WECM: an evidential subspace clustering algorithm

Van Tri Do[1[0000−0002−8966−8132]], Violaine Antoine[1[0000−0002−0981−3505]], and Jonas Koko[1[0000−0002−0970−5002]]

Université Clermont Auvergne, Clermont Auvergne INP, UMR 6158 CNRS, LIMOS, F-63000 Clermont-Ferrand, France firstname.secondname@uca.fr

Abstract. This paper introduces WECM, a novel evidential and subspace clustering algorithm. It is based on the Evidential c-means, a variant of the k-means designed to produce a credal partition, allowing a better representation of the partial knowledge regarding the class membership of objects. The WECM algorithm integrates weights on features and clusters to enhance the clustering separability and interpretability. Experiments conducted on synthetic and real data show the positive effects of the weights on the clustering performances.

Keywords: evidential c-means · subspace clustering · weights · belief function theory.

1 Introduction

Clustering is an unsupervised learning approach of machine learning, used in various fields such as medicine [4], computer vision [25], IoT [18], etc. to unveil hidden patterns within datasets. The primary objective of clustering algorithms is to create groups of objects, ensuring that the similarity among elements in a group surpasses the similarity between different groups. There exists three categories of clustering: the hierarchical clustering, the density-based clustering, and the prototype-based clustering [23]. The prototype-based algorithms are widely used because of their simple computation, easy interpretation, and their clearly defined objective function that they strive to optimize. Among prototype-based algorithms, the most famous algorithm is the k-means algorithm. It generates a hard partition which assign objects to exactly a single cluster. In many real-world cases, however, clusters are overlapping, and the objects in these between-clusters areas are uncertain to belong to a specific cluster. Under such scenarios, the hard-partitioning of k-means forces a crisp assignment that can lead to poor performance. For this reason, variants of k-means creating soft partitions have been proposed. Based on the concept of partial membership described in the fuzzy sets theory [29], the fuzzy c-means (FCM) algorithm [2] generates a fuzzy partition where each object has a degree of membership to each cluster. Possibilistic extensions of k-means [15, 20] have been introduced to better manage noise and outliers. The Evidential C-means (ECM) [17] treats overlapping regions between clusters as new clusters and take advantages of the belief function

theory to express the membership value of each object with respect to all clusters. The outcome of ECM is a credal partition, considered to express in a richer way the partial knowledge concerning the assignment of an object to a cluster. Indeed, this credal partition can be converted into hard, fuzzy, or possibilistic partitions through the use of transformation functions.

For above-mentioned algorithms and their variants, the assumption is made that all features of a given dataset contribute equally to the construction of optimal clusters. However, in some cases, some features may hold greater significance in providing clustering information than others. Consequently, features with higher relevance play a more crucial role in achieving the optimal clustering outcome compared to those with lower relevance. By identifying and assessing these relevant features through clustering algorithms, improvements in accuracy and computational efficiency can be achieved. Several approaches have been developed for this purpose. First methods include soft clustering algorithms using Mahalanobis distances [9, 10, 19, 16]. This distance, adapted for each cluster, enables the representation of importance and correlations among features. However, defining both importance and correlations adds complexity in the minimization process, occasionally leading to inconclusive results, especially with challenging or large datasets. The feature-weighting techniques, also referred to as subspace clustering techniques, propose optimizing the weights within objective function simultaneously with the partition [7]. There exists two types of method: the global method assigns the same weight vector to all clusters [13, 26], while the local method involves assigning a different weight vector to each cluster [8, 14, 21, 11, 28]. Subspace clustering makes the assumption of independence among the attributes of a dataset. It is commonly used with Euclidean distance and allows good interpretability of the results.

As we are aware, there is limited existing research on feature-weighted ECM in the literature. This study specifically concentrates on incorporating feature-weights into ECM. The new algorithm, termed feature-weighted ECM (WECM), automatically calculates feature weights for different features and generates a credal partition. The remainder of this paper is organized as follows. Section 2 recalls some backgrounds related to ECM clustering algorithm. Section 3 is our proposed approach feature-weighting ECM algorithm. The experiments and results are given in Section 4, and finally Section 5 is our conclusion for this research.

