Pattern Recognition 159 (2025) 111108

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

MvWECM: Multi-view Weighted Evidential *C*-Means clustering

Kuang Zhou [a,b], Yuchen Zhu [a], Mei Guo [a,*], Ming Jiang [a]

[a] School of Mathematics and Statistics, Northwestern Polytechnical University, Xi'an 710129, PR China
[b] MOE Key Laboratory for Complexity Science in Aerospace, Northwestern Polytechnical University, Xi'an 710129, PR China

ARTICLE INFO

Keywords:
Multi-view clustering
Credal partitions
Uncertainty
View weights

ABSTRACT

Traditional multi-view clustering algorithms, designed to produce hard or fuzzy partitions, often neglect the inherent ambiguity and uncertainty in the cluster assignment of objects. This oversight may lead to performance degradation. To address these issues, this paper introduces a novel multi-view clustering method, termed MvWECM, capable of generating credal partitions within the framework of belief functions. The objective function of MvWECM is introduced considering the uncertainty in the cluster structure included in the multi-view dataset. We take into account inter-view conflict to effectively leverage coherent information across different views. Moreover, the effectiveness is heightened through the incorporation of adaptive view weights, which are customized to modulate their smoothness in accordance with their entropy. The optimization method to get the optimal credal membership and class prototypes is derived. The view wights can be also provided as a by-product. Experimental results on several real-word datasets demonstrate the effectiveness and superiority of MvWECM by comparing with some state-of-the-art methods.

1. Introduction

Clustering is a prevalent unsupervised learning technique that seeks to partition the samples in the datasets into groups based on certain criteria. This process is designed to ensure that samples in the same group exhibit a high level of similarity, while those in different groups manifest a low level of similarity. The rapid development of multimedia technology has allowed us to gather real-world data from multiple viewpoints, leading to the emergence of multi-view data [1]. For example, a news story might have commentaries available in English, French, or German. Similarly, in image analysis, a single image can be represented using different visual descriptors such as HOG, SIFT, GIST, and LBP. This type of data is referred to as multi-view data, where each set of features represents a distinct viewpoint [2].

Data from multiple views often display both consistency and complementarity. Effectively leveraging multi-view data while accounting for its consistency and complementarity is fundamental to advancing multi-view clustering techniques. Traditional clustering methods are typically designed for single-view data. Although it is possible to merge all feature sets from multiple views into a single one and then apply a single-view clustering method, this strategy disregards the relationships among the views and fails to effectively manage the rich information inherent in multi-view data [3].

As an emerging field of research, Multi-View Clustering (MVC) has received extensive attention recently [4]. A variety of algorithms have been proposed for MVC, encompassing classic methods like spectral clustering, subspace clustering, and manifold learning, as well as more recent approaches based on deep learning. Zhong et al. [3] provided two versions of Self-taught Multi-view Spectral Clustering (SMSC) based on convex combination and centroid graph fusion schemes. Nie et al. [5] made a multi-view extension of the spectral rotation technique raised in single view spectral clustering research, and proposed an adaptive weighted procrustes to consider the varying clustering capacities of different views. Khan and Hu et al. [6] introduced a new approach to multi-view subspace clustering. This method is capable of learning a joint representation that takes into account both the common aspects shared across multiple views and the unique characteristics of each independent view. In [7], the authors proposed a multi-view clustering methodology with consensus manifold regularization, built upon the concept factorization technique. Some researchers have proposed deep learning-based multi-view clustering methods, which allow them to utilize additional techniques like the auto-attention mechanism [8], aimed at calculating the similarity between different samples, and graph representation [9], which incorporates an intrinsic contrastive network to ensure the consistency of node representations.

Prototype-based clustering, such as *k*-means [10], is one of the most widely adopted clustering methods due to its straightforward implementation and intuitive nature. It represents each cluster with

* Corresponding author.
E-mail addresses: kzhoumath@nwpu.edu.cn (K. Zhou), h1nkik@mail.nwpu.edu.cn (Y. Zhu), guomei@mail.nwpu.edu.cn (M. Guo), xming@mail.nwpu.edu.cn (M. Jiang).

https://doi.org/10.1016/j.patcog.2024.111108
Received 6 December 2023; Received in revised form 20 September 2024; Accepted 3 October 2024
Available online 1 November 2024

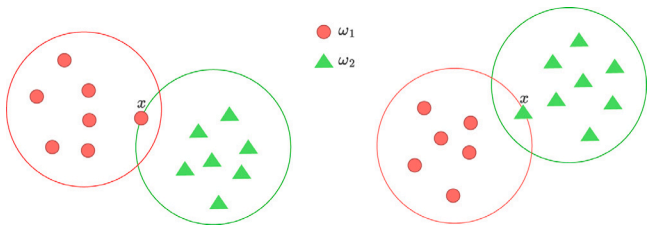


Fig. 1. An illustrative example of the uncertain cluster assignment in multi-view clustering.

a prototype and utilizes a relocation mechanism to iteratively reassign data points to clusters. This approach has also been adapted for multi-view data. Zhang et al. [11] proposed the Two-level Weighted COllaborative K-Means algorithm (TW-CO-KM). It enhanced the traditional k-means algorithm by introducing a novel mechanism for view and feature weights in the objective function. Yang et al. [12] extended Fuzzy C-Means (FCM) [13] clustering and proposed the Collaborative Feature-Weighted Multi-View FCM (Co-FW-MVFCM). It can generate fuzzy partitions for multi-view datasets, enabling the tolerance of uncertain data points, which hard methods are incapable of handling. Han et al. [14] proposed a multi-view k-means clustering with adaptive sparse memberships and weight allocation.

As can be seen, these multi-view clustering algorithms typically rely on either hard or fuzzy partitions. However, given the conflicting information inherent in multi-view data, objects might or might not clearly belong to a specific cluster [15]. Fig. 1 illustrates a clustering task on a two-view dataset consisting of fourteen samples. As shown in the figure, the sample x is classified into class ω_1 in the first view but into class ω_2 in the second view. From the figure we can see this sample is situated in the overlapping area between these two classes. With the information available, it is difficult to determine which specific class it should be assigned to.

Most existing multi-view clustering methods often struggle with handling uncertainty and imprecision in the cluster structure. Credal partitions, as delineated within the framework of belief functions [16,17], permit ambiguity, uncertainty, or doubt in the assignment of objects to clusters [18]. Evidential C-Means (ECM) [19], as an extension of FCM in belief functions frame, is one of the most commonly used evidential clustering methods to create credal partitions. Following the framework of ECM, numerous clustering approaches have subsequently been introduced. The constrained ECM (CECM) was proposed to introduce the prior knowledge into the objective function [20]. In [21], credal c-means (CCM) was introduced to redefine the distance between prototypes and instances. This adjustment aimed to prevent unreasonable results when the prototypes of an imprecise cluster are close to the centers of some specific classes. Gong et al. [22] introduced belief-peaks evidential clustering, expanding on traditional density peaks clustering within the belief function framework. Jiao et al. [23] developed a decision tree based evidential clustering (DTEC) algorithm. In [24], an evidential transfer clustering method was introduced for the clustering task when the data are uncertain or insufficient. Unfortunately, these methods inadequately address the challenges posed by multi-view data.

In this paper[1], we propose the Multi-view Weighted Evidential C-Means (MvWECM) clustering algorithm as an extension of ECM, enabling its application to multi-view datasets. We outline the main contributions of our work as follows:

- We introduce an evidential clustering method for multi-view data based on the theory of belief functions. This method can effectively utilize information from multiple sources to accomplish the task of clustering. The resulting global credal partition effectively addresses uncertainties that may stem from potential disagreements between different views.
- In the objective function, the view weights as well as their associated entropy are incorporated to quantify the degree of contributions of various views. At the same time, we introduce a term of collaborative strength which can capture shared information and complementarity between different views to exploit explicit cluster structure.
- Using the Coordinate Descent method and Lagrange multiplier optimization method, the optimal basic belief assignments, view weights and class prototypes are derived, respectively. In addition, we discuss the complexity of the proposed clustering algorithm.
- Experimental results demonstrate that MvWECM outperforms the state-of-the-art methods on seven benchmarks, as statistically evidenced by three evaluation metrics.

The remainder of this paper is organized as follows. In Section 2, we introduce some basic concepts and related works. Section 3 provides a detailed understanding of the proposed method. In Section 4, performance analysis through extensive experiments is reported. In Section 5, we conclude the paper.

2. Background

This section will briefly introduce the concepts related to belief function theory (also known as Dempster–Shafer theory (DST) or evidence theory) and ECM. Some related works of MVC are introduced as well.

2.1. Theory of belief functions

Consider a problem with c distinct and mutually exclusive elements. The frame of discernment can be defined as

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_c\}.$$

The set formed by all subsets of Ω is called the power set of Ω (containing 2^c elements) and is represented as follows:

$$2^{\Omega} = \{\emptyset, \{\omega_1\}, \{\omega_2\}, \dots, \{\omega_c\}, \{\omega_1, \omega_2\}, \dots, \{\omega_1, \omega_2, \dots, \omega_i\}, \dots, \Omega\}. \quad (1)$$

In the application of evidential theory, the obtained uncertain information is called evidence. It is usually necessary to represent the evidence in a power set framework, which enables DST to give estimates not only for an individual event but also for a collection of events. This approach significantly enhances the capacity to express uncertain information effectively. There are three general ways of evidence representation, including Basic Belief Assignment Function (BBA), Belief Function (Bel) and Plausibility Function (Pl).

If there is a set mapping function $m : 2^{\Omega} \to [0, 1]$ defined on 2^{Ω} that satisfies

$$\sum_{A \in 2^{\Omega}} m(A) = 1, m(A) \geq 0, \quad (2)$$

m is called a mass function, also known as BBA. For a subset $A \subseteq \Omega$, A is said to be a focal element if it satisfies $m(A) > 0$. We remark here that $m(\emptyset)$ may be positive. When $m(\emptyset) = 0$, m is called a normalized mass function. The nature of $m(\emptyset) > 0$ in the open world assumption is discussed in [26]. In this paper, we use $\emptyset$ to denote the noisy cluster, which can be thought of as representing outliers within the dataset.

