Modified fuzzy c -mean for custom-sized clusters

DEBJANI CHAKRABORTY and SUMAN DAS*

Department of Mathematics, Indian Institute of Technology Kharagpur, Kharagpur 721302, India
e-mail: debjani@maths.iitkgp.ac.in; pnksmdas@gmail.com

MS received 13 March 2019; revised 5 May 2019; accepted 31 May 2019

Abstract. Fuzzy c -mean (FCM) is one of the widely used data clustering methods. FCM method not only divides a data set into several clusters but also determines the potential belongingness of each data in different clusters. The size of clusters generated by FCM cannot be controlled by the inherent mechanism. However, sometimes real life situations demand that the clusters should have some pre-specified size. In this study, the FCM method is further extended to obtain clusters with specified size. In the first step of the proposed method, FCM algorithm is executed; later the potential belongingness matrix passes through an optimization model to yield clusters with specified sizes. In the proposed technique, the centres of the clusters obtained from FCM are considered but the boundary elements are redistributed to achieve equal or custom-sized clusters. The methodology has been explained further with examples.

Keywords. Fuzzy c -mean; size constraint; customized clustering; data mining; equal size clustering.

1. Introduction

Data clustering methods mainly divide a data set into different groups; each of these groups is labelled as a cluster. Generally, clustering is done with the given data set information only and no external information is provided during the clustering [1]. Clustering has several useful applications in big data clustering [2], document clustering [3], image segmentation [4], land cover classification [5], vehicle networking [6], mechanical engineering [7] and sensor clustering in wireless sensor networks [8, 9]. Several algorithms on data clustering have been proposed by researchers, a comprehensive review of which may be found in [10–12].

K -mean (KM) clustering [13] is one of the widely used clustering methodology. This method gives an optimal clustering with prior knowledge on number of clusters iteratively by updating the position of centroids until no significant change in centroids is observed. Though KM method is proved to be a very effective and useful clustering procedure, the nature of strict belongingness of each data in a particular cluster leaves no further scope of analysis or adjustment to the decision maker. Also KM method fails to accommodate the prior knowledge of cluster sizes, which may be known from the background information or due to some restriction. Later on, several researches tried to modify KM method to incorporate size constraint on clusters, which may be found in [14–17].

In 1973, Dunn [18] developed fuzzy c -mean (FCM) clustering method and later it was improved by Bezdek [19] in

1981. To capture hesitation in imprecise data, intuitionistic fuzzy c -mean (IFCM) was proposed by Xu and Wu [20] in 2010. Chaira [21] proposed a new IFCM, which is being used for testing medical images. Further developments on IFCM were also done by Kumar and Harish [22] in 2018 and Verma *et al* [23] in 2019. Kumar and Harish [22] used Bustince intuitionistic fuzzy generators for construction of intuitionistic fuzzy sets and Verma *et al* [23] incorporated the hesitation degree in their modification. Both of the modifications are tested with medical image segmentation.

The main difference between FCM method and the other clustering methods, like KM method, is that FCM method does not allocate data strictly into one cluster. FCM method calculates a matrix ' U ' (say) of potential degree of belongingness for each data in every possible cluster. In general, the assignment of data x_j depends on its highest belongingness to a particular cluster. Thus the knowledge on potential of belongingness of each data makes the FCM method more flexible. A decision maker can further analyse the characteristic of data using the matrix U according to the situation. A study is needed to find the possibility of adjustments in assigning data into clusters by analysing the matrix U if situation demands. In literature, there is no mathematical model from which the adjustment in the process of assignment with the knowledge of U can be done optimally for pre-assigned cluster-sized problems. Our work is focused to introduce a systematic way through which the adjustment in assignment process can be done in an optimum way for equal-sized or custom-sized clusters.

Size-constrained (SC) clustering is needed where all clusters are required to have some specified number of

*For correspondence
Published online: 17 July 2019

objects. Several real life situations require clusters of fixed sizes. For example, in a vehicle routine problem, each vehicle is bound to pick up passengers from their houses of a particular locality and every vehicle has its own defined capacity. Similarly, a job scheduling problem, where a number of jobs are assigned to different machines/processes, can be modelled as a data clustering problem with size constraints when different machines/processes have different capacities. Many other problems such as document clustering where each cluster has a fixed storage space and spatial clustering where each cluster can accommodate specific number of spatial objects can be naturally formulated as data clustering problems with size constraints [24].