2 Background

2.1 Belief function theory

The belief function theory [24] is a mathematical framework that enables the representation of uncertain and imprecise information. Let us consider $\Omega = \{\omega_1, \ldots \omega_c\}$ a finite set of events and ω the true event occurring in the context of a system. The basic belief assignment (bba) $m : 2^{\Omega} \rightarrow [0, 1]$ quantifies the

partial knowledge regarding the event ω. It satisfies:

$$\sum_{A \subseteq \Omega} m(A) = 1.$$

Subsets $A \subseteq \Omega$ such that $m(A) > 0$ are called focal sets. The bbas have various interpretations following their distributions. If the focal sets are all singletons, m is a Bayesian bba. If only one singleton holds all the belief, i.e. $m(\omega) = 1$, then the bba expresses a full certainty. Inversely, $m(\Omega) = 1$ corresponds to a total ignorance of the real value of ω. Eventually, when $m(\emptyset) = 0$, the bba is said to be normal. Inversely, $m(\emptyset) > 0$ raises the possibility that the event ω does not belong to the frame of discernment Ω.

There exists various transformations of a mass function in order to obtain a probability distribution, a possibility distribution, or to make a crisp decision regarding the actual value of ω. The pignistic transformation converts a normal bba m into a probability distribution:

$$BetP(\omega) = \sum_{\omega \in A} \frac{m(A)}{|A|},$$

where $|A|$ denotes the cardinality of $A \subseteq \Omega$. In case of a subnormal bba, a normalization can be achieved by dividing $1 - m(\emptyset)$ among the elements of A [27].

2.2 Evidential c-means

The Evidential c-means clustering algorithm (ECM) is an adaptation of the traditional k-means in the framework of belief function theory. It provides a credal partition, an informative partition where each subset of clusters $A_j \subseteq \Omega$ is associated with a belief mass function. This function indicates the degree of belief that each object belongs to a cluster included A_j. This representation enables to express various situations, ranging from complete ignorance to total certainty.

Let $\boldsymbol{X} = (\boldsymbol{x}_i) = (x_{ip}) \in \mathbb{R}^{n \times q}$ be the set of n objects characterized by q attributes, $\Omega = \{\omega_1, \ldots, \omega_c\}$ be the clusters, $\boldsymbol{V} = (\boldsymbol{v}_k) = (v_{kp}) \in \mathbb{R}^{c \times q}$ be the centroids of the c clusters, and $\boldsymbol{M} = (m_{ij}) \in \mathbb{R}^{n \times 2^c}$ be the credal partition. Similarly to k-means, ECM involves the use of centroids to represent clusters. Then, each subset A_j is associated with a centroid $\boldsymbol{v}_j$, which is computed as the center of mass of the prototypes associated with the classes comprising A_j. The Euclidean distance is finally employed to measure the dissimilarity between an object $\boldsymbol{x}_i$ and the centroid $\boldsymbol{v}_j$. Since no centroid can be defined to the empty set, a fixed distance δ is incorporated, similarly to the noise-clustering algorithm [6].

The ECM algorithm searches to minimize the intra-cluster distances with respect to the centroids $\boldsymbol{V}$ and the credal partition $\boldsymbol{M}$:

$$J_{ECM}(\boldsymbol{M}, \boldsymbol{V}) = \sum_{i=1}^{n} \sum_{\{j/A_j \neq \emptyset, A_j \subseteq \Omega\}} |A_j|^{\alpha} m_{ij}^{\beta} d_{ij}^2 + \sum_{i=1}^{n} m_{i\emptyset}^{\beta} \delta^2,$$

such that

$$\sum_{\{j/A_j \subseteq \Omega\}} m_{ij} = 1 \qquad \forall i \in \{1, \ldots, n\}, \tag{1}$$

$$m_{ij} \geq 0 \qquad \forall i \in \{1, \ldots, n\}, \forall A_j \subseteq \Omega, \tag{2}$$

where α and β are exponents controlling the imprecision and the fuzziness of the credal partition, $m_{i\emptyset}$ refers to the mass of $\boldsymbol{x}_i$ given to the empty set, and d_{ij} corresponds to the Euclidean distance between the object $\boldsymbol{x}_i$ and the centroid $\boldsymbol{v}_j$.

