8,9,10,11,12,13,18\}$ and $T_2 = \{14,15,16,17,19,20\}$. Secondly to find the center of T_1 and T_2 using average of points, these centers is connected and again another minimal spanning tree CMST2 is constructed. The minimum edge of CMST2 is $E_{min} = 0.750$ and the maximum edge of CMST2 is $E_{max} = 0.750$ then to

compute cluster separation value is 1. If the CS is greater than 0.8 then we conclude the subtrees or clusters created are well separated. Next to identify another longest edge weight from Table 3 is 0.299 connecting the data points 10 and 8 is finding to be inconsistent one. By removing the inconsistent edge from the CMST1, data points partitioned into three sub trees or clusters T_1 , T_2 and T_3 namely $T_1 = \{1,2,4,5,6,7,9,10,11,12,13,18\}$, $T_2 = \{14,15,16,17,19,20\}$ and $T_3 = \{3,8\}$. To compute the center of T_1, T_2 and T_3 using average of points, these centers is connected and again another minimal spanning tree CMST2 is constructed.

Table 4. Dissimilarity matrix-II.

	Cluster-I	Cluster-II	Cluster-III
Cluster-I	0.000	0.381	0.317
Cluster-II		0.000	0.611
Cluster-III			0.000

Table 5. Canberra distance based CMST2 edges.

S. No	Edges	Canberra measure
1	(1,3)	0.317
2	(1,2)	0.381

The minimum edge of CMST2 is $E_{min} = 0.317$ and the maximum edge of WMST2 is $E_{max} = 0.381$ then to compute cluster separation value is 0.832. If the CS is greater than 0.8 then we conclude the subtrees or clusters created are valid. Continuing this process, next to identify another longest edge weight from Table 3 is 0.298 connecting the data points 3 and 8 is finding to be inconsistent one. By removing the inconsistent edge from the CMST1, data points partitioned into three sub trees or clusters T_1 , T_2 , T_3 and T_4 namely $T_1 = \{1,2,4,5,6,7,9,10,11,12,13,18\}$, $T_3 = \{3\}$ and $T_4 = \{8\}$. To compute the center of T_1, T_2, T_3 and T_4 using average of points, these centers is connected and again another minimal spanning tree CMST2 is constructed.

Table 6. Dissimilarity matrix-III.

	Cluster-I	Cluster-II	Cluster-III	Cluster-IV
Cluster-I	0.000	0.381	0.508	0.267
Cluster-II		0.000	0.723	0.523
Cluster-III			0.000	0.298
Cluster-IV				0.000

Table 7. Canberra distance based CMST2 edges.

S. No	Edges	Canberra measure
1	(1,4)	0.267
2	(4,3)	0.298
3	(4,2)	0.523

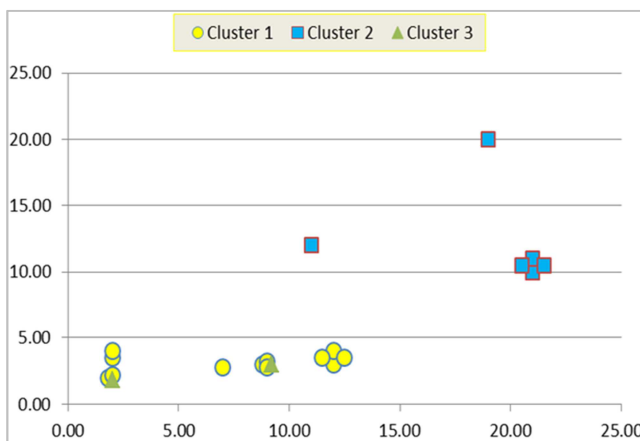The minimum edge of CMST2 is $E_{min} = 0.267$ and the maximum edge of CMST2 is $E_{max} = 0.523$ then to compute cluster separation value is 0.510 If the CS is less than 0.8 then we conclude the subtrees or clusters created are not valid. Finally CMST produces three cluster centers. Canberra minimal spanning tree algorithm creates three cluster centers

