Pattern Recognition 119 (2021) 108064

Contents lists available at ScienceDirect

Pattern Recognition

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

Collaborative feature-weighted multi-view fuzzy c-means clustering

Miin-Shen Yang[a,*], Kristina P. Sinaga[a,b]

[a] Department of Applied Mathematics, Chung Yuan Christian University, Chung-Li, Taoyuan 32023, Taiwan
[b] Information Systems Management Department, BINUS Graduate Program – Master of Information Systems Management, Bina Nusantara University, Jakarta, Indonesia

ARTICLE INFO

Article history:
Received 8 August 2020
Revised 12 January 2021
Accepted 19 May 2021
Available online 28 May 2021

Keywords:
Clustering
Fuzzy c-means (FCM)
Multi-view FCM (MVFCM)
Collaborative learning
Feature weights
Collaborative feature-weighted MVFCM (Co-FW-MVFCM)
Feature reduction

ABSTRACT

Fuzzy c-means (FCM) clustering had been extended for handling multi-view data with collaborative idea. However, these collaborative multi-view FCM treats multi-view data under equal importance of feature components. In general, different features should take different weights for clustering real multi-view data. In this paper, we propose a novel multi-view FCM (MVFCM) clustering algorithm with view and feature weights based on collaborative learning, called collaborative feature-weighted MVFCM (Co-FW-MVFCM). The Co-FW-MVFCM contains a two-step schema that includes a local step and a collaborative step. The local step is a single-view partition process to produce local partition clustering in each view, and the collaborative step is sharing information of their memberships between different views. These two steps are then continuing by an aggregation way to get a global result after collaboration. Furthermore, the embedded feature-weighted procedure in Co-FW-MVFCM can give feature reduction to exclude redundant/irrelevant feature components during clustering processes. Experiments with several data sets demonstrate that the proposed Co-FW-MVFCM algorithm can completely identify irrelevant feature components in each view and that, additionally, it can improve the performance of the algorithm. Comparisons of Co-FW-MVFCM with some existing MVFCM algorithms are made and also demonstrated the effectiveness and usefulness of the proposed Co-FW-MVFCM clustering algorithm.

© 2021 Elsevier Ltd. All rights reserved.

1. Introduction

Clustering is a technique for grouping a set of objects based on their similarities between objects. It is an unsupervised learning in pattern recognition and machine learning. In clustering, fuzzy c-means (FCM) introduced by Dunn [1] and developed by Bezdek [2] is a popularly used clustering procedure in the literature, where FCM had been widely extended and applied in various real-world problems, such as pattern recognition, image segmentation, medical diagnostic, economics, cell formation, gene expression, and data mining [3–8]. However, FCM and its extensions are mostly designed for clustering a single-view real world scenario that generally cannot handle multi-view data. To tackle this issue, many researchers had extended the single-view FCM scenario into multi-view FCM scenarios, and then extended FCM to multi-view FCM (MVFCM) [9–12].

Multi-view learning can be divided into supervised, semi-supervised, and unsupervised learning approaches [13]. Sun et al. [14] proposed a supervised learning method, called multiview uncorrelated linear discriminant analysis (MULDA) in which they embedded the generalized discriminant analysis and kernel canonical correlation analysis to solve a nonlinear multi-view problem. Chen et al. [15] gave a semi-supervised learning method, called semi-supervised generalized correlation analysis (**S^{2}GCA**). The S^{2}GCA can perform semi-supervised learning on two or more views data simultaneously, and thus it can capture the latent knowledge in data more sufficiently. Xin et al. [16] proposed a semi-supervised learning model which incorporates deep feature learning and pseudo label estimation. In general, multi-view clustering is an unsupervised learning approach. Brbić and Kopriva [17] constructed an affinity matrix shared among all views to multi-view subspace clustering based on spectral clustering, namely kernel multi-view low-rank sparse subspace clustering (KMLRSC). The KMLRSC can solve nonlinear problems by reproducing kernel Hilbert space. Zhang et al. [18] introduced the entropy weight approach into single k-means clustering to handle multi-view data, called TW-Co-k-means.