2.3 Soft subspace clustering with local weights

Several soft subspace clustering algorithms using locally feature weights have been proposed. In this framework, the weights are introduced for each clusters considering a weighted squared Euclidean distance. This distance is defined as follows:

$$d_{ik}^2 = \sum_{p=1}^{q} w_{kp}^s (x_{ip} - v_{kp})^2, \quad \forall i = \{1, \ldots, n\}, \forall k \in \{1, \ldots, c\}, \tag{3}$$

where $w_{kp} \in [0, 1]$ denotes the weight of the cluster k for the feature p, $s > 1$ is an hyper-parameter that controls the fuzziness of the weight. The weights $\boldsymbol{W} = (w_{kp}) \in \mathbb{R}^{c \times q}$ can then be adjusted through the optimization of an objective function.

Most of the methods are extensions of FCM [8, 14, 21, 11]. Variations between fuzzy subspace clustering include the introduction of a weight entropy [14] or the consideration of both feature weighting and cluster weighting [11]. Recently, few extensions have also been proposed for the possibilistic c-means algorithms PFCM [22] and PCM [28]. It shows the interest of the community for algorithms generating various types of partial knowledge. In this vein, we propose to extend the ECM algorithm to handle local weights.

3 Locally weighted Evidential c-means

3.1 Objective function

Similarly to soft subspace clustering, we define weights $\boldsymbol{W} = (w_{kp})$ for all clusters and features. Then, we characterize the weights of subsets A_j as the average of the weights of clusters included in A_j:

$$w_{jp} \triangleq \frac{1}{|A_j|} \sum_{\omega_k \in A_j} w_{kp}, \quad \forall p = \{1, \ldots, q\}, \forall A_j \subseteq \Omega. \tag{4}$$

The squared Euclidean distance between an object $\boldsymbol{x}_i$ and a subset A_j is then calculated using the weights and the centroid associated to the subset. The

objective function of the subspace evidential c-means, referred to as weighted evidential c-means (WECM) is as follows:

$$J_{WECM}(\boldsymbol{M},\boldsymbol{V},\boldsymbol{W}) = \sum_{i=1}^{n} \sum_{A_j \neq \emptyset, A_j \subseteq \Omega} |A_j|^{\alpha} m_{ij}^{\beta} \sum_{p=1}^{q} w_{jp}^2 (x_{ip} - v_{jp})^2 + \sum_{i=1}^{n} m_{i\emptyset}^{\beta} \delta^2, \quad (5)$$

such that constraints (1), (2) on masses are respected, and

$$\sum_{p=1}^{q} w_{kp} = 1, \qquad \forall \omega_k \in \Omega, \forall i \in \{1,\ldots,n\}, \quad (6)$$

$$w_{kp} \geq 0 \qquad \forall \omega_k \in \Omega, \forall p \in \{1,\ldots q\}. \quad (7)$$

Similarly to the centroids, only weights associated to cluster has to be optimized using the objective function. It is trivial to show that constraints on weights regarding subsets $|A_j| > 1$ are equivalent to constraints regarding weights on clusters:

Proposition 1 (Constraints on weights). *The weights $w_{jp} \forall A_j \subseteq \Omega$ respect the positivity constraint and*

$$\sum_{p=1}^{q} w_{jp} = 1, \quad \forall A_j \subseteq \Omega, |A_j| > 1, \forall i \in \{1,\ldots,n\}. \quad (8)$$

Proof. Since the weights $w_{jp}\ \forall A_j \subseteq \Omega$ are defined as the mean of cluster weights (4) and all cluster weights are positive (7), then $w_{jp} \geq 0\ \forall A_j \subseteq \Omega$. For the sum constraint, let us develop (4) in the sum of the weights:

$$\begin{aligned} \sum_{p=1}^{q} w_{jp} &= \sum_{p=1}^{q} \left(\frac{1}{|A_j|} \sum_{\omega_k \in A_j} w_{kp} \right), \\ &= \frac{1}{|A_j|} \sum_{\omega_k \in A_j} \sum_{p=1}^{q} w_{kp}. \end{aligned} \quad (9)$$

Integrating constraint (1) in (9) gives then (8).