Equi-sized clustering is a special case of SC clustering. In application there are many situations where equi-sized or uniform clustering is required. A few more examples may be found in photo query systems [25] and retail chain problems. Another promising application is in the energy load balance of wireless sensor networks [26], where unbalanced cluster structure may cause unbalanced energy consumption and shorten the network lifetime. Bradley *et al* [17] proposed a constrained k -mean clustering in 2000, which ensures a minimum number of objects in each cluster. Wagstaf *et al* also proposed constrained k -mean clustering method by imposing ‘must-link’ and ‘cannot-link’ types of constraints. In 2014, Malinen and Fränti [15] proposed balanced k -mean clustering by minimizing mean square error. Ganganath *et al* [14] also proposed a modification of k -mean algorithm to impose size constraints on each cluster. They selectively initialize centroid of each cluster with prior knowledge, which reduces the chances of getting local minima. A different type of SC clustering method was proposed by Zhu *et al* [24]. They modelled the SC clustering problem as linear programming problem and used a heuristic algorithm to solve the problem. An improvement of FCM method with size constraining was attempted by Höppner and Klawonn in [27, 28]. A study on spatial information of given data set that determines the number of cluster may be found in [29].

The algorithm proposed by Höppner and Klawonn is an iterative method where the size constraints are considered by equating the sum of membership values with the cluster sizes. Lagrangian method has been used to obtain the required assignments. One of the shortcomings of this method here is that the Lagrangian constants are obtained by solving a large system of linear equations iteratively. Moreover, size constraints do not assure the requirement of having j^{th} cluster with m_j number of objects or mathematically, $\sum_{i=1}^n u_{ij} = m_j$, where $U = \{u_{ij}\}$. In this algorithm the modification of the objective function of FCM method makes the algorithm more complicated, which affects the simplicity and efficiency of FCM method. Also, the consideration of size constraints from the first step of the algorithm changes the position and behaviour of clusters obtained from FCM method. Moreover, the procedure

defined in [27, 28] does not always lead to the desired cluster size at all.

In this paper, we take the centroids obtained from FCM method and then calculate the distances of each data from each centroid. Then the distance matrix passes through an integer linear optimization problem with size constraints to obtain the desired clustering. This paper proposes an extension of FCM method that can accommodate the restriction on size of clusters.

The rest of the paper is arranged as follows. Section 2 briefs the FCM method with algorithm. Section 3 presents the proposed methodology for equi-sized clustering with algorithm. Section 4 illustrates two examples on SC clustering and section 5 concludes our work.

2. FCM method

FCM clustering method divides a data set $X := \{x_1, x_2, x_3, \dots, x_n\} \subset \mathbb{R}^p$ into c number of different clusters. Every cluster has a representative prototype or centroid $p_j \in \mathbb{R}^p$. The major difference between FCM method and other hard partitional methods is that FCM does not restrict a data x_i ($i = 1, 2, 3, \dots, n$) to fall into a particular cluster p_j ($j = 1, 2, 3, \dots, c$). FCM method determines degree of belongingness $U_i = (u_{i1}, u_{i2}, \dots, u_{ic})$ of a data x_i into possible c number of clusters. In the first step of the algorithm, a set of centroids $\mathbf{P}^{(0)} = \{p_1^{(0)}, p_2^{(0)}, \dots, p_c^{(0)}\}$ is often initialized randomly. An overall study about other initialization methods was done by [30].

At k^{th} step the membership function matrix $\mathbf{U}^{(k)} = \{u_{ij}^{(k)}\}$ is obtained and the individual membership functions are calculated as follows:

$$u_{ij}^{(k)} = \frac{1}{\sum_{k=1}^c \left(\frac{\|x_i - p_j^{(k)}\|}{\|x_i - p_k^{(k)}\|} \right)^{\frac{2}{m-1}}}. \quad (1)$$

Here $\|\cdot\|$ is the norm expressing the dissimilarity or Euclidian distance between the observed data and the centre and ‘ m ’ is called fuzziness coefficient index or fuzziness parameter, which ranges from 1 to infinity. In general it is chosen as $m = 2$.