for the given data points. Then the center of the cluster and its convergence of standard FCM and EKFCM are determined under successive interactions of experiments using data points. The standard FCM algorithm and the numbers of updated centers are high under the objective function of Euclidean distance measures. This takes more iteration to converge the termination value of algorithm. With the new efficient objective function based kernel distance measure the termination value is achieved, with very less iteration and with much better performance in getting membership (Table 8) than standard FCM. Table 9 gives the number of iteration to achieve the results of cluster on the data points by standard FCM and EKFCM. It is clear from the final cluster, membership (Table 8), scatter diagram (Fig 2), that our proposed EKFCM is much faster than the standard FCM and the method is converged fast to terminate condition with less run time. To test the effectiveness of EKFCM, the weighted minimal spanning tree based FCM is used as center. This is done to find out the fuzzy membership and appropriate number of clusters. Thus, we have concluded the final optimal clusters formed as 3. This algorithm has also reduced the number of iterations. Best result is achieved by this measure fuzzy partition coefficient V_{pc} maximum and validation function V_p minimum (Table 10). The EKFCM clustering algorithm has the following membership value intimacy (Table 8).

Table 8. *Final membership of three clusters of EKFCM method and object allocation.*

Co-ordinate(x, y)			intensity				appropriate cluster
S. No	x	y	I(v)	Mem-1	Mem-2	Mem-3	
1	1.80	2.00	0.50	0.8542	0.0026	0.1432	1
2	2.00	2.20	0.91	0.9940	0.0022	0.0038	1
3	2.00	1.80	0.12	0.0094	0.0002	0.9905	3
4	2.00	3.50	0.40	0.5591	0.0057	0.4352	1
5	8.80	3.00	0.50	0.9462	0.0092	0.0446	1
6	9.00	3.20	0.38	0.5872	0.0259	0.3869	1
7	9.00	2.80	0.60	0.9955	0.0015	0.0030	1
8	9.20	3.00	0.12	0.0131	0.0021	0.9848	3
9	7.00	2.80	0.80	0.9912	0.0050	0.0037	1
10	12.00	3.00	0.90	0.9660	0.0316	0.0024	1
11	12.00	4.00	0.80	0.9545	0.0412	0.0043	1
12	11.50	3.50	0.45	0.8067	0.0536	0.1397	1
13	12.50	3.50	0.55	0.9191	0.0457	0.0352	1
14	21.00	10.00	0.65	0.0255	0.9738	0.0007	2
15	21.00	11.00	0.25	0.0100	0.9753	0.0147	2
16	20.50	10.50	0.35	0.0097	0.9849	0.0053	2
17	21.50	10.50	0.75	0.0382	0.9614	0.0004	2
18	2.00	4.00	0.70	0.9826	0.0035	0.0139	1
19	19.00	20.00	0.60	0.0016	0.9983	0.0001	2
20	11.00	12.00	0.40	0.1090	0.8430	0.0480	2

Figure 2. *Scatter diagram for CMST based FCM, final cluster three.*

Table 9. *Comparison of iteration count.*

	No. of iterations	No. of clusters
Standard FCM	16	3
IKFCM	3	3
CMSTFCM	2	3

Table 10. *Cluster validity function.*

	V_{pc}	V_p
Standard FCM	0.8372	0.2532
IKFCM	0.8423	0.2623
CMSTFCM	0.8705	0.1253

6. Conclusion

Thus, this paper has proposed a new Canberra measure minimal spanning tree algorithm and the efficient kernel induced fuzzy c-means based on Gaussian function for image data analyzing. The algorithm was formulated by introducing kernel function, Gaussian function and Lagrangian methods, with basic objective function of the FCM algorithm to have proper effective segmentation of the image data and to overcome the noise sensitiveness of conventional FCM clustering algorithm. The main contribution of this algorithm is to incorporate the spatial neighbourhood information into the standard FCM algorithm by CMST. It can be automatically decided in the algorithm based on the membership function of the center pixel and its neighbouring

pixels and also compares the results with standard FCM segmentation. It is clear from our comparison that EKFCM performed better than FCM. The results accuracy and validation functions used in measuring the efficiency in clustering algorithm with Gaussian. In future we will explore and test our proposed algorithm in various domains.

Acknowledgements

References

[1] Bezdek. J. C, Ehrlich. R and Full. W, FCM: the fuzzy c-means clustering algorithm, Computers & Geosciences, Vol. 10, No. 2-3, pp. 191-203(1984).

[2] Mac Queen. J, Some methods for classification and analysis of multivariate observations, In Proceedings of the 5th Berkeley symposium on mathematical statistics and probability, University of California press, USA, pp. 14(1967).