Collaborative learning is one type of unsupervised learning that has been mostly used in FCM to analyze multi-view data. An originally collaborative idea used in FCM for (single-view) data, called Co-FC, was first introduced by Pedrycz et al. [19]. The ultimate goal of collaborative learning is to first identify structures in each of local data sets and then they are merged by exchanging information between their local partition matrices [20]. Some strategies

* Corresponding author.
E-mail address: msyang@math.cycu.edu.tw (M.-S. Yang).

https://doi.org/10.1016/j.patcog.2021.108064
0031-3203/© 2021 Elsevier Ltd. All rights reserved.

can be applied to collaborative learning, such as combining concatenation and centralized strategies. The main idea of concatenation and centralized strategies in FCM for multi-view procedure was proposed by Cleuziou et al. [9], called a centralized method for multiple-view clustering (Co-FKM). Co-FKM can be used to improve the performance of a single-view scenario to a multi-view scenario. However, Co-FKM treated the importance of each view equally. Since the importance of each view in multi-view data is necessary to be identified, Jiang et al. [10] used collaborative learning and automatically recognize different importance of each view in their procedure, called weighted view collaborative FCM (WV-Co-FCM). Their experimental results showed that the view weighting strategy effectively improved the performance of MVFCM clustering. However, WV-Co-FCM achieves promising results by combining multiple parameters selection. In this sense, WV-Co-FCM is unstable and complex in terms of parameter setting. Furthermore, Zeng et al. [11] proposed the collaborative multi view clustering algorithm based on kernel learning process. Zeng et al. [11] called their procedure as a collaborative multi-kernel fuzzy clustering.

Wang and Chen [12] introduced the multi-view minimax-FCM clustering algorithm based on minimax optimization with a general consensus manner which directly concatenates the memberships of views during clustering processing and ignores the connection between views during clustering processing. Minimax-FCM is fast with a time-efficient strategy because it directly implements the consensus membership into their objective function. However, their initialization started from randomly selected cluster centers in each view and then minimized the maximum intra-clustering error in each view. Furthermore, the exponent parameter to control the distribution of view weights in the Minimax-FCM algorithm needs a priory set, and initializations started from randomly selected cluster centers in each view, and different values of parameters always affect clustering results. On the other hand, Liu et al. [21] developed a multi-view clustering based on a joint nonnegative matrix factorization (NMF) technique, called MultiNMF. The NMF technique in MultiNMF was introduced to normalize the coefficient matrix from each view into a common, and a consistent matrix was considered as a potential representation of the original data. In general, when we deal multi-view data sets, each view in multi-view data usually involves large numbers of features. Different feature components in different views have different numbers of dimensions. Thus, in learning process, it is desirable to consider different importance of feature components in each view. Until now, few researches consider different feature weights in MVFCM clustering. Generally, if there exist irrelevant features during clustering processes, the clustering algorithm must take more computational time and even yields incorrect clustering results, especially for multi-view data. Thus, a feature-weighted schema for MVFCM becomes essential.

Recently, Yang and Nataliani [22] proposed a feature-weighted schema with a feature-reduction behavior for (single-view) FCM. However, the feature-weighted procedure in Yang and Nataliani [22] cannot be used as a MVFCM clustering. In this paper, we propose a novel feature-weighted mechanism for MVFCM based on a collaborative learning that can help to identify different importance of features in each view and then automatically discard these redundant feature components with higher clustering performance. The collaborative learning related to all views is allowed to collaborate and share information between their local partition/membership to reach an agreement on the partitioning of multi-view data sets. The rest of the paper is organized as follows. In Section 2, some related works, such as Co-FC, Co-FKM, WV-Co-FCM, Minimax-FCM and MultiNMF, are reviewed and discussed. In Section 3, a novel MVFCM clustering algorithm with feature weights based on collaborative learning, called collaborative FW-MVFCM (Co-FW-MVFCM), is proposed. We further analyze these parameter estimations. We also set up the schema for feature-reduction behaviors. In Section 4, some numerical and real data sets are used to compare the proposed Co-FW-MVFCM clustering algorithm with some existing MVFCM algorithms. Experimental results actually demonstrate the effectiveness and usefulness of the Co-FW-MVFCM algorithm for clustering multi-view data. Finally, conclusions are stated in Section 5.