3.2 Optimization

The optimization of WECM involves iteratively optimizing the cluster centroids $\boldsymbol{V}$, the credal partition $\boldsymbol{M}$, and the cluster weights $\boldsymbol{W}$ until convergence. The two first steps result to similar update than ECM [17], except a squared Euclidean distance is used with fixed weights. The update of $\boldsymbol{W}$, implying to fix $\boldsymbol{M}$ and $\boldsymbol{V}$ in the objective function, is obtained using the projected gradient method [5].

First, the derivative of J_{WECM} with respect to $w_{k'p'}$ is computed :

$$\frac{\partial J_{WECM}}{\partial w_{k'p'}} = \sum_{i=1}^{n} \sum_{\substack{\{j/A_j \subseteq \Omega, \\ A_j \cap \omega_{k'} \neq \emptyset\}}} |A_j|^{\alpha-2} m_{ij}^{\beta} \left(2w_{k'p'} + 2 \sum_{\substack{\omega_k \in A_j \\ \omega_k \neq \omega_{k'}}} w_{kp'} \right) (x_{ip'} - v_{jp'})^2.$$

Let $\nabla J_1(\boldsymbol{W}) = (\frac{\partial J_{WECM}}{\partial w_{k'p'}})$ be the matrix $\mathbb{R}^{c \times q}$ containing the derivatives $\forall k' \in \{1 \dots c\}$ and $\forall p' \in \{1 \dots q\}$. Constraint $\sum_{p=1}^{q} w_{jp} = 1$ can be rewritten as $e^T \boldsymbol{w}_k = 1$ where $e^\top = [1\ 1 \cdots 1]$ is a $q \times 1$ vector. To keep the $\boldsymbol{w}_k$ inside the set defined by the above constraint, during the descent process, we project $\nabla J_1(\boldsymbol{W})$ onto the kernel of the linear subspace defined by the constraint. The projection matrix is

$$P = \mathbb{I} - \frac{1}{||e||^2} ee^T = \mathbb{I} - \frac{1}{q} ee^T.$$

Since $e^\top e = q$, $\mathbb{I}$ is the identity matrix. The projection matrix $P = (p_{ij})$ is

$$p_{ij} = \begin{cases} 1 - \frac{1}{q} & \text{if } i = j \\ -\frac{1}{q} & \text{otherwise.} \end{cases}$$

The projected gradient iteration is as follows

$$\boldsymbol{w}_k^{(l+1)} \leftarrow \boldsymbol{w}_k^{(l)} - \gamma P(\nabla J_1(\boldsymbol{W}^{(l)}))_k, \tag{10}$$

where γ is a small positive number (e.g. $0.001 - 0.01$). We repeatedly update the value of weight and value of $J_1(\boldsymbol{W})$ until reaching stopping condition, if $\| P \nabla J_1(\boldsymbol{W}^l) \|$ becomes sufficiently small.

4 Results analysis

4.1 Experimental protocol

The WECM algorithm was evaluated using two synthetic data sets and five data sets from the UCI machine learning[1]. The synthetic data sets, named Toys2D and Toys6D, are composed of points generated from two multivariate Gaussian distributions. Figure 1 presents the Toys2D data set.

The Toys6D data set has a multivariate Gaussian distribution for its two classes. The first distribution is defined with $\boldsymbol{\mu}_1 = [1, 1, 1, 1, 1, 1]$ and $\boldsymbol{\Sigma}_1 = [1, 8, 3, 7, 9, 4] \times \boldsymbol{I}$, where $\boldsymbol{I}$ represents the identity matrix, whereas the second distribution is characterized by $\boldsymbol{\mu}_2 = [4.5, 4.5, 4.5, 4.5, 4.5, 4.5]$ and $\boldsymbol{\Sigma}_2 = [9, 2, 10, 3, 1, 8] \times \boldsymbol{I}$.

The characteristics of the data sets are illustrated Table 1. Remark that the LettersIJL corresponds to letters data set, where only the three letters {I,J,L} has been selected [3].