[3] Bezdek J. C., Hall L. O., Clarke L. P, Review of MR image segmentation techniques using pattern recognition, medical physics 20(4), pp. 1033-1048(1993).

[4] Ferahta N., Moussaoui A., Benmahammed K., Chen V, New fuzzy clustering algorithm applied to RMN image segmentation, international journal of soft computing 1(2), pp. 137-142 (2006).

[5] Noordam J. C., Van Den Broek W. H. A. M., Buydens L. M. C, Geometrically guided fuzzy C-means clustering for multivariate image segmentation, In. proceedings15-th International conference on pattern recognition, Vol. pp. 462-465(2000).

[6] Mahipal singh choudhry and Rajiv Kapoor, A novel fuzzy energy based level set method for medical image segmentation, Cogent engineering, vol. 5(2018).

[7] Perumal K and Latha C, Probability based fuzzy c-means for image segmentation, International Journal of Pure and Applied Mathematics, vol. 18 (17), pp. 779-789(2018).

[8] Hall L. O., Bensaid A. m., Clarke L. P., Velthuizen R. P., Silbiger M. S., Bezdek J. C, A comparison of Neural network and Fuzzy clustering Techniques in segmenting magnetic Resonance images of the Brain, IEEE Transactions, Neural networks 3(5), pp. 672-682(1992).

[9] Xue J. H., Pizurica A., Philips W., Kerre E., Walle R. V., Lemahieu I, An integrated method of Adaptive Enhancement for unsupervised segmentation of MRI Brain images, Pattern Recognition Letters 24(15), pp. 2549-2560(2003).

[10] Krishnan N., Nelson Kennedy Babu C. V., Joseph rajapandian V., Richard Devarajan, A Fuzzy image segmentation using feed forward Neural networks with supervised Learning, In. Proceedings of the International conference on cognition and recognition, pp. 396-402.

[11] Moussaoui A., Benmahammed K., Ferahta N., Chen V, A new MR Brain image segmentation using an optimal semi supervised Fuzzy C-means and pdf Estimation, Electronic letters on computer vision and image Analysis 5(4), pp. 1-11(2005).

[12] Aruna Kumar SV and Harish BS, A novel fuzzy clustering based system for medical image segmentation, Int. J. Computational Intelligence Studies, vol. 7(1), pp. 33-66(2018).

[13] Xiaohong Jia, Yanning Zhang, Superpixel-based fast fuzzy c-means clustering for color image segmentation, IEEE transactions on fuzzy systems, DOI: 10.1109/TFUZZ.2018.28890182018

[14] Hruschka. E. R, Campello. R. J. G. B, Freitas. A. A and De carvalho. A. C. P. L. F, A survey of evolutionary algorithms for clustering, IEEE-Transactions on systems, man and cybernetics, part C, Applications and review, Vol. 39, No. 2, pp. 133-155(2009).

[15] Duda. R. O, Hart. P. E and Stork. D. G, Pattern classification, Wiley-Interscience, New York, USA, 2nd edition (2001).

[16] Girvan M., Newman M. E. J., Community structure in social and biological networks, In Proceeding of the National Academy of Sciences, USA. pp. 8271-8276 (2002).

[17] Prim R., Shortest connection networks and some generalization, Bell systems technical journal Vol: 36, pp. 1389-1401 (1957).

[18] Kruskal J, On the shortest spanning subtree and the travelling salesman problem In proceedings of the American Mathematical Society, pp. 48-50 (1956).

[19] Nesetril J, Milkova E, Nesetrilova H., Otakar Boruvka, On minimal spanning tree problem: Translation of both the 1926 papers, comments, history. DMATH. Discrete Mathematics, pp. 233(2001).

[20] Edward W. Packel, Functional Analysis, Intext Educational Publishers, New York (1974).

[21] Feng Luo, Latifur Kahn, Farokh Bastani, T Ling Yen and Jizhong Zhon, A dynamically growing self-organizing tree (DGOST) for hierarchical gene expression profile, Bio Informatics, Vol. 20, No. 16, pp-2605-2617(2004).

[22] Dunn J. C, A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters, Journal of cybernetics 3(3), pp. 32-57 (1973).

[23] Bezdek J. C, pattern recognition with fuzzy objective function algorithms, plenum press, New York (1981).