Assuming that $m(\cdot)$ is the basic belief assignment function, the belief function and the plausibility function $Pl : 2^{\Omega} \to [0, 1]$ are defined as

$$Bel(A) = \sum_{B \subseteq A, B \neq \emptyset} m(B), \quad \forall A \subseteq \Omega \quad (3)$$

[1] This paper is an extension and revision of our previous work Zhou et al. (2021)[25].

and

$$Pl(A) = \sum_{B \cap A \neq \emptyset} m(B), \quad \forall A \subseteq \Omega \tag{4}$$

respectively. For any subset $A \in 2^{\Omega}$, the $Bel(A)$ is the sum of the basic belief assignments corresponding to all subsets of A, describes the total amount of support for the proposition A, and constitutes a lower bound on the support for the proposition A. The $Pl(A)$ is the sum of the basic belief assignments corresponding to all subsets intersecting A that are not empty, which describes the degree of non-rejection of the proposition A and constitutes the upper bound of the support of the proposition A. That is, $[Bel(A), Pl(A)]$ is the support interval for proposition A, and the relationship between the belief function and the plausibility function can be expressed as

$$Pl(A) = 1 - m(\emptyset) - Bel(\overline{A}), \tag{5}$$

where $\overline{A}$ denotes the complement of A in Ω.

Smets [27] developed the transferable belief model, which is a two-level mental model. At the credal level, beliefs are formed using mass functions, while at the Pignistic level, decisions are made by deriving a probability function from the mass functions. Smets also proposed a way to convert the mass function on a power set into a probability distribution over Ω. For a given mass function m defined on 2^{Ω}, the corresponding Pignistic probability, denoted by $BetP$, is defined as

$$BetP(\omega_i) = \sum_{A \ni \omega_i} \frac{m(A)}{|A|(1 - m(\emptyset))}, \omega_i \in \Omega, \tag{6}$$

where $|A|$ denotes the number of elements contained in the set $A \subseteq \Omega$.

2.2. Evidential C-means

Consider a dataset with n samples, which is to be grouped into c clusters. Define the sample set and the cluster set by

$$\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$$

and

$$\Omega = \{\omega_1, \ldots, \omega_c\}$$

respectively. Let p denote the dimension of the data. Under the discriminative framework Ω, the basic belief assignment of the i_{th} sample is noted as

$$m_i = \left\{ m_i\left(A_k\right) : A_k \subseteq \Omega \right\}, i = 1, 2, \ldots, n.$$

The ECM algorithm obtains the optimal belief assignment matrix $\boldsymbol{M} = \{m_i(A_k)\}_{n \times 2^c}$ which is known as a credal partition and the class prototype matrix $\boldsymbol{V} = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_c}\}$ by minimizing the following objective function:

$$J_{\text{ECM}}(\boldsymbol{M}, \boldsymbol{V}) = \sum_{i=1}^{n} \sum_{A_k \subseteq \Omega, A_k \neq \emptyset} |A_k|^{\alpha} m_i(A_k)^{\beta} d_{ik}^2 + \sum_{i=1}^{n} \delta^2 m_i(\emptyset)^{\beta}, \tag{7}$$

s.t.

$$\sum_{A_k \subseteq \Omega, A_k \neq \emptyset} m_i(A_k) + m_i(\emptyset) = 1, \tag{8}$$

$$m_i\left(A_k\right) \geq 0, \quad m_i\left(\emptyset\right) \geq 0. \tag{9}$$

The parameters α, β and δ are adjustable. α penalizes the focal element with a high number of elements. As a smoothing factor, similar to the exponential weight parameter in the FCM algorithm, β controls the usual degree of affiliation [28]. The parameter δ is adjusted according to the presence or absence of extraneous points in the data. In the objective function J_{ECM}, d_{ik} denotes the distance (Euclidean distance) between the sample $\boldsymbol{x_i}$ and the focal element class center/class prototype (denoted as $\overline{\boldsymbol{v}}_{\boldsymbol{k}}$), which is calculated by the following equation:

$$d_{ik} = \|\boldsymbol{x_i} - \overline{\boldsymbol{v}}_{\boldsymbol{k}}\|_2, \tag{10}$$

where the focal element class center $\overline{\boldsymbol{v}}_{\boldsymbol{k}}$ is calculated as

$$\overline{\boldsymbol{v}}_{\boldsymbol{k}} = \frac{1}{|A_k|} \sum_{h=1}^{c} s_{hk} \boldsymbol{v_h}, \quad \text{with} \quad s_{hk} = \begin{cases} 1 & \text{if } \omega_h \in A_k \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

In Eq. (11), the symbol $\boldsymbol{v_h}$ denotes the geometric centers of all sample points in the category ω_h. Let $m_{ij} \triangleq m_i(A_j)$, which denotes the basic belief assignment of the point $\boldsymbol{x}_i$ with respect to the non-empty focal element A_j. Let $m_{i\emptyset} \triangleq m_i(\emptyset)$ denote the basic belief assignment of the sample $\boldsymbol{x}_i$ classified to the empty set.

To minimize J_{ECM}, an alternating optimization scheme based on the Lagrange multiplier method can be employed, similar to the approach used in FCM algorithms. First, we consider that $\boldsymbol{V}$ is fixed. To address the constrained minimization problem with respect to $\boldsymbol{M}$, the n Lagrange multipliers λ_i can be introduced to construct the Lagrangian:

$$\mathscr{L}(\boldsymbol{M}, \lambda_1, \ldots, \lambda_n) = J_{\text{ECM}}(\boldsymbol{M}, \boldsymbol{V}) - \sum_{i=1}^{n} \lambda_i \left(\sum_{A_j \subseteq \Omega, A_j \neq \emptyset} m_{ij} + m_{i\emptyset} - 1 \right) \tag{12}$$

By differentiating the Lagrangian with respect to the m_{ij}, $m_{i\emptyset}$, and λ_i and setting the derivatives to zero, the equations for updating the basis belief assignment matrix for the credal membership of objects can be derived as follows:

$$m_{ik} = \frac{|A_k|^{-\alpha/(\beta-1)} d_{ik}^{-2/(\beta-1)}}{\sum_{A_h \neq \emptyset} |A_h|^{-\alpha/(\beta-1)} d_{ih}^{-2/(\beta-1)} + \delta^{-2/(\beta-1)}}, \tag{13}$$

$$m_{i\emptyset} = 1 - \sum_{A_k \neq \emptyset} m_{ik}, \quad i = 1, 2, \ldots, n. \tag{14}$$

Then we can consider that $\boldsymbol{M}$ is fixed to update the prototype matrix $\boldsymbol{V}$, which can be seen as an unconstrained optimization problem. Setting the partial derivatives of J_{ECM} with respect to the centers to zero, we can get the following linear system to update the class prototype matrix $\boldsymbol{V}$:

$$\boldsymbol{H}\boldsymbol{V} = \boldsymbol{B}, \tag{15}$$

where the symbol $\boldsymbol{H}$ represents a matrix of $(c \times c)$, defined as

$$\boldsymbol{H}_{lk} = \sum_{i=1}^{n} \sum_{A_k \supseteq \{\omega_k, \omega_l\}} |A_k|^{\alpha-2} m_{ik}^{\beta}. \tag{16}$$

The symbol $\boldsymbol{B}$ represents a matrix of $(c \times p)$, defined as

$$\boldsymbol{B}_{lq} = \sum_{i=1}^{n} x_{iq} \sum_{A_k \ni \omega_l} |A_k|^{\alpha-1} m_{ik}^{\beta}. \tag{17}$$

To demonstrate the benefits of credal partitions in addressing the challenges caused by uncertainty in clustering, we provide the following artificial example.

Example. Consider the dataset shown in Fig. 2. We denote 2 clusters as $\Omega = \{\omega_1, \omega_2\}$ with their class centers by green dots. It is clear that the two clusters significantly exhibit overlapping. We consider three data points $\{x_1, x_2, x_3\}$ marked by dark triangles in the figure. It is easy to see, the three points have an equal distance to the centers of the two classes. By the method of ECM and FCM, we can get their credal membership and fuzzy membership denoted as $\{m_1, m_2, m_3\}$ and $\{u_1, u_2, u_3\}$, respectively. The results are shown in Table 1.

As we can see, for the uncertain data points x_1, x_2, x_3, it is hard to assign them directly to a specific cluster. For the fuzzy partition, we have $u_i(\{\omega_1\}) = u_i(\{\omega_2\}) = 0.5, i = 1, 2, 3$. However, since x_2 and x_3 are closer to the boundary of the cluster $\{\omega_1\}$, they should not have the same affiliation as x_1. Instead of relying exclusively on singletons, credal partitions can express our knowledge pertaining to the membership of samples for composite clusters and the empty set. As the distance of the samples x_1, x_2, x_3 from the center of the singletons (the green ones) and the center of the composite cluster (the orange one) increases, making the affiliation decrease with it, while

Table 1
The credal and fuzzy partitions of specific example.

Focal sets	m_1	m_2	m_3	u_1	u_2	u_3
$\emptyset$	0	0.3	0.7	–	–	–
$\{\omega_1\}$	0.3	0.2	0.1	0.5	0.5	0.5
$\{\omega_2\}$	0.3	0.2	0.1	0.5	0.5	0.5
$\Omega = \{\omega_1, \omega_2\}$	0.4	0.3	0.1	–	–	–

the affiliation of the assignment to the empty set $\emptyset$ increases with it. In specific, as for x_3, it can be considered as an outlier instead of being assigned to $\{\omega_1\}$ or $\{\omega_2\}$. In sum, credal partitions have a more tolerant ability for uncertain information.

2.3. Multi-view clustering

As previously mentioned, multi-view clustering stands out as an intriguing subject in machine learning. Here some compelling approaches within the realm of MVC are concisely introduced.