[1] https://archive.ics.uci.edu

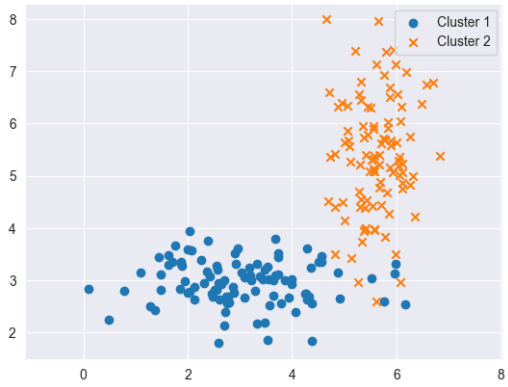

Fig. 1. Toys2D data set.

Table 1. Characteristics of the data sets.

	n	p	c
Toys2D	200	2	2
Toys6D	400	6	2
Iris	150	4	3
LettersIJL	227	16	3
Lung	32	56	3
Seeds	210	6	3
Wine	178	13	3

We set the hyperparameters $\alpha = 1$, $\beta = 2$, and $\delta = 100$, as is commonly performed by default [16, 1].

To measure the performance of our algorithm, we choose internal and external evaluation criteria. The internal evaluation criteria is the non-specificity. It measures how uncertain is the credal partition obtained: $N(\boldsymbol{M}) = \frac{1}{n}\sum_{i=1}^{n} N(m_i)$, with

$$N(m_i) = \sum_{A_j \subseteq \Omega} m_i(A_j)\log_2(|A_j|) + m_{i\emptyset}\log_2(|\Omega|).$$

The Adjusted Rand Index (ARI) [12] is an external measure used to evaluate the similarity between two crisp clustering partitions, considering both the agreement and disagreement of cluster assignments. The ARI is equal to 1 (respectively 0) when there exists a total concordance (respectively a total dissimilarity). In order to compute the ARI, the credal partition of the evidential clustering is first transformed into a hard partition using the maximal pignistic transformation. The second partition corresponds to the true labels available of the data sets.

Centroids are randomly initialized for both ECM and WECM, and the weights in WECM are initialized to $1/q$, in order to start the algorithm with a balance

between features. We run 10 trials of each clustering algorithm and then we select the solution having the lowest objective function.

4.2 Performance of WECM

The interest of WECM compared to other soft subspace clustering is its ability to produce a credal partition. Figure 2 presents the hard credal partition, which assigns each object to the subset of classes with the highest mass, for (a) ECM and (b) WECM. It shows that there exists an imprecise area where making a decision without further prior knowledge might be harzardous. With WECM, this imprecise area is smaller, and its center is slightly shifted upwards compared to ECM, owing to the different weights assigned to the two features.

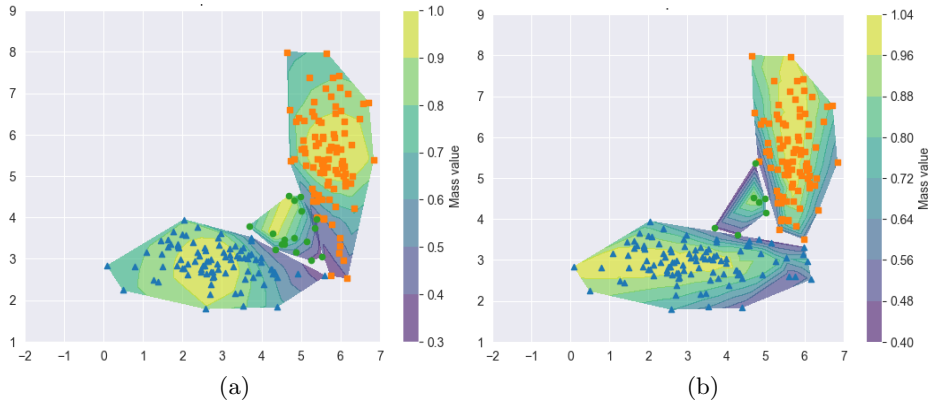

Fig. 2. Credal partition obtained by (a) ECM and (b) WECM on Toys2D. Hard credal partition is represented by the symbols and colors, and the mass values are represented by contours.