2. Related works

In this section, we review some related works to FCM and multi-view FCM (MVFCM) clustering algorithms. Let $\mathbf{X} = \{x_1, \ldots, x_n\}$ be a data set in a D-dimensional Euclidean space $\mathbb{R}^d$ and $A = \{a_1, \ldots, a_c\}$ be the set of the c cluster centers with its Euclidean norm denoted by $d_{ik} = \|x_i - a_k\|$. Let $U = [\mu_{ik}]_{n\times c}$be the fuzzy c-membership matrix where μ_{ik} with $\mu_{ik} \in [0, 1]$ and $\sum_{k=1}^{c} \mu_{ik} = 1 \forall i$ is the degree of membership of the data point i in the cluster centers k, and $m>1$ is the fuzziness index. Thus, the FCM objective function was formulated as follows:

$$J_{FCM}(U, A) = \sum_{i=1}^{n}\sum_{k=1}^{c} \mu_{ik}^{m}\|x_i - a_k\|^2 \tag{1}$$

The FCM algorithm is iterated through necessary conditions for minimizing the FCM objective function $J_{FCM}(U, A)$ with updating equations for cluster centers and memberships as $a_{kj} = \sum_{i=1}^{n} \mu_{ik}^{m} x_{ij} / \sum_{i=1}^{n} \mu_{ik}^{m}$ and $\mu_{ik} = (\sum_{k'=1}^{c} (d_{ik}/d_{ik'})^{\frac{2}{m-1}})^{-1}$ with $d_{ik} = \|x_i - a_k\|$.

A collaborative fuzzy clustering (Co-FC) algorithm was first proposed by Pedrycz [19] in which several subsets of patterns can be processed together to find a common structure to all of them. To reveal this structure, different subsets of the initial data are processed independently based on the standard FCM clustering algorithm. The fuzzy c-membership matrix will be used to quantify how far the collaboration affects the final clustering results. In Co-FC, Pedrycz [19] formulated the objective function as follows:

$$J_{CoFC}(h) = J_{FCM}^{h} + \sum_{h'=1, h'\neq h}^{s} \alpha_{hh'} \sum_{i=1}^{n}\sum_{k=1}^{c} \left(\mu_{ik}^{h} - \mu_{ik}^{h'}\right)^2 \sum_{j=1}^{d_h} \left(x_{ij}^{h} - a_{kj}^{h}\right)^2 \tag{2}$$

where $J_{FCM}^{h} = \sum_{i=1}^{n}\sum_{k=1}^{c} (\mu_{ik}^{h})^m \sum_{j=1}^{d_h} (x_{ij}^{h} - a_{kj}^{h})^2$ is the FCM objective function for the hth database, and μ_{ik}^{h} denotes the fuzzy membership degree of the data point x_i belonging to the cluster k of the hth database, and $\alpha_{hh'}$ is a disagreement values between the hth database and the h'th database. In Co-FC, Pedrycz [19] set the value of $\alpha_{hh} = 0$, and μ_{ik}^{h} should satisfy $\sum_{h=1}^{s} \mu_{ik}^{h} = 1$,and $\mu_{ik}^{h} \in [0, 1]$. As it can be seen, the Co-FC objective function (2) is defined by two terms. The first term is the standard FCM. The second term is a new penalty term which is a collaborative learning process among different databases that was implemented by exchanging information between local membership matrices.