The centroid $p_j^{(k+1)}$ at $(k+1)^{\text{th}}$ step is then updated as follows:

$$p_j^{(k+1)} = \frac{\sum_{i=1}^n u_{ij}^{(k+1)m} x_j}{\sum_{i=1}^n u_{ij}^{(k+1)m}}. \quad (2)$$

Iteration of FCM algorithm continues with the goal of minimizing weighted sum of least-square errors with the following objective function:

$$J(\mathbf{U}, \mathbf{P}) = \sum_{i=1}^n \sum_{j=1}^c u_{ij}^m \|x_i - p_j\|^2. \quad (3)$$

The iteration goes on either to a fixed number of times or until the condition $\|J^{(k+1)}(U, P) - J^{(k)}(U, P)\| < \epsilon$ holds for a small preassigned $\epsilon > 0$. After the termination of iterations, the matrix of belongingness U is obtained. Inclusion of data x_i into j th cluster is decided from the measure of belongingness, i.e. if $u_{ij} \geq u_{ik}$ for $k = 1, 2, \dots, c$, then x_i is included in j th cluster whose centroid is p_j . Now in this process of assignment, the user does not have any control over the size of clusters. However, in the following sections, the sizes of clusters have been customized by introducing an optimization model with a size constraint.

In a nutshell, the algorithm of FCM method may be given as follows:

Algorithm 1 Algorithm of FCM

Require: Data set X ; clusters number c ; fuzziness index $m > 1$; termination threshold ϵ

- 1: Initialize the set of centroids $P^{(0)}$ randomly
- 2: **while** $\|J^{(k+1)} - J^{(k)}\| > \epsilon$
- 3: Update the membership matrix $U^{(k+1)}$ using Eq. (1)
- 4: Update the set of centroid $P^{(k+1)}$ using Eq. (2)
- 5: Calculate the value of the objective function $J^{(k+1)}$ using Eq. (3)
- 6: Calculate $\|J^{(k+1)} - J^{(k)}\|$
- 7: **end while**
- 8: Find $v_i = \max\{u_{i1}, u_{i2}, \dots, u_{ic}\}$
- 9: If $v_i = u_{ij}$, assign x_i into j th cluster

3. The proposed method

In each iteration of FCM method the potential belongingness of each data is calculated with the updated centroid and finally a potential belongingness matrix U is obtained. The assignment of a data x_i into a cluster p_j is then decided from the matrix U . This assignment procedure gives the flexibility to the decision maker for the last time adjustment to include a data into a particular cluster, which makes the FCM method more preferable than any other clustering methods. Adjustment for inclusion of a few number of data may be controlled but a systematic procedure to get a desired size clusters in a optimal way is needed.

If the situation demands clustering with strict restriction on size of clusters, then the assignment of data into clusters should be done in such a manner that total sum of distances of each data from the cluster centroids of the cluster to which the data belongs should be minimum.

Hence, the objective is to minimize $\sum_{i,j} d_{ij}$ with size constraint $|p_j| = m_j$, where $d_{ij} = D(x_i, p_j)$ and D stands for Euclidean distance. Here the summation runs over the indices i, j for which $x_i \in p_j$ and $|p_j|$ represents the numbers data in the j th cluster.

Thus, the optimization model through which the assignment of a data into a cluster should be done with cluster size constraint is given as follows.

Problem-1

$$\text{Maximize } f(y) = \sum_{i=1}^n \sum_{j=1}^c d_{ij} y_{ij} \quad (4)$$

$$\text{subject to } \sum_{j=1}^c y_{ij} = 1 \quad \text{for } i = 1, 2, \dots, n \quad (5)$$

$$\sum_{i=1}^n y_{ij} = m_j \quad \text{for } j = 1, 2, \dots, c \quad (6)$$

y_{ij} are all binary, i.e. $y_{ij} \in \{0, 1\}$.