- MCGC [29]: It learns consensus information by using a new disagreement cost function for regularizing graphs and a constrained Laplacian matrix without any post-processing like k-means.
- AWP [5]: It extends spectral rotation for multi-view data and applies an adaptively weighted Procrustes technique which aims to overcome the deficiency of that extension. And AWP is parameter-free method, which makes it more applicable.
- TW-CO-KM [11]: It extends k-means by considering both view weights and feature weights. A penalty term in objective function is designed to measure the disagreement in a collaborative manner.
- CO-FW-MVFCM [12]: This approach contains a two-step schema under a framework of fuzzy, including a local step and a collaborative step. The former produces local single-view partition, while the latter shares information between views.
- MVASM [14]: It focuses on constructing a sparse membership matrix and learning the centroid matrix and its corresponding view weights. Simultaneously, it converts hard partitions into fuzzy ones.
- MCHC [30]: It is a multi-view adjacency-constrained hierarchical clustering approach. The method incorporates the fusion distance matrices with extreme weights, adjacency-constrained nearest neighbor clustering and the internal evaluation index to achieve the promising clustering performance.
- CHOC-MVC [31]: It is a multi-view subspace clustering method which can unify the consistency and specificity of data in tensor manner. This model can effectively capture the high-order consistent information while eliminating classification barriers caused by specific information.

3. Multi-view weighted evidential clustering

In this section, we begin by providing a detailed presentation of the proposed algorithm MvWECM. Subsequently, the optimization process is discussed in-depth. The complexity is also analyzed at the end of this section.

3.1. The objective function

Let $\boldsymbol{X} = \{\boldsymbol{X}[1], \boldsymbol{X}[2], \ldots, \boldsymbol{X}[t], \ldots, \boldsymbol{X}[T]\}$ denote the dataset with T views, and each view has n samples, that is,

$$\boldsymbol{X}[t] = \{\boldsymbol{x}_1[t], \boldsymbol{x}_2[t], \ldots, \boldsymbol{x}_i[t], \ldots, \boldsymbol{x}_n[t]\}, t = 1, 2, \ldots, T.$$

Suppose the dimension of the data in the t_{th} view be q_t. As before, we assume that the dataset has c classes denoted by $\Omega = \{\omega_1, \omega_2, \ldots, \omega_c\}$. Let $\boldsymbol{w} = \{w[t], t = 1, 2, \ldots, T\}$ denote the set of view weights, where

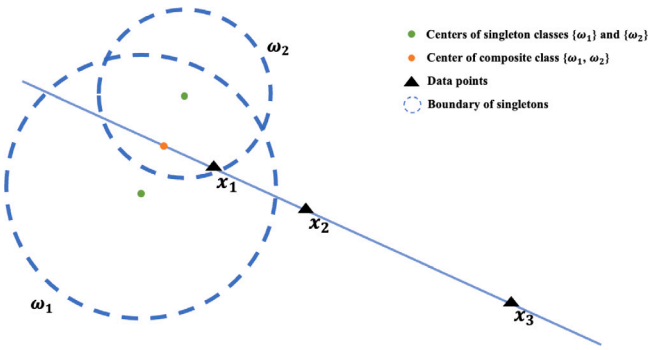


Fig. 2. An illustrative example comparing fuzzy partition and credal partition.

$w[t](w[t] \geq 0)$ is the weight of the t_{th} view. $m_{ij}[t]$ illustrates the basic belief assignment of $\boldsymbol{x}_i$ for the class A_j in the t_{th} view, while $m_{i\emptyset}[t]$ illustrates the basic belief assignment to the empty set in this view. The matrix of belief assignment of all the samples for the t_{th} view is denoted as $\boldsymbol{M}[t]$. Similar to δ in ECM algorithm, the parameter $\delta[t]$ reflects whether the outliers exist in the t_{th} view. $d_{ij}[t]$ denotes the Euclidean distance between the i_{th} sample and the center of class A_j in t_{th} view. The Euclidean distance is defined as

$$d_{ij}[t] = \|\boldsymbol{x}_i[t] - \overline{\boldsymbol{v}}_j[t]\|_2, \tag{18}$$

where $\overline{\boldsymbol{v}}_j[t]$ is the prototype of class A_j ($A_j \subset \Omega, A_j \neq \emptyset$) in the t_{th} view. It can be derived as

$$\overline{\boldsymbol{v}}_j[t] = \frac{1}{c_j} \sum_{h=1}^{c} s_{hj} \boldsymbol{v}_h[t], \quad \text{with} \quad s_{hj} = \begin{cases} 1 & \text{if } \omega_h \in A_j \\ 0 & \text{otherwise.} \end{cases} \tag{19}$$

In the above equation, c_j represents the cardinality of the focal set A_j, and $\boldsymbol{v}_h[t]$ is the prototype of the specific class $\omega_h \in \Omega$. When $c_j = 1$, A_j represents a specific class, and $\overline{\boldsymbol{v}}_j[t]$ is the center of all samples within that class. When $c_j > 1$, A_j represents an imprecise class, and $\overline{\boldsymbol{v}}_j[t]$ is the center of the prototypes of all included specific classes.

Under the framework of belief functions, with multi-view data, the objective function of MvWECM can be defined as

$$\begin{aligned} J_{\text{MvWECM}}(\boldsymbol{M}, \boldsymbol{V}, \boldsymbol{w}) = & \sum_{t=1}^{T} w[t] \left(\sum_{i=1}^{n} \sum_{\{j | A_j \subseteq \Omega, A_j \neq \emptyset\}} c_j^{\alpha} m_{ij}^2[t] d_{ij}^2[t] \right. \\ & \left. + \sum_{i=1}^{n} \delta^2[t] m_{i\emptyset}^2[t] \right) \\ & + \eta \sum_{t=1}^{T} \sum_{s \neq t}^{T} K[t, s] \sum_{i=1}^{n} \sum_{\{j | A_j \subseteq \Omega, A_j \neq \emptyset\}} \left(m_{ij}[t] - m_{ij}[s] \right)^2 d_{ij}^2[t] \\ & + \beta \sum_{t=1}^{T} w[t] \log w[t], \end{aligned} \tag{20}$$

s.t.

$$\sum_{t=1}^{T} w[t] = 1, \ w[t] \in (0, 1], \ t = 1, 2, \ldots, T, \tag{21}$$

$$\sum_{\{j | A_j \subseteq \Omega, A_j \neq \emptyset\}} m_{ij}[t] + m_{i\emptyset}[t] = 1, \ \forall A_j \subseteq \Omega, i = 1, 2, \ldots, n; t = 1, 2, \ldots, T. \tag{22}$$

As we can see, there are three terms in the objective function. The first term computes the weighted sum of squared distances for samples within each view from the class prototype. This is achieved through a strategy of assigning weights to the views. The objective is to minimize the distance between the samples within a view and the class prototype after clustering, while maximizing the weights of views that yield the best clustering results.

The second term aims to minimize the disagreements between different views. We introduce the factor $K[t,s] \in [0,1]$. We assume that there is no collaboration between the view and itself. Thus, $K[t,s]$ is set to 0 when $t = s$. The default value of $K[t,s]$ is set to 1 when $t \neq s$ for simplification. The parameter η is used to measure the degree of influence of this term.

In the third term, to control for extreme changes in view weights, the Shannon entropy of view weights is introduced to improve the smoothness of the weight distribution. β is the Shannon's entropy of weight parameter, which assists in adaptively regulating the distribution of view weights so that they are as evenly distributed as possible while ensuring the optimal clustering of view weights. In brief, it regulates the impact of this term on the clustering results.

3.2. Update and optimization

In this subsection, we present the solution to the optimization problem from an analytical perspective. We employ the well-established Coordinate Descent (CD) algorithm [32], known for its simplicity and efficiency in non-gradient optimization, to minimize the defined objective function.

Mathematically, focusing $\min_{x_1,x_2,\ldots,x_n \in \Theta} f\left(x_1, x_2, \ldots, x_n\right)$, where Θ is an arbitrary constraint and f is the objective function. If $x^k = \left(x_1^k, \ldots, x_{i-1}^k, x_i^k, x_{i+1}^k, \ldots, x_n^k\right)$ is given as the current iteration at the k_{th} iteration, the algorithm generates the next iteration x^{k+1} according to the solutions of the following n subproblems

$$x_i^{k+1} \leftarrow \operatorname{argmin}_y f\left(x_1^{k+1}, \ldots, x_{i-1}^{k+1}, y, x_{i+1}^k, \ldots, x_n^k\right), i = 1, 2, \ldots, n.$$

This makes sure that f does not increase. Back to our $\min_{\boldsymbol{M},\boldsymbol{V},\boldsymbol{w}} J_{\text{MvWECM}}$ with linear constraints, we update one variable while holding others fixed, leveraging the method of Lagrange multipliers to solve these subproblems. Through this alternating cyclic process, achieving convergence to a local optimum is ensured. The following outlines the detailed computational process.

- Update the credal membership

We start by defining the update rule for the belief assignment $\boldsymbol{M}[t]$ by assuming the view weight $w[t]$ and class center $\boldsymbol{V}[t]$ is fixed. To solve the constrained minimization problem with respect to $\boldsymbol{M}[t]$, $n \times T$ Lagrange multipliers $\lambda_i[t]$ are introduced. The resulting Lagrange function $L(\boldsymbol{M}[t], \lambda[t])$ is

$$L(\boldsymbol{M}[t], \lambda[t]) = J_{\text{MvWECM}} - \sum_{t=1}^{T} \sum_{i=1}^{n} \lambda_i[t] \left(\sum_{\{j | A_j \subseteq \Omega, A_j \neq \emptyset\}} m_{ij}[t] + m_{i\emptyset}[t] - 1 \right). \tag{23}$$

Differentiating the above equation with respect to $m_{ij}[t]$, $m_{i\emptyset}[t]$ and $\lambda_i[t]$ (determined i, j, t) and letting the differential equation be zero, one obtains

$$\frac{\partial L(\boldsymbol{M}[t], \lambda[t])}{\partial m_{ij}[t]} = 2w[t] c_j^{\alpha} m_{ij}[t] d_{ij}^2[t] + 2\eta \sum_{s \neq t}^{T} K[t,s](m_{ij}[t] - m_{ij}[s]) d_{ij}^2[t] - \lambda_i[t] = 0, \tag{24}$$

$$\frac{\partial L(\boldsymbol{M}[t], \lambda[t])}{\partial m_{i\emptyset}[t]} = 2w[t]\delta^2[t] m_{i\emptyset}[t] - \lambda_i[t] = 0, \tag{25}$$

and

$$\frac{\partial L(\boldsymbol{M}[t], \lambda[t])}{\partial \lambda_i[t]} = - \sum_{\{j | A_j \subseteq \Omega, A_j \neq \emptyset\}} m_{ij}[t] - m_{i\emptyset}[t] + 1 = 0. \tag{26}$$