Table 2 summarizes the results obtained using ECM and WECM. As it can be observed, the WECM algorithm achieves better partitions for four data sets and comparable results for one other data set. However, it yields lower performance for the Iris and Seeds data set, suggesting that the data may not naturally exhibit clear subspaces. Finally, it can be noticed that most of the time, Non-Specificity and ARI are correlated although it is not a general rule.

For the Toys data sets, it is possible to visualize the weights obtained with respect to the variances of the multivariate distributions (cf. Figure 3). As expected, in the case of good performance of the algorithm, there exists a negative correlation between weights and variances. Indeed, a low variance implies a denser area, increasing the chance of separating clusters.

Table 2. ARI and Non specificity for ECM and WECM

	ARI		N	
	ECM	WECM	ECM	WECM
Toys2D	0.85	**0.90**	0.87	**0.70**
Toys6D	0.84	**0.90**	0.99	**0.86**
Iris	**0.60**	0.43	**1.5**	1.67
LettersIJL	0.18	**0.21**	1.43	**1.37**
Lung	0.14	**0.21**	1.45	1.45
Seeds	**0.68**	0.47	1.26	**0.73**
Wine	**0.83**	0.82	1.38	1.38

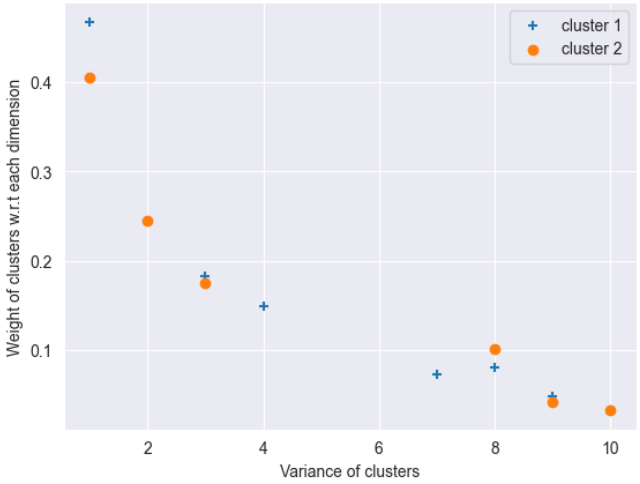

Fig. 3. Toys6D variances again the weights obtained by WECM.

5 Conclusion

In this paper, we proposed WECM, a new subspace clustering algorithm capable of generating a credal partition. The weights optimized by the algorithm allow to effectively handle the varying importance across the dimensions of a data set, and enhances the interpretability of clustering results. By integrating the principles of belief function theory into the clustering process, this algorithm enables the rich representation partial information regarding the class membership of an object. Preliminary results show that WECM can outperform the ECM algorithm for some data sets.

In the future, several validation studies and upgrades can be undertaken. This includes investigating the influence of the parameters α and β on WECM compared to ECM, as well as specific optimizations for updating ECM with respect to the weights. The definition of centroids for subsets with cardinalities higher than one can also be modified to better locate the imprecise regions. Finally, the

goal is to apply the method to a medical application, enabling experts to observe both the imprecision between groups and the importance of the features.