Equation (6) represents the size constraint and Eq. (5) makes sure that a data is included into only one cluster. This model is a mixed integer linear optimization problem (MILOP) with ' $n + c$ ' number of constraints and ' $n \times c$ ' number of variables, where each variable is of binary type. The constraints $y_{ij} \in \{0, 1\}$ can be easily implemented in this MILOP by restricting the lower and upper bounds of the variables by 0 and 1, respectively, i.e. $0 \leq y_{ij} \leq 1$ for $i = 1, 2, \dots, n; j = 1, 2, \dots, c$.

After solving this mixed integer linear optimization model, a matrix $y = [y_{ij}]_{n \times c}$ of elements with 0 and 1 is obtained. The assignment of an object x_i into j th cluster is then obtained using the following rule:

$$x_i \in p_j \text{ if and only if } y_{ij} = 1. \quad (7)$$

Balanced clustering or clustering with all clusters having same sizes is a special case of SC clustering. In this case, the value of m_j in Eq. (6) is to be replaced with $m_j = \frac{n}{c}$ for $j = 1, 2, \dots, c$. It is to be noted that in case of balanced clustering, the size of the data ' n ' is assumed to be divisible by the number of clusters ' c '. The optimization model defined in **Problem 1** with these constraints reallocates an object x_i in a cluster. Mostly the boundary elements of the FCM clusters are redistributed, keeping the other elements near cluster centre at their own positions.

For a fixed number of variables, Lenstra [31] showed that MILOP can be solved in polynomial time. If for a large number of data set, the branch and bound method fails to solve the MILOP then advancement of heuristic methods such as tabu search [32] may be used in this case.

The algorithm of the proposed method is given as follows.

Algorithm 2 Algorithm of FCM with size constraint

Require: Data set X ; clusters number c ; fuzziness index $m > 1$; termination threshold ϵ ; size of clusters $\{m_1, m_2, \dots, m_c\}$

- 1: Execute FCM **Algorithm 1**.
 - 2: Obtain the potential belongingness matrix U
 - 3: Solve the optimization problem with size constraint $\{m_1, m_2, \dots, m_c\}$
 - 4: Obtain the solution set $[y_{ij}]_{n \times c}$
 - 5: Assign data x_i in cluster p_j with rule (7)
-

4. Experimental result

In this section we present experimental results of our proposed method in comparison to the existing method. Example 4.1 presents comparison of our proposed method to FCM and HK method on a two-dimensional randomly distributed data set. Example 4.2 compares proposed method to FCM and HK method on a data set of car sharing problem [33]. The second part of this section presents accuracy comparison of proposed method on data sets taken from UCI machine learning repository.

Example 4.1 In this example a two-dimensional data set X consists of $n = 2200$ points, which are randomly distributed over $(x, y) \in [0, 16] \times [0, 14]$. After applying FCM algorithm for $c = 5$ clusters, in an instance the following clusters are obtained: $p_1(12.50, 1.73)$; $p_2(6.21, 10.83)$; $p_3(2.72, 4.91)$; $p_4(12.59, 10.30)$; $p_5(13.36, 5.46)$ with sizes $m_1 = 465$; $m_2 = 523$; $m_3 = 424$; $m_4 = 333$; $m_5 = 455$, respectively (see figure 1).

In order to distribute the same data set X to the pre-assigned sized clusters, algorithm 2 may be applied. The proposed algorithm 2 readjusts the data among different clusters in an optimum way for the given cluster sizes $m = [m_1, m_2, m_3, m_4, m_5] = [500, 380, 420, 470, 430]$. The results obtained from algorithm 2 with the size constraints $m = [m_1, m_2, m_3, m_4, m_5] = [500, 380, 420, 470, 430]$ are shown in figure 2. It is worth noting that the data description is preserved here as the centres of the clusters, obtained from FCM, are unchanged. Only the boundary data of the clusters are redistributed within the clusters.

In a more comprehensive manner it can be said that the proposed algorithm 2 can redistribute the data to the same-sized clusters in an optimum way. Figure 3 depicts the clustering obtained from algorithm 2 with equal size constraints where 440 data points are allocated in each cluster.

Figure 1. FCM clustering without size constraint.