From Eqs. (24) and (25), it follows that:

$$m_{ij}[t] = \frac{\eta \varphi_{ij}[t]}{w[t] c_j^{\alpha} + \eta \psi[t]} + \frac{\lambda_i[t]}{2\left(w[t] c_j^{\alpha} + \eta \psi[t]\right) d_{ij}^2[t]} \tag{27}$$

and

$$m_{i\emptyset}[t] = \frac{\lambda_i[t]}{2w[t]\delta^2[t]}, \tag{28}$$

where

$$\psi[t] = \sum_{s \neq t}^{T} K[t,s], \tag{29}$$

$$\varphi_{ij}[t] = \sum_{s \neq t}^{T} K[t,s] m_{ij}[s]. \tag{30}$$

Substituting Eqs. (27) and (28) into Eq. (26), the Lagrange multiplier $\lambda_i[t]$ can be represented by the view weight and class center as

$$\lambda_i[t] = \frac{1 - \sum_{\{j | A_j \subseteq \Omega, A_j \neq \emptyset\}} \frac{\eta \varphi_{ij}[t]}{w[t] c_j^{\alpha} + \eta \psi[t]}}{\sum_{\{j | A_j \subseteq \Omega, A_j \neq \emptyset\}} \frac{1}{2\left(w[t] c_j^{\alpha} + \eta \psi[t]\right) d_{ij}^2[t]} + \frac{1}{2w[t]\delta^2[t]}}. \tag{31}$$

Substituting Eq. (31) into Eq. (27), and by Eq. (26), the basic belief assignment $\boldsymbol{M}[t]$ is updated to

$$m_{ij}[t] = \frac{\eta \varphi_{ij}[t]}{w[t] c_j^{\alpha} + \eta \psi[t]} + \frac{\left(1 - \sum_{\{j | A_j \subseteq \Omega, A_j \neq \emptyset\}} \frac{\eta \varphi_{ij}[t]}{w[t] c_j^{\alpha} + \eta \psi[t]}\right) \frac{1}{\left(w[t] c_j^{\alpha} + \eta \psi[t]\right) d_{ij}^2[t]}}{\sum_{\{j | A_j \subseteq \Omega, A_j \neq \emptyset\}} \frac{1}{\left(w[t] c_j^{\alpha} + \eta \psi[t]\right) d_{ij}^2[t]} + \frac{1}{w[t]\delta^2[t]}}, \tag{32}$$

$$m_{i\emptyset}[t] = 1 - \sum_{\{j | A_j \subseteq \Omega, A_j \neq \emptyset\}} m_{ij}[t], \ i = 1, 2, \ldots, n; t = 1, 2, \ldots, T. \tag{33}$$

- Update the view weight

By fixing the basic belief assignment $\boldsymbol{M}[t]$ and class center $\boldsymbol{V}[t]$, the view weight $w[t]$ is updated using the Lagrange multiplier method, with the corresponding Lagrange function given as

$$L(w[t], \mu) = J_{\text{MvWECM}} - \mu \left(\sum_{t=1}^{T} w[t] - 1 \right), \tag{34}$$

where μ is the Lagrange multiplier. Differentiating $L(w[t], \mu)$ with respect to $w[t]$ and μ, respectively, it follows that:

$$\frac{\partial L(w[t], \mu)}{\partial w[t]} = \sum_{i=1}^{n} \sum_{\{j | A_j \subseteq \Omega, A_j \neq \emptyset\}} c_j^{\alpha} m_{ij}^2[t] d_{ij}^2[t] + \sum_{i=1}^{n} \delta^2[t] m_{i\emptyset}^2[t] + \beta(1 + \log w[t]) - \mu = 0, \tag{35}$$

and

$$\frac{\partial L(w[t], \mu)}{\partial \mu} = - \sum_{i=1}^{T} w[t] + 1 = 0. \tag{36}$$

Denote

$$\Delta[t] = \sum_{i=1}^{n} \sum_{\{j | A_j \subseteq \Omega, A_j \neq \emptyset\}} c_j^{\alpha} m_{ij}^2[t] d_{ij}^2[t] + \sum_{i=1}^{n} \delta^2[t] m_{i\emptyset}^2[t]. \tag{37}$$

Therefore, from Eq. (35), one obtains

$$\Delta[t] + \beta(\log w[t] + 1) - \mu = 0, \tag{38}$$

Eq. (38) is equivalent to

$$w[t] = \exp\left\{\frac{-\Delta[t] - \beta}{\beta}\right\} \exp\left\{\frac{\mu}{\beta}\right\}. \tag{39}$$

Substituting the above equation into Eq. (36), we have

$$\sum_{t=1}^{T} \exp\left\{\frac{-\Delta[t] - \beta}{\beta}\right\} \exp\left\{\frac{\mu}{\beta}\right\} = 1. \tag{40}$$

From the above equation, it is obtained that

$$\exp\left\{\frac{\mu}{\beta}\right\} = \frac{1}{\sum_{t=1}^{T}\exp\left\{\frac{-\Delta[t]-\beta}{\beta}\right\}}. \tag{41}$$

Substituting Eq. (41) into Eq. (39), we obtain the updated formula for the view weight $w[t]$

$$w[t] = \frac{\exp\left\{\frac{-\Delta[t]-\beta}{\beta}\right\}}{\sum_{i=1}^{T}\exp\left\{\frac{-\Delta[i]-\beta}{\beta}\right\}},\ t = 1, 2, \ldots, T. \tag{42}$$

- Update the class prototypes

Fix the basic belief assignments $\boldsymbol{M}[t]$ and view weight $w[t]$, and we can find the best clustering center $\boldsymbol{V}[t]$ by differentiating J_{MvWECM} with respect to the class center:

$$\frac{\partial J}{\partial \boldsymbol{v}_l[t]} = w[t]\sum_{i=1}^{n}\sum_{\{j|A_j\subseteq\Omega, A_j\neq\emptyset\}} c_j^{\alpha} m_{ij}^2[t]\frac{\partial d_{ij}^2[t]}{\partial \boldsymbol{v}_l[t]} + \eta\sum_{s\neq t}^{T} K[t,s]\sum_{i=1}^{n}\sum_{\{j|A_j\subseteq\Omega, A_j\neq\emptyset\}}(m_{ij}[t]-m_{ij}[s])^2\frac{\partial d_{ij}^2[t]}{\partial \boldsymbol{v}_l[t]}, \tag{43}$$

where

$$\frac{\partial d_{ij}^2[t]}{\partial \boldsymbol{v}_l[t]} = -2\frac{1}{c_j}s_{lj}\boldsymbol{x}_i[t] + 2\frac{1}{c_j^2}\sum_{h=1}^{c} s_{lj}s_{hj}\boldsymbol{v}_h[t], \tag{44}$$

$l = 1, 2, \ldots, c;\ t = 1, 2, \ldots, T.$

The following linear equation for solving $\boldsymbol{v}_h[t]$ can be obtained by making the differential Eq. (43) equal to zero:

$$\begin{aligned}&\sum_{h=1}^{c}\boldsymbol{v}_h[t]\sum_{s\neq t}^{T}\eta K[t,s]\sum_{i=1}^{n}\sum_{\{j|\{w_l,w_k\}\subseteq A_j\}}(m_{ij}[t]-m_{ij}[s])^2\frac{1}{c_j^2}\\ &+\sum_{h=1}^{c}\boldsymbol{v}_h[t]\sum_{i=1}^{n}\sum_{\{j|\{w_l,w_k\}\subseteq A_j\}} w[t]c_j^{\alpha-2}m_{ij}^2[t]\\ &=\sum_{i=1}^{n}\boldsymbol{x}_i[t]\sum_{\{j|w_l\in A_j\}} w[t]c_j^{\alpha-1}m_{ij}^2[t]\\ &+\sum_{i=1}^{n}\boldsymbol{x}_i[t]\sum_{s\neq t}^{T}\eta K[t,s]\sum_{\{j|w_l\in A_j\}}(m_{ij}[t]-m_{ij}[s])^2\frac{1}{c_j}.\end{aligned} \tag{45}$$

Let $\boldsymbol{B}[t]$ be the matrix of $(c\times q_t)$ defined by

$$B_{lk}[t] = \sum_{i=1}^{n} x_{ik}[t]\sum_{\{j|\omega_l\subseteq A_j\}} c_j^{\alpha-1}m_{ij}^2[t]w[t], l = 1, 2, \ldots, c; k = 1, 2, \ldots, q_t. \tag{46}$$

Let $\boldsymbol{B}[t,s]$ be the matrix of $(c\times q_t)$ defined by

$$B_{lk}[t,s] = \sum_{i=1}^{n} x_{ik}[t]\sum_{\{j|\omega_l\in A_j\}}(m_{ij}[t]-m_{ij}[s])^2\frac{1}{c_j}, \quad l = 1, 2, \ldots, c; k = 1, 2, \ldots, q_t. \tag{47}$$

Let $\boldsymbol{H}[t]$ be the matrix of $(c\times c)$ defined by

$$H_{lh}[t] = \sum_{i=1}^{n}\sum_{\{j|\{\omega_l,\omega_k\}\subseteq A_j\}} w[t]c_j^{\alpha-2}m_{ij}^2[t],\ l, h = 1, 2, \ldots, c. \tag{48}$$

Let $\boldsymbol{H}[t,s]$ be the matrix of $(c\times c)$ defined by

$$H_{lh}[t,s] = \sum_{i=1}^{n}\sum_{\{j|\{\omega_l,\omega_k\}\subseteq A_j\}}(m_{ij}[t]-m_{ij}[s])^2\frac{1}{c_j^2},\ l, h = 1, 2, \ldots, c. \tag{49}$$

Based on the above matrix definition, the class center update formula for the t_{th} view can be obtained as

$$\boldsymbol{V}[t] = \left(\boldsymbol{H}[t] + \sum_{s\neq t}^{T}\eta K[t,s]\boldsymbol{H}[t,s]\right)^{-1}\left(\boldsymbol{B}[t] + \sum_{s\neq t}^{T}\eta K[t,s]\boldsymbol{B}[t,s]\right). \tag{50}$$

Using the iterative update process described above, we can determine the optimal credal partition for each view. The final global credal partition is then derived by applying a weighted sum to the basic belief assignments from the various views. The formula is as follows:

$$\boldsymbol{M}^* = \sum_{t=1}^{T} w[t]\boldsymbol{M}[t]. \tag{51}$$

The proposed MvWECM method is summarized in Algorithm 1 in detail.