References

1. Antoine, V., Guerrero, J.A., Xie, J.: Fast semi-supervised evidential clustering. International Journal of Approximate Reasoning **133**, 116–132 (2021)
2. Bezdek, J.C., Ehrlich, R., Full, W.: Fcm: The fuzzy c-means clustering algorithm. Computers & geosciences **10**(2-3), 191–203 (1984)
3. Bilenko, M., Basu, S., Mooney, R.J.: Integrating constraints and metric learning in semi-supervised clustering. In: Proceedings of the twenty-first international conference on Machine learning. p. 11 (2004)
4. Cai, W., Zhai, B., Liu, Y., Liu, R., Ning, X.: Quadratic polynomial guided fuzzy c-means and dual attention mechanism for medical image segmentation. Displays **70**, 102106 (2021)
5. Calamai, P.H., Moré, J.J.: Projected gradient methods for linearly constrained problems. Mathematical programming **39**(1), 93–116 (1987)
6. Dave, R.N.: Robust fuzzy clustering algorithms. In: [Proceedings 1993] Second IEEE International Conference on Fuzzy Systems. pp. 1281–1286. IEEE (1993)
7. Deng, Z., Choi, K.S., Jiang, Y., Wang, J., Wang, S.: A survey on soft subspace clustering. Information sciences **348**, 84–106 (2016)
8. Frigui, H., Nasraoui, O.: Unsupervised learning of prototypes and attribute weights. Pattern recognition **37**(3), 567–581 (2004)
9. Gath, I., Geva, A.B.: Unsupervised optimal fuzzy clustering. IEEE Transactions on pattern analysis and machine intelligence **11**(7), 773–780 (1989)
10. Gustafson, D., Kessel, W.: Fuzzy clustering with a fuzzy covariance matrix. In: 1978 IEEE conference on decision and control including the 17th symposium on adaptive processes. pp. 761–766. IEEE (1979)
11. Hashemzadeh, M., Oskouei, A.G., Farajzadeh, N.: New fuzzy c-means clustering method based on feature-weight and cluster-weight learning. Applied Soft Computing **78**, 324–345 (2019)
12. Hubert, L., Arabie, P.: Comparing partitions. Journal of classification **2**, 193–218 (1985)
13. Hung, W.L., Yang, M.S., Chen, D.H.: Bootstrapping approach to feature-weight selection in fuzzy c-means algorithms with an application in color image segmentation. Pattern Recognition Letters **29**(9), 1317–1325 (2008)
14. Jing, L., Ng, M.K., Huang, J.Z.: An entropy weighting k-means algorithm for subspace clustering of high-dimensional sparse data. IEEE Transactions on knowledge and data engineering **19**(8), 1026–1041 (2007)
15. Krishnapuram, R., Keller, J.M.: A possibilistic approach to clustering. IEEE transactions on fuzzy systems **1**(2), 98–110 (1993)
16. Masson, M.H.: Cecm: Constrained evidential c-means algorithm. Computational Statistics & Data Analysis **56**(4), 894–914 (2012)
17. Masson, M.H., Denœux, T.: Ecm: An evidential version of the fuzzy c-means algorithm. Pattern Recognition **41**(4), 1384–1397 (2008)
18. Mehta, D., Saxena, S.: Hierarchical wsn protocol with fuzzy multi-criteria clustering and bio-inspired energy-efficient routing (fmcb-er). Multimedia Tools and Applications **81**(24), 35083–35116 (2022)

19. Ojeda-Magana, B., Ruelas, R., Corona-Nakamura, M., Andina, D.: An improvement to the possibilistic fuzzy c-means clustering algorithm. In: 2006 World Automation Congress. pp. 1–8. IEEE (2006)
20. Pal, N.R., Pal, K., Keller, J.M., Bezdek, J.C.: A possibilistic fuzzy c-means clustering algorithm. IEEE transactions on fuzzy systems **13**(4), 517–530 (2005)
21. Pimentel, B.A., de Souza, R.M.: Multivariate fuzzy c-means algorithms with weighting. Neurocomputing **174**, 946–965 (2016)
22. Pimentel, B.A., de Souza, R.M.: A generalized multivariate approach for possibilistic fuzzy c-means clustering. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems **26**(06), 893–916 (2018)
23. Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O.P., Tiwari, A., Er, M.J., Ding, W., Lin, C.T.: A review of clustering techniques and developments. Neurocomputing **267**, 664–681 (2017)
24. Shafer, G.: A mathematical theory of evidence, vol. 42. Princeton university press (1976)
25. Wang, C., Pedrycz, W., Li, Z., Zhou, M.: Residual-driven fuzzy c-means clustering for image segmentation. IEEE/CAA Journal of Automatica Sinica **8**(4), 876–889 (2020)
26. Xing, H.J., Ha, M.H.: Further improvements in feature-weighted fuzzy c-means. Information Sciences **267**, 1–15 (2014)
27. Yager, R.R.: On the normalization of fuzzy belief structures. International Journal of Approximate Reasoning **14**(2-3), 127–153 (1996)
28. Yang, M.S., Benjamin, J.B.: Feature-weighted possibilistic c-means clustering with a feature-reduction framework. IEEE Transactions on Fuzzy Systems **29**(5), 1093–1106 (2020)
29. Zadeh, L.A.: Fuzzy sets. Information and control **8**(3), 338–353 (1965)