Figure 2. Proposed clustering with customized size.

Figure 3. Proposed clustering with equal size.

The result obtained from the proposed method is compared to the result obtained from the method proposed by [27] (Method HK) for equal-sized clustering. HK method gives the cluster centroids at $p_1(11.51, 5.66)$; $p_2(6.04, 7.14)$; $p_3(9.41, 7.38)$; $p_4(7.73, 8.31)$; $p_5(11.99, 4.39)$ with cluster sizes $m = [m_1, m_2, m_3, m_4, m_5] = [181, 505, 632, 435, 447]$. Not only the positions of the centroids vary from those of the centroids obtained in method FCM but also the required size constraints are not satisfied in HK method. Figure 4 depicts the clustering obtained from HK method with equal size constraints. This result clearly shows that the proposed method fulfills the requirement of SC clustering but the method proposed in [27] fails to meet the same.

Example 4.2 Let us consider a situation in which a car sharing problem in a city is modelled. This shared

Figure 4. HK method clustering.

transportation problem is for workplace commute that considers the journey of the commuters from home to work and back again. Similar type of vehicles are assigned to do the job. Vehicles are shared by the customers. The objective is to plan an allotment so that all the vehicles are in work and every customer is served in the allotment process. If all the vehicles have same sitting capacities and the geographic locations of the intended passengers are known with their longitude and latitude coordinates, then an optimal planning needs to be designed for it. The primary task would be to assign n customers to m number of vehicles uniformly. Uniformity has to be maintained as the capacities of the vehicles are same. Hence, an appropriate clustering algorithm has to be designed so that assignments of the vehicles to customers are optimal. The situation should be viewed in a more elaborate fashion by considering more constraints, like the distance covered by each vehicle must be minimum, the waiting time of the customers should be minimum, etc. However, here the problem is considered in a restricted manner that is a part of the whole actual scenario. The optimal allotment of intended passengers to a limited number of vehicles is only the main issue of this clustering example. The initial hundred data points are considered from [33] and it is assumed that ten vehicles each with capacity of ten people are to be assigned to hundred customers with known geographical coordinates optimally.

After applying FCM algorithm for $c = 10$ vehicles, in an instance, 10 clusters are obtained with centroids at $p_1(38.91, -77.02)$; $p_2(38.89, -76.98)$; $p_3(38.86, -76.98)$; $p_4(38.91, -77.01)$; $p_5(38.92, -77.09)$; $p_6(38.92, -77.04)$; $p_7(38.83, -77.00)$; $p_8(38.96, -77.01)$; $p_9(38.93, -76.98)$; $p_{10}(38.88, -76.94)$; the cluster sizes are obtained as $m = [6, 11, 7, 17, 7, 11, 13, 9, 13, 6]$, respectively (see figure 5). The obtained clusters are not of equal sizes but it is necessary to distribute 100 passengers in 10 vehicles with equal capacity of 10 each.

Figure 5. FCM method.

The same data set of locations of passengers is then passed through the HK [27] method with requirement of 10 equal-sized clustering. The HK method converges by providing 10 cluster centroids as $p_1(38.92, -77.03)$; $p_2(38.92, -77.02)$; $p_3(38.93, -77.02)$; $p_4(38.87, -76.98)$; $p_5(38.90, -76.98)$; $p_6(38.88, -76.98)$; $p_7(38.91, -77.01)$; $p_8(38.91, -77.02)$; $p_9(38.92, -76.02)$; $p_{10}(38.91, -77.01)$ with cluster sizes $m = [22, 7, 13, 24, 12, 3, 3, 2, 6, 8]$, respectively (see figure 6). Hence, the HK method also fails to meet the requirement of the situation.

The proposed method, given in algorithm 2 is applied (see figure 7) here. It is observed that the clusters obtained are all of equal sizes, the cluster centres are almost same as FCM cluster centres and only the boundary elements of FCM clusters are redistributed to meet the requirement. The centres of the new clusters with 10 cluster sizes are as follows: $p_1(38.91, -77.03)$; $p_2(38.90, -76.99)$; $p_3(38.86, -76.97)$; $p_4(38.92, -77.01)$; $p_5(38.92, -77.08)$; $p_6(38.92, -77.04)$; $p_7(38.84, -77.00)$; $p_8(38.96, -77.02)$; $p_9(38.94, -76.99)$; $p_{10}(38.89, -76.94)$.