3.3. Complexity

As MvWECM is to provide credal partitions for each object, it distributes a fraction of the unit mass to each element of 2^{Ω} with $\Omega = \{\omega_1, \ldots, \omega_c\}$. When updating the belief assignment $\boldsymbol{M}$ and view weight $\boldsymbol{w}$, the complexity is $O(nT2^c)$. The complexity of other processes is below $O(nT2^c)$, where n, T, c are the number of objects, views and classes, respectively. Notice that the complexity grows exponentially with the number of class clusters c, it leads to more time cost. We can reduce the complexity of the method by considering a limited number of focal sets [19,33]. For example, the focal sets can be constrained to be either Ω, $\emptyset$, or to be composed of at most two classes, thereby reducing the complexity from 2^c to c^2. In this case, the complexity will be $O(nTc^2)$.

Algorithm 1 MvWECM.

Input: The dataset $\boldsymbol{X} = \{\boldsymbol{X}[1], \boldsymbol{X}[2], \cdots, \boldsymbol{X}[T]\}$, 4 non-negative parameters α, β, δ and η, the number of clusters c, the maximum number of iterations $i_{\max}$, the threshold ϵ.

1: **Initialization:** iteration $i = 0$, basic belief assignment matrix $\boldsymbol{M}[t] = \boldsymbol{M}_0[t]$, class prototype $\boldsymbol{V}[t] = \boldsymbol{V}_0[t]$, view weight $w[t] = 1/T$, strength of collaboration $K[t,t] = 0$, $\forall t = 1, 2, \cdots, T$.
2: Calculate the objective function J via Eq. (20).
3: **repeat**
4: $i \leftarrow i + 1$.
5: $J_{\text{old}} = J$.
6: Update the basic belief assignment $\boldsymbol{M}$ via Eq. (32) and Eq. (33).
7: Update the view weight $\boldsymbol{w}$ via Eq. (42).
8: Update the class center $\boldsymbol{V}$ via Eq. (50).
9: Update the objective function J via Eq. (20).
10: **until** $|J - J_{\text{old}}| < \epsilon$ or $i \geq i_{\max}$.
11: Calculate the global credal partition $\boldsymbol{M}^*$ via Eq. (51).

Output: Global credal partition $\boldsymbol{M}^*$, view weight $\boldsymbol{w}$, and class center $\boldsymbol{V}$ of each view.

4. Experiments

In this section, we test the effectiveness of the MvWECM approach by conducting thorough experiments on seven multi-view datasets, each involving eight competing methods. Three evaluation metrics are utilized, including two for hard partitions and one for credal partitions, making them comprehensively assess results of different methods. Our implementation is available at: https://github.com/H1nkik/MvWECM.

4.1. Datasets

We first briefly introduce seven real datasets used in our experiments.

- WebKB dataset4 (Webkb) [34]: It consists of 203 web-pages of 4 classes. Each web-page is described by the content of the page, the anchor text of the hyper-link, and the text in its title.

Table 2
Description of Benchmarks.

Dataset	Objects	Clusters	Views	Dimensions of features					
				View 1st	View 2nd	View 3rd	View 4th	View 5th	View 6th
Prok	551	4	3	438	3	393	–	–	–
Webkb	203	4	3	1703	203	203	–	–	–
IS	2100	7	2	9	10	–	–	–	–
Caltech07	1474	7	6	48	40	254	1984	512	928
3sources	169	6	3	3560	3631	3068	–	–	–
Reuters-1500	1500	6	5	21 531	24 893	34 279	15 506	11 547	–
Reuters-18758	18 758	6	2	10	10	–	–	–	–

- 3sources[2]: It contains 169 of 948 news articles dataset from three well-known online news sources: BBC, Reuters, and The Guardian, where each source is seen as one view.
- Prokaryotic phyla (Prok) [35]: It contains 551 prokaryotic species described with heterogeneous multi-view data including textual data and different genomic representations. Textual data consists of bag-of-words representation of documents describing prokaryotic species, while genomic representations include the proteome composition and gene repertoire representations. Proteome composition is encoded as the relative frequencies of amino acids and gene repertoire is encoded as the presence/absence indicators of gene families in a genome.
- Image Segmentation[3](training) (IS): 2100 of 2310 instances are drawn from a database of 7 outdoor images randomly and compose the training set. The images are hand segmented to create a classification for every pixel.
- Caltech101-7 (Caltech07) [36]: It contains 1474 pictures of objects belonging to 7 classes (*i.e.*, Face, Dollar-Bill, Motorbikes, Stop-Sign, Garfield, Snoopy, and Windsor-Chair). All images are described with six types of features (254 CENTRIST, 48 Gabor, 512 GIST, 1984 HOG, 928 LBP, and 40 wavelet moments).
- Reuters-1500 [36]: It collects 1500 documents which are expressed in five different languages (Italian, Spanish, French, German and English). All the documents are categorized into 6 classes.
- Reuters-18758 [37]: This dataset comprises 18758 samples as a subset of Reuters. It includes the English version along with translations in different languages. The dataset contains six classes, consistent with those in Reuters-1500.

The summary of these datasets is presented in Table 2.

4.2. Evaluation metrics

Drawing from prior research, we use three evaluation metrics to measure the clustering performance: Adjusted Rand Index (ARI) [38], Normalized Mutual Information (NMI) [39] and Evidential Precision (EP) [24].

As ARI and NMI are used to measure the quality of hard partitions, when evaluating the credal partitions provided by MvWECM and ECM, we derive the corresponding hard partitions by assigning each sample to the cluster with the highest Pignistic probability (see Eq. (6)). For these two metrics, the higher value indicates better clustering performance. In credal partitions, imprecise classes are introduced to capture the uncertainty associated with sample categories. When the available information does not permit the determination of a specific class, samples are classified into imprecise classes. This mechanism ensures the accuracy of the partition results for samples assigned into specific classes. To account for this, EP has been introduced to evaluate the assignment precision of credal partitions, which is defined as follows [24]:

$$EP = \frac{n_{er}}{N_e}, \tag{52}$$

where N_e denotes the number of pairs partitioned into the same specific class by credal partitions, while n_{er} is the number of relevant instance pairs out of these specifically clustered pairs. For hard partitions, where there is no imprecise classes, EP is equivalent to the classical Precision, which measures the ratio of relevant pairs (those in the same group in the benchmark) to the pairs found (those in the same group within the discovered clusters). The closer the EP is to 1, the better the result is.

4.3. Compared algorithms and parameters setup

- **Algorithms**. The proposed MvWECM is compared with ECM [19], and 7 SoTA multi-view clustering algorithms including 1 MVC based on learning consensus graph (MCGC) [29], 1 MVC based on hard clustering and weights of views and features (TW-CO-KM) [11], 1 MVC based on collaborative fuzzy clustering and sparse membership (MVASM) [14], 1 MVC based on fuzzy clustering and features reduction (CO-FW-MVFCM) [12], 1 MVC based on a weighted Procrustes analysis (AWP) [5], 1 MVC within a unified framework of fusion distance, adjacency-constrained nearest neighbor clustering and internal evaluation index (MCHC) [30], 1 MVC based on tensor learning and self-expressiveness learning (CHOC-MVC) [31]. As ECM is designed to handle single-view data, we merge the multi-view data in the order of the views to become a large dataset on which ECM runs.
- **Parameters**. All numbers of iterations, errors and parameters of all comparison methods are default as their papers (procedures) are. In detail, ECM follows default settings, namely, $\alpha = 1, \delta = 10$. MCGC and CO-FW-MVFCM have only one parameter β and m, respectively. Their values are said to be $\beta = 0.6$ and $m = 2$, consistent with the specifications in their papers. TW-CO-KM has three parameters. We apply to make parameters same with theirs, namely, $\beta = 10, \alpha = 70, \eta = 0.45$ for IS and $\beta = 8, \alpha = 45, \eta = 0.45$ for the others. MVASM has its unique range of parameters, and we have searched for the optimal parameters following the illustration, namely, $q \in [1, 6]$, step $= 0.01$; $\gamma \in [0, 6]$, step $= 0.1$. AWP is a non-parameters method that requires no additional parameters settings. For MCHC, no changes have been made from the default codes. For CHOC-MVC, we search the optimal parameters β and λ in the ranges of $[10^2, 10^5]$ and $[10^{-1}, 10^5]$ respectively, as suggested by the authors.

4.4. Experiment analysis

For one given dataset, each algorithm is repeated by 5 times which performed on an Intel(R) Core(TM) i7-13700 @2.10 GHz CPU with 24 GB RAM. The average values and standard deviations are reported. At the same time, the paired t-test is applied at significance level $\alpha = 0.05$ to verify whether the MvWECM outperforms statistically. The detailed results are shown in Table 3, Tables 4 and 5. In the tables,

[2] http://mlg.ucd.ie/datasets/3sources.html.
[3] https://archive.ics.uci.edu/dataset/50/image+segmentation.

Table 3
Adjusted Rand Index of different algorithms. ▲/▽ indicates our proposed algorithm is statistically superior or inferior to other algorithms at $\alpha = 0.05$. "OOM" means the method raises the out-of-memory failure.

	MvWECM	MCGC	ECM	TW-CO-KM	MVASM	CO-FW-MVFCM	AWP	CHOC-MVC	MCHC
Webkb	$0.3727_{.00}$	$0.3564_{.00}$▲	$-0.0574_{.00}$▲	$0.2022_{.27}$	$0.3597_{.02}$	$0.1098_{.15}$▲	$0.2959_{.00}$▲	$0.2561_{.00}$▲	$0.0000_{.00}$▲
3sources	$0.3324_{.03}$	$-0.0382_{.00}$▲	$0.0434_{.10}$▲	$0.1189_{.18}$▲	$0.2329_{.10}$▲	$0.1948_{.04}$▲	$0.2947_{.00}$▲	$0.0548_{.00}$▲	$0.4285_{.00}$▽
Prok	$0.2119_{.01}$	$-0.0062_{.00}$▲	$-0.0133_{.01}$▲	$0.1980_{.03}$	$0.1300_{.00}$▲	$0.1198_{.06}$▲	$0.0876_{.00}$▲	$0.1722_{.00}$▲	$0.1256_{.00}$▲
IS	$0.1113_{.00}$	$0.0000_{.00}$▲	$0.0725_{.00}$▲	$0.0965_{.01}$▲	$0.0797_{.00}$▲	$0.0292_{.02}$▲	$0.0660_{.00}$▲	$0.0674_{.00}$▲	$0.0002_{.00}$▲
Caltech07	$0.5395_{.04}$	$0.3987_{.00}$▲	$0.3371_{.00}$▲	$0.3510_{.03}$▲	$0.3991_{.00}$▲	$0.3446_{.01}$▲	$0.4993_{.00}$▲	$0.5388_{.00}$	$0.3002_{.00}$▲
Reuters-1500	$0.0764_{.01}$	$-0.0053_{.00}$▲	$0.0000_{.00}$▲	$0.0580_{.08}$	$0.0472_{.11}$	$0.0790_{.00}$	$-0.0004_{.00}$▲	$0.0000_{.00}$▲	$0.0060_{.00}$▲
Reuters-18758	$0.1107_{.01}$	$0.0000_{.00}$▲	$0.0804_{.00}$▲	$0.0885_{.09}$	$0.1078_{.01}$	$0.0000_{.00}$▲	$-0.0031_{.00}$▲	OOM▲	$0.0441_{.00}$▲

Table 4
Normalized Mutual Information of different algorithms. ▲/▽ indicates our proposed algorithm is statistically superior or inferior to other algorithms at $\alpha = 0.05$. "OOM" means the method raises the out-of-memory failure.

	MvWECM	MCGC	ECM	TW-CO-KM	MVASM	CO-FW-MVFCM	AWP	CHOC-MVC	MCHC
Webkb	$0.2994_{.00}$	$0.2923_{.00}$▲	$0.0330_{.01}$▲	$0.2046_{.20}$	$0.2947_{.01}$	$0.0985_{.12}$▲	$0.2714_{.00}$▲	$0.2555_{.00}$▲	$0.0000_{.00}$▲
3sources	$0.3567_{.03}$	$0.1418_{.00}$▲	$0.5148_{.04}$▽	$0.2532_{.19}$	$0.3959_{.07}$	$0.2572_{.03}$▲	$0.4157_{.00}$▽	$0.2067_{.00}$▲	$0.5468_{.00}$▽
Prok	$0.3313_{.01}$	$0.0147_{.00}$▲	$0.0723_{.07}$▲	$0.3128_{.02}$▲	$0.2422_{.00}$▲	$0.2109_{.04}$▲	$0.2285_{.00}$▲	$0.1885_{.00}$▲	$0.2923_{.00}$▲
IS	$0.1812_{.01}$	$0.0255_{.00}$▲	$0.1330_{.00}$▲	$0.1730_{.02}$	$0.1776_{.00}$	$0.0481_{.03}$▲	$0.1746_{.00}$▲	$0.1004_{.00}$▲	$0.0398_{.00}$▲
Caltech07	$0.3948_{.02}$	$0.4993_{.00}$▽	$0.3534_{.00}$▲	$0.4912_{.03}$▽	$0.4355_{.01}$▽	$0.3774_{.03}$	$0.5967_{.00}$▽	$0.7614_{.00}$▽	$0.4654_{.00}$▽
Reuters-1500	$0.0775_{.01}$	$0.1012_{.00}$▽	$0.0000_{.00}$▲	$0.1060_{.07}$	$0.0656_{.11}$	$0.1267_{.00}$▽	$0.1368_{.00}$▽	$0.0000_{.00}$▲	$0.0785_{.00}$
Reuters-18758	$0.0966_{.01}$	$0.0073_{.00}$▲	$0.0538_{.00}$▲	$0.1193_{.11}$	$0.1431_{.01}$▽	$0.0000_{.00}$▲	$0.0162_{.00}$▲	OOM▲	$0.1490_{.00}$▽

Table 5
Evidential Precision (Precision) of different algorithms. ▲/▽ indicates our proposed algorithm is statistically superior or inferior to other algorithms at $\alpha = 0.05$. "OOM" means the method raises the out-of-memory failure.

	MvWECM	MCGC	ECM	TW-CO-KM	MVASM	CO-FW-MVFCM	AWP	CHOC-MVC	MCHC
Webkb	$0.5800_{.00}$	$0.5734_{.00}$▲	$0.5916_{.04}$	$0.4972_{.01}$	$0.5598_{.04}$	$0.4484_{.07}$▲	$0.6152_{.00}$▽	$0.5876_{.00}$	$0.0000_{.00}$▲
3sources	$0.4009_{.02}$	$0.2166_{.00}$▲	$0.0000_{.00}$▲	$0.2922_{.09}$▲	$0.6446_{.04}$	$0.3589_{.01}$▲	$0.3203_{.00}$▲	$0.7323_{.00}$▽	$0.6030_{.00}$▽
Prok	$0.6551_{.01}$	$0.3923_{.00}$▲	$0.0000_{.00}$▲	$0.5625_{.03}$▲	$0.3924_{.00}$▲	$0.4763_{.05}$▲	$0.4647_{.00}$▲	$0.6011_{.00}$▲	$0.4822_{.00}$▲
IS	$0.2545_{.01}$	$0.1425_{.00}$▲	$0.1979_{.00}$▲	$0.2196_{.02}$▲	$0.2219_{.00}$▲	$0.1670_{.01}$▲	$0.1824_{.00}$▲	$0.1976_{.00}$▲	$0.1425_{.00}$▲
Caltech07	$0.7223_{.03}$	$0.7488_{.00}$▽	$0.0000_{.00}$▲	$0.8453_{.03}$▽	$0.6270_{.20}$	$0.7050_{.04}$	$0.7707_{.00}$▽	$0.4984_{.00}$▲	$0.7420_{.00}$
Reuters-1500	$0.3044_{.06}$	$0.2136_{.00}$▲	$0.0000_{.00}$▲	$0.2448_{.04}$▲	$0.2395_{.05}$▲	$0.2632_{.00}$	$0.2155_{.00}$▲	$0.0000_{.00}$▲	$0.2194_{.00}$▲
Reuters-18758	$0.3054_{.06}$	$0.2140_{.00}$▲	$0.2854_{.06}$	$0.1881_{.17}$	$0.2845_{.01}$	$0.2140_{.00}$▲	$0.2127_{.00}$▲	OOM▲	$0.2466_{.00}$

the top-ranked result is highlighted in red, and the second-ranked result is highlighted in blue. To be more intuitive and be illustrated conveniently, the t-SNE method [40] is utilized to check over the clustering effects of views on 2D. The figures of IS, 3sources and Prok are shown in Fig. 3, Figs. 4 and 5, respectively.

The results by ECM show that traditional evidential clustering algorithm cannot handle multi-view data effectively because every single view provides different data structure (*e.g.* one is sparse matrix, another one is high rank matrix). Compared to other multi-view methods, the proposed MvWECM generally outperforms them. In sum, MvWECM is statistically superior in 67.9%(114/168) and equal in 19.6% (33/168) . In addition, those values of ARI that are close to 0 mean that each partition is essentially equivalent to a random partition basically. When MCGC cannot learn the graph structure and embedding matrix correctly, the performance is approximately equal to random clustering. Thus, negative values indicate that the performance is worse than that of a random distribution. In specific, AWP outperforms on Caltech07 which has 6 views because it considers capacity difference of different views by a unified matrix. For the Reuters-1500 dataset, the abundant high-dimensional features dilute the complementary information between views. In the Reuters-18758 dataset, the large number of samples requires substantial computational resources. Both datasets pose significant challenges for existing multi-view clustering methods. MvWECM attains the highest EP on these datasets, demonstrating efficiency and enhanced accuracy in cluster determination for complex datasets.

In addition, certain approaches may exhibit superior performance when ARI is employed as a metric, while falling short when NMI or EP metrics are considered. For example, the TW-CO-KM algorithm and the MVASM algorithm perform opposite on the IS dataset and the Caltech07 dataset. However, in general, the larger the ARI value, the larger the NMI or EP value. This is exactly what was mentioned earlier, *i.e.*, the three metrics measure the clustering effect in three aspects, thus evaluating the algorithm performance more comprehensively.

As we can see from Figs. 3 and 5, it is evident that the IS and Prok datasets exhibit a high degree of overlap. The notable superiority of MvWECM over other state-of-the-art methods across all evaluation metrics in these two datasets is particularly pronounced. This can be attributed to the incorporation of credal partitions, enhancing our method's ability to address uncertain information, potentially originating from overlapping data points.

4.5. Parameter analysis

First, by fixing $\alpha = 1, \eta = 1$, we conduct analyzes with δ ranges from 2 to 20 and β ranges from 1 to 350. From Fig. 6, it can be inferred that both of them play important roles in ARI. The change of β, δ can make measurement higher or lower gradually or sharply. In fact, choosing a proper value of β is important for tasks. We have

$$\frac{w[l]}{w[m]} = \frac{\exp\left\{\frac{-\Delta[l]-\beta}{\beta}\right\}}{\exp\left\{\frac{-\Delta[m]-\beta}{\beta}\right\}} = \exp\left\{\frac{\Delta[m]-\Delta[l]}{\beta}\right\}, \forall m \neq l.$$

When $\beta \to \infty$,

$$\lim_{\beta\to\infty} \frac{w[l]}{w[m]} = \lim_{\beta\to\infty} \exp\left\{\frac{\Delta[m]-\Delta[l]}{\beta}\right\} = \exp\left\{\lim_{\beta\to\infty} \frac{\Delta[m]-\Delta[l]}{\beta}\right\} = 1.$$

It implies that every view has equal importance which is not suitable in real applications. Besides, The clustering algorithm tends to assign relatively greater weights to the view with the smallest $\Delta[t]$ when $\beta \to 0$, and the final global clustering result learns information from mainly one view, which may ignore some critical information in multi-view clustering tasks.