Figure 6. HK's method.

Figure 7. Proposed method.

Table 1. Data description.

Data set	Dimension	No. of classes	Class size
Iris	4	3	50,50,50
Wine	13	3	59,71,48
Balance	4	3	228,228,49
Ionosphere	34	2	126,225

4.1 Accuracy comparison

To check the effectiveness of proposed modified fuzzy c -mean (MFCM) method, accuracy of MFCM is compared to that of KM [13], SC [24] and FCM method [18] on four UCI data sets. The characteristics of the data sets are summarized in table 1 and more description of the data sets can be found in [34].

Table 2. Accuracy comparison.

Data set	Methods	Accuracy
Iris	KM	0.8133
	SC	0.8467
	FCM	0.8666
	MFCM	0.9200
Wine	KM	0.7022
	SC	0.7079
	FCM	0.6853
	MFCM	0.6236
Balance	KM	0.4976
	SC	0.5248
	FCM	0.5248
	MFCM	0.5520
Ionosphere	KM	0.7126
	SC	0.8034
	FCM	0.7179
	MFCM	0.8091

The best accuracies for each data set are highlighted in bold

Accuracy of clustering technique reflects the strength of technique to capture the actual relationship between given classes and obtained cluster of a data set. It is measured as summation of matching degree of all clusters and classes. Formulation of accuracy of a clustering method is given as follows:

$$accuracy = \frac{\sum_k \max \left\{ T(C_k, L_m) \right\}}{N} \quad (8)$$

where C_k and L_m denote the k^{th} and m^{th} cluster and class, respectively, of data set. $T(C_k, L_m)$ is the number of data of L_m class belonging to C_k . Accuracy comparison on different data sets is presented in table 2.

5. Conclusion

FCM method does not accommodate the size restrictions on clusters. In this paper, the FCM method has been extended with an optimization model, which helps adjusting the size of each cluster after the execution of the same. Thus, this method does not disturb the simplicity and efficiency of FCM but widens the applicability of FCM method in various real life problems where the cluster sizes are restricted with external information. Simulation of this algorithm is needed for a large data set.

Job scheduling problem where each agent or machine has to be allocated with a predefined numbers of jobs may be considered as a future real life application of the proposed method.

References

- [1] Tan P N, Steinbach M and Kumar V 2005 *Introduction to data mining*, 1st ed. Boston: Addison-Wesley Longman Publishing Co. Inc.
- [2] Tong H and Kang U 2013 Big data clustering. In: *Data Clustering: Algorithms and Applications*, p. 370
- [3] Karypis M S G, Kumar V and Steinbach M 2000 A comparison of document clustering techniques. In: *Proceedings of the Text Mining Workshop at KDD2000*
- [4] Pham DL 2001 Spatial models for fuzzy clustering. *Comput. Vis. Image Underst.* 84(2): 285–297
- [5] Sowmya B and Sheelarani B 2011 Land cover classification using reformed fuzzy C-means. *Sadhana* 36(2): 153–165
- [6] Çalhan A 2015 A fuzzy logic based clustering strategy for improving vehicular ad-hoc network performance. *Sadhana* 40(2):351–367
- [7] Ghosh A, Pratihari D K, Amarnath M V V, Dittrich G and Mueller J 2012 Fuzzy clustering of mechanisms. *Sadhana* 37(5): 539–556
- [8] Abbasi A A and Younis M 2007 A survey on clustering algorithms for wireless sensor networks. *Comput. Commun.* 30(14–15): 2826–2841