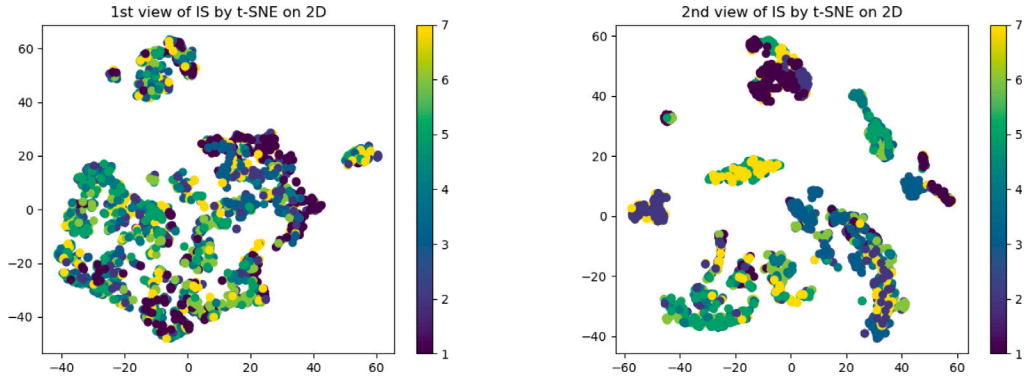


Fig. 3. Clustering results of IS visualized by t-SNE on 2D.

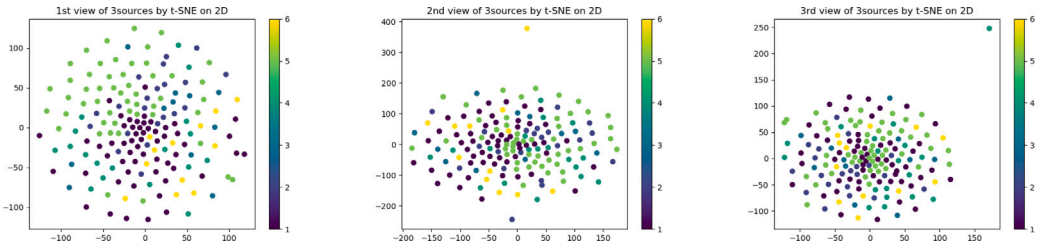


Fig. 4. Clustering results of 3sources visualized by t-SNE on 2D.

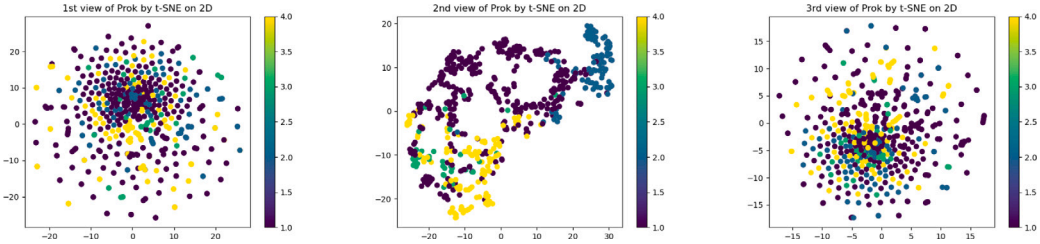
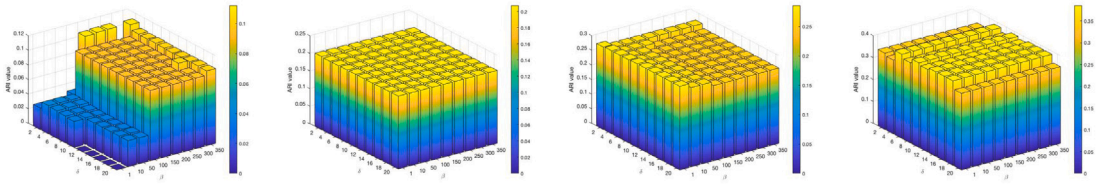
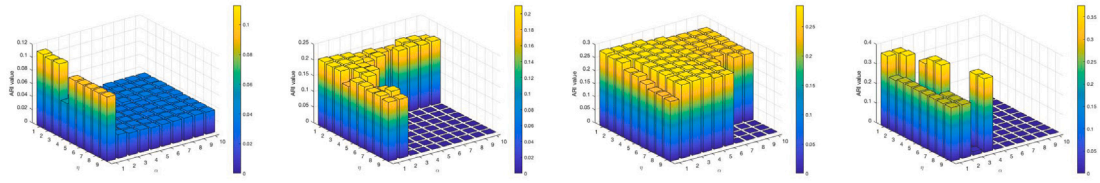


Fig. 5. Clustering results of Prok visualized by t-SNE on 2D.



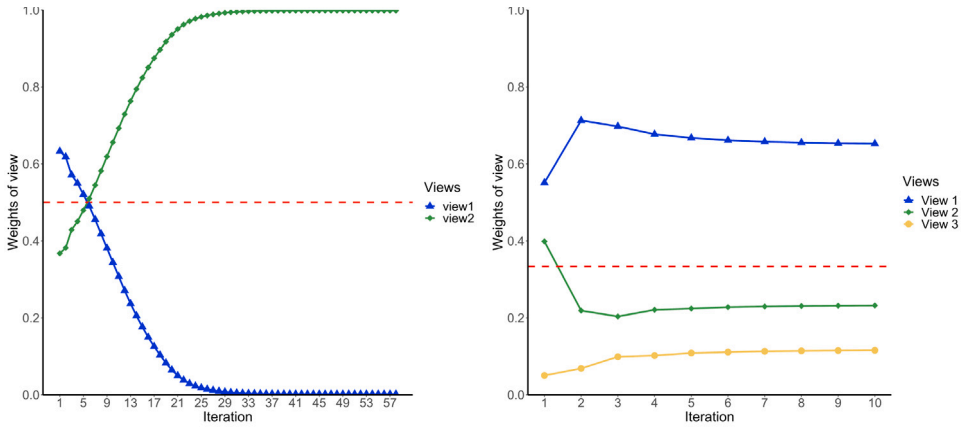
(a) IS (b) Prok (c) 3sources (d) Webkb

Fig. 6. Clustering results of ARI with fixed α and η.



(a) IS (b) Prok (c) 3sources (d) Webkb

Fig. 7. Clustering results of ARI with fixed δ and β.



(a) Views change in each iteration on IS (b) Views change in each iteration on 3sources

Fig. 8. Visualization of weights of view in each iteration on IS and 3sources.

Second, by fixing β, δ which lead to the best performance before, we conduct analyzes and discover that these two parameters may decide whether the MvWECM works (see Fig. 7). We can see that increasing the cardinality will make a larger value of $\Delta[t]$ in Eq. (37), which may lead to accuracy overflows due to exponential function.

In sum, we suggest that parameters settings are $\alpha \in [1, 2]$, $\eta \in [1, 9]$, $\beta \in [1, 350]$ and $\delta \in [2, 20]$.

After analyzing the effect of the parameters on the indicator values, we further analyze the effect of the algorithm on the view weights. In terms of weights of different views, the view which gains better clustering results matches higher weights. The specific change of view weights is indicated by the line graph below. We use the red dashed line to indicate the uninformative view weights, *i.e.*, each view has the same importance.

As can be seen in Fig. 8, view weights change gradually at the beginning. These view weights will be stable when number of iteration overpasses half. It is clear from Fig. 3 that the data in the first view of the IS is more dispersed, *e.g.* the yellow, light green and dark green data points are all more dispersed compared to these three categories in the second view. In contrast, the data in the second view is clearer overall. This is consistent with a greater value of weights of the second view as our algorithm converges. In terms of 3sources, the data of the first view is more divisible, *e.g.*, the yellow ones are mainly concentrated in the lower right of Fig. 4, and the green ones are more concentrated in the upper left of the figure. The data distributions of the second and third views are similar and both have special outliers, thus both views obtain approximate weights. Therefore, the first view gets a higher weight, while other views get lower weights instead.

To experimentally demonstrate the convergence property of MvWECM, we present the convergence curves of MvWECM on the Webkb, Reuters-1500, Reuters-18758, 3sources, and IS datasets in Fig. 9. The values of the objective function generally decrease as the iterations progress. Nevertheless, it is important to note that the algorithm may become stuck in a local minimum. As demonstrated by the curve for the 3sources dataset, the objective function value sharply decreases during the first two iterations, indicating that it is likely trapped in a local optimum. The experimental results for the 3sources dataset further illustrate this phenomenon.

In sum, compared to those SoTA based on crisp or fuzzy partitions, the experimental results illustrate that MvWECM within the framework of belief functions is more reliable in handling overlapping data that lead to ambiguity. This superiority is particularly evident in the results obtained from the Prok and IS datasets. The visualizations also validate the reasonableness of the view weight assignment and fast convergence of our algorithm. Besides, the thorough analysis of the parameters show the rationale behind different terms in the objective function.

5. Conclusion

In this study, we introduce a novel clustering method for multi-view uncertain data, named MvWECM, within the framework of belief functions. The objective function of MvWECM is defined based on the concept of credal partitions. As a result, it can well consider the uncertainty in the cluster structure. Moreover, the collaborative mechanism for effectively taking advantage of the information from different views is designed by considering adaptive view weights. The entropy of view weights is included to ensure their smooth transition. The Lagrange multiplier method combining with Coordinate Descent method is employed to iteratively update the cluster membership and prototypes. Subsequently, the global credal membership is obtained by a weighted sum of the optimal one in each view. The clustering performance of MvWECM is demonstrated by experiments on different real datasets. In the future, we plan to extend the method by incorporating the sparse structure of features. This enhancement will render it more applicable to high-dimensional datasets.