- [9] Cheng C T, Chi K T and Lau F C 2011 A clustering algorithm for wireless sensor networks based on social insect colonies. *IEEE Sensors J.* 11(3): 711–721
- [10] Jain A K, Murty M N and Flynn P J 1999 Data clustering: a review. *ACM Comput. Surv.* 31(3): 264–323
- [11] Madhulatha T S 2012 *An overview on clustering methods*. arXiv preprint [arXiv:1205.1117](https://arxiv.org/abs/1205.1117)
- [12] Omran M G, Engelbrecht A P and Salman A 2007 An overview of clustering methods. *Intell. Data Anal.* 11(6): 583–605
- [13] MacQueen J 1967 Some methods for classification and analysis of multivariate observations. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, June, vol. 1, No. 14, pp. 281–297
- [14] Ganganath N, Cheng C T and Chi K T 2014 Data clustering with cluster size constraints using a modified k-means algorithm. In: *Proceedings of the 2014 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*, IEEE, pp. 158–161
- [15] Malinen M I and Fränti P 2014 Balanced k-means for clustering. In: *Proceedings of the Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, August. Berlin–Heidelberg: Springer, pp. 32–41
- [16] Wagstaff K, Cardie C, Rogers S and Schrödl S 2001 Constrained k-means clustering with background knowledge. In: *Proceedings of ICML*, June, vol. 1, pp. 577–584
- [17] Bradley P S, Bennett K P and Demiriz A 2000 Constrained k-means clustering. MSR-TR-2000-65, Microsoft Research, Redmond, 20
- [18] Dunn J C 1973 A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *J. Cybern.* 3(3): 32–57. <https://doi.org/10.1080/01969727308546046>
- [19] Bezdek J C 1981 Objective function clustering. In: *Pattern Recognition with Fuzzy Objective Function Algorithms*. Boston: Springer, pp. 43–93
- [20] Xu Z and Wu J 2010 Intuitionistic fuzzy C-means clustering algorithms. *J. Syst. Eng. Electron.* 21(4): 580–590
- [21] Chaira T 2011 A novel intuitionistic fuzzy C means clustering algorithm and its application to medical images. *Appl. Soft Comput.* 11(2): 1711–1717
- [22] Kumar S A and Harish B S 2018 A modified intuitionistic fuzzy clustering algorithm for medical image segmentation. *J. Intell. Syst.* 27(4): 593–607
- [23] Verma H, Gupta A and Kumar D 2019 A modified intuitionistic fuzzy c-means algorithm incorporating hesitation degree. *Pattern Recognit. Lett.* 122: 45–52
- [24] Zhu S, Wang D and Li T 2010 Data clustering with size constraints. *Knowl. Based Syst.* 23(8): 883–889
- [25] Althoff T, Ulges A and Dengel A 2011 Balanced clustering for content-based image browsing. *Series of the Gesellschaft für Informatik* 1: 27–30
- [26] Du Z, Liu Y and Qian D 2009 An energy-efficient balanced clustering algorithm for wireless sensor networks. In: *Proceedings of the 2009 5th International Conference on Wireless Communications, Networking and Mobile Computing*, September 2009, IEEE, pp. 1–4
- [27] Höppner F and Klawonn F 2008 Clustering with size constraints. In: *Computational Intelligence Paradigms*. Berlin: Springer, pp. 167–180
- [28] Klawonn F and Höppner F 2006 October. Equi-sized, homogeneous partitioning. In: *Proceedings of the International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*. Berlin: Springer, pp. 70–77
- [29] Daisy V R and Nirmala S 2018 Stability-integrated fuzzy C means segmentation for spatial incorporated automation of number of clusters. *Sadhana* 43(3): 40
- [30] Celebi M E, Kingravi H A and Vela P A 2013 A comparative study of efficient initialization methods for the k-means clustering algorithm. *Expert Syst. Appl.* 40(1): 200–210
- [31] Lenstra Jr H W 1983 Integer programming with a fixed number of variables. *Math. Oper. Res.* 8(4): 538–548
- [32] Glover F 1986 Future paths for integer programming and links to artificial intelligence. *Comput. Oper. Res.* 13(5): 533–549
- [33] Nour M A 2018 C5 bike station locations. http://ropafiles.ucoz.site/C5_bike_station_locations.csv (1st September)
- [34] Asuncion A and Newman D 2007 UCI machine learning repository. University of California, School of Information and Computer Science, Irvine, CA. <http://www.ics.uci.edu/mllearn/MLRepository.html>.