CRediT authorship contribution statement

Kuang Zhou: Writing – review & editing, Supervision, Investigation, Funding acquisition. **Yuchen Zhu:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Formal analysis. **Mei Guo:** Formal analysis. **Ming Jiang:** Investigation.

Declaration of competing interest

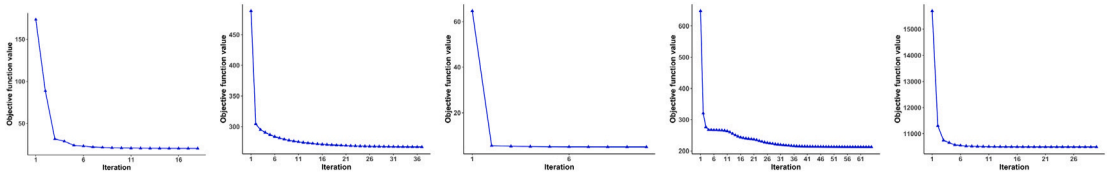
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was sponsored by the National Social Science Fund of China (No. 23BTJ053), the Aeronautical Science Foundation of China (No. 20182053023), the National Natural Science Foundation of China (No. 92371101, 61701409) and the Practice and Innovation Funds for Graduate Students of Northwestern Polytechnical University (No. PF2024115).

Data availability

Data will be made available on request.



(a) Webkb (b) Reuters-1500 (c) 3sources (d) IS (e) Reuters-18758

Fig. 9. Convergence curves of MvWECM.

References

[1] C. Cui, Y. Ren, J. Pu, J. Li, X. Pu, T. Wu, Y. Shi, L. He, A novel approach for effective multi-view clustering with information-theoretic perspective, in: Thirty-Seventh Conference on Neural Information Processing Systems, 2023.

[2] K. Houfar, D. Samai, F. Dornaika, A. Benlamoudi, K. Bensid, A. Taleb-Ahmed, Automatically weighted binary multi-view clustering via deep initialization (AW-BMVC), Pattern Recognit. 137 (2023) 109281.

[3] G. Zhong, C.-M. Pun, Self-taught multi-view spectral clustering, Pattern Recognit. 138 (2023) 109349.

[4] U. Fang, M. Li, J. Li, L. Gao, T. Jia, Y. Zhang, A comprehensive survey on multi-view clustering, IEEE Trans. Knowl. Data Eng. (2023).

[5] F. Nie, L. Tian, X. Li, Multiview clustering via adaptively weighted procrustes, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, pp. 2022–2030.

[6] G.A. Khan, J. Hu, T. Li, B. Diallo, S. Du, Multi-view subspace clustering for learning joint representation via low-rank sparse representation, Appl. Intell. 53 (19) (2023) 22511–22530.

[7] G.A. Khan, J. Hu, T. Li, B. Diallo, H. Wang, Multi-view clustering for multiple manifold learning via concept factorization, Digit. Signal Process. 140 (2023) 104118.

[8] B. Diallo, J. Hu, T. Li, G.A. Khan, X. Liang, H. Wang, Auto-attention mechanism for multi-view deep embedding clustering, Pattern Recognit. 143 (2023) 109764.

[9] Y. Liu, W. Shan, X. Wang, Z. Xiao, L. Geng, F. Zhang, D. Du, Y. Pang, Cross-scale contrastive triplet networks for graph representation learning, Pattern Recognit. 145 (2024) 109907.

[10] A.M. Ikotun, A.E. Ezugwu, L. Abualigah, B. Abuhaija, J. Heming, K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data, Inform. Sci. 622 (2023) 178–210.

[11] G.-Y. Zhang, C.-D. Wang, D. Huang, W.-S. Zheng, Y.-R. Zhou, TW-Co-k-means: Two-level weighted collaborative k-means for multi-view clustering, Knowl.-Based Syst. 150 (2018) 127–138.

[12] M.-S. Yang, K.P. Sinaga, Collaborative feature-weighted multi-view fuzzy c-means clustering, Pattern Recognit. 119 (2021) 108064.

[13] J.C. Bezdek, R. Ehrlich, W. Full, FCM: The fuzzy c-means clustering algorithm, Comput. Geosci. 10 (2–3) (1984) 191–203.

[14] J. Han, J. Xu, F. Nie, X. Li, Multi-view K-means clustering with adaptive sparse memberships and weight allocation, IEEE Trans. Knowl. Data Eng. 34 (2) (2022) 816–827.

[15] G.A. Khan, J. Hu, T. Li, B. Diallo, Y. Zhao, Multi-view low rank sparse representation method for three-way clustering, Int. J. Mach. Learn. Cybern. 13 (2022) 233–253.

[16] A.P. Dempster, Upper and lower probabilities induced by a multivalued mapping, in: Classic Works of the Dempster-Shafer Theory of Belief Functions, Springer, 2008, pp. 57–72.

[17] G. Shafer, A Mathematical Theory of Evidence, vol. 42, Princeton University Press, 1976.

[18] T. Denœux, M.-H. Masson, EVCLUS: evidential clustering of proximity data, IEEE Trans. Syst. Man Cybern. B 34 (1) (2004) 95–109.

[19] M.-H. Masson, T. Denoeux, ECM: An evidential version of the fuzzy c-means algorithm, Pattern Recognit. 41 (4) (2008) 1384–1397.

[20] V. Antoine, B. Quost, M.-H. Masson, T. Denoeux, CECM: Constrained evidential C-means algorithm, Comput. Statist. Data Anal. 56 (4) (2012) 894–914.

[21] Z.-g. Liu, Q. Pan, J. Dezert, G. Mercier, Credal c-means clustering method based on belief functions, Knowl.-based Syst. 74 (2015) 119–132.

[22] C. Gong, Z.-g. Su, P.-h. Wang, Q. Wang, An evidential clustering algorithm by finding belief-peaks and disjoint neighborhoods, Pattern Recognit. 113 (2021) 107751.

[23] L. Jiao, H. Yang, F. Wang, Z.-g. Liu, Q. Pan, DTEC: Decision tree-based evidential clustering for interpretable partition of uncertain data, Pattern Recognit. 144 (2023) 109846.

[24] K. Zhou, M. Guo, A. Martin, Evidential prototype-based clustering based on transfer learning, Internat. J. Approx. Reason. 151 (2022) 322–343.

[25] K. Zhou, M. Guo, M. Jiang, Evidential weighted multi-view clustering, in: Belief Functions: Theory and Applications: 6th International Conference, BELIEF 2021, Shanghai, China, October 15–19, 2021, Proceedings 6, Springer, 2021, pp. 22–32.

[26] P. Smets, R. Kennes, The transferable belief model, in: Classic Works of the Dempster-Shafer Theory of Belief Functions, 2008, pp. 693–736.

[27] P. Smets, Decision making in the TBM: the necessity of the pignistic transformation, Int. J. Approx. Reason. 38 (2) (2005) 133–147.

[28] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Springer Science & Business Media, 2013.

[29] K. Zhan, F. Nie, J. Wang, Y. Yang, Multiview consensus graph clustering, IEEE Trans. Image Process. 28 (3) (2019) 1261–1270.

[30] J. Yang, C.-T. Lin, Multi-view adjacency-constrained hierarchical clustering, IEEE Trans. Emerg. Top. Comput. Intell. 7 (4) (2023) 1126–1138.

[31] X. You, H. Li, J. You, Z. Ren, Consider high-order consistency for multi-view clustering, Neural Comput. Appl. 36 (2) (2024) 717–729.

[32] F. Nie, J. Xue, D. Wu, R. Wang, H. Li, X. Li, Coordinate descent method for k-means, IEEE Trans. Pattern Anal. Mach. Intell. 44 (5) (2021) 2371–2385.

[33] Z.-g. Su, T. Denoeux, BPEC: Belief-peaks evidential clustering, IEEE Trans. Fuzzy Syst. 27 (1) (2018) 111–123.

[34] Q. Lu, L. Getoor, Link-based classification, in: International Conference on Machine Learning, Washington, DC, USA, 2003.

[35] M. Brbić, M. Piškorec, V. Vidulin, A. Kriško, T. Šmuc, F. Supek, The landscape of microbial phenotypic traits and associated genes, Nucleic Acids Res. (2016) gkw964.

[36] Y. Li, F. Nie, H. Huang, J. Huang, Large-scale multi-view spectral clustering via bipartite graph, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 29, 2015.

[37] Z. Huang, J.T. Zhou, X. Peng, C. Zhang, H. Zhu, J. Lv, Multi-view spectral clustering network, in: IJCAI, Vol. 2, 2019, p. 4.

[38] V. Robert, Y. Vasseur, V. Brault, Comparing high-dimensional partitions with the co-clustering adjusted rand index, J. Classification 38 (2021) 158–186.

[39] A. Strehl, J. Ghosh, Cluster ensembles—a knowledge reuse framework for combining multiple partitions, J. Mach. Learn. Res. 3 (Dec) (2002) 583–617.

[40] L. Van der Maaten, G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (11) (2008).

Kuang Zhou received the B.E. degree from Chang'an University, Xi'an, China, in 2010, the M.S. degree from Northwestern Polytechnical University, Xi'an, China, in 2013, and the Ph.D. degree from the University of Rennes 1, France, in 2016. He is currently an associate professor at the School of Mathematics and Statistics, Northwestern Polytechnical University, Xi'an, China. His research interests include uncertain statistical mining and uncertain data analysis.

Yuchen Zhu received the B.E. degree from Northwest University, Xi'an, China, in 2023. He is currently working toward the master's degree with the School of Mathematics and Statistics, Northwestern Polytechnical University, Xi'an, China. His research interests include data clustering and multi-view data analysis.

Mei Guo received the B.E. degree from Shanxi University, Taiyuan, China, in 2020 and the M.S. degree from Northwestern Polytechnical University, Xi'an, China, in 2023. Her research interests include transfer clustering and multi-view data clustering.

Ming Jiang received the B.E. degree and the M.S. degree from the School of Mathematics and Statistics, Northwest Polytechnical University, Xi'an, China, in 2021 and 2024 respectively. His research interests include transfer learning and domain adaptation.