Cleuziou et al. [9] seems to be a first try to extend FCM for multi-view data in the literature. Since multi-view data have multiple views in a data set, Cleuziou et al. [9] analyzed different strategies for extracting more information between these different views, and then proposed the multi-view FCM, called Co-FKM. The Co-FKM algorithm proposed by Cleuziou et al. [9] considered the collaborative idea of Pedrycz [19] with a concatenation strategy. The collaboration is by penalizing the disagreement between any pairs of views, and the concatenation strategy consists of membership matrices for all views. The goal of Co-FKM is to search the clustering patterns that perform a consensus between the patterns from different views. The Co-FKM objective function [9] was formulated as follows:

$$J_{Co-FKM} = \sum_{h=1}^{s} J_{FCM}^{h} + \eta \frac{1}{s-1} \sum_{h'=1,h' \neq h}^{s} \sum_{i=1}^{n} \sum_{k=1}^{s} \left\{ \left(\mu_{ik}^{h'}\right)^{m} - \left(\mu_{ik}^{h}\right)^{m} \right\} \sum_{j=1}^{d_h} \left(x_{ij}^{h} - a_{kj}^{h}\right)^{2} \tag{3}$$

where$J_{FCM}^{h} = \sum_{i=1}^{n}\sum_{k=1}^{c} (\mu_{ik}^{h})^{m} (d_{ik}^{h})^{2}$ is the FCM objective function for the hth view,$d_{ik}^{h} = \|x_i^h - a_k^h\| = \sqrt{\sum_{j=1}^{d_h} (x_{ij}^{h} - a_{kj}^{h})^{2}}$ is the Euclidean norm between the ith multi-view data point x_i and the kth multi-view cluster center a_k in the hth view with $x_i = \{x_i^h\}_{h=1}^{s}$, $x_i^h \in \mathbb{R}^{d_h}$, and $x_i^h = \{x_{ij}^h\}_{j=1}^{d_h}$, and$\frac{1}{s-1} \sum_{h=1,h \neq h'}^{s} \sum_{i=1}^{n}\sum_{k=1}^{s} \{(\mu_{ik}^{h'})^{m} - (\mu_{ik}^{h})^{m}\}(d_{ik}^{h})^{2}$ is the additional collaborative penalty term, andη is a parameter which allows to control the penalty related to the disagreement. If we further simplify the objective function (3), it becomes the following form $J_{Co-FKM} = \sum_{h=1}^{s}\sum_{i=1}^{n}\sum_{k=1}^{c} \tilde{\mu}_{ik,\eta}^{h} (d_{ik}^{h})^{2}$, where $\tilde{\mu}_{ik,\eta}^{h} = (1-\eta)(\mu_{ik}^{h})^{m} + \frac{\eta}{s-1} \sum_{h'=1,h'\neq h}^{s} (\mu_{ik}^{h'})^{m}$ is used to denote the weighted mean of the usual fuzzy memberships $(\mu_{ik}^{h})^{m}$ obtained from each view. In order to get a global clustering $\tilde{\mu}_{ik}$, the fuzzy memberships μ_{ik}^{h} of each view need to be merged by calculating the geometric mean of memberships of all views with $\tilde{\mu}_{ik} = \sqrt[s]{\prod_{h=1}^{s} \mu_{ik}^{h}}$.

In Co-FKM, Cleuziou et al. [9] considered equal weights for all views in which Jiang et al. [10] demonstrated it may degrade clustering performance. To identify different importance for each view, Jiang et al. [10] proposed a weighted view collaborative fuzzy c-means (WV-Co-FCM). They proposed the WV-Co-FCM based on a generalized FCM objective function of Zhu et al. [23] with several types of collaborative learning rules under view weights with the following WV-Co-FCM objective function:

$$J_{WV-Co-FCM} = \sum_{h=1}^{s} \upsilon_h \left(J_{FCM}^{h} + \Delta_h\right) + \beta \sum_{h=1}^{c} \upsilon_h \ln \upsilon_h \tag{4}$$

where$J_{FCM}^{h} = \sum_{i=1}^{n}\sum_{k=1}^{c} (\mu_{ik}^{h})^{m} (d_{ik}^{h})^{2}$ with$d_{ik}^{h} = \sqrt{\sum_{j=1}^{d_h} (x_{ij}^{h} - a_{kj}^{h})^{2}}$,$\Delta_h = \sum_{i=1}^{n} \alpha_{ik}^{h} \sum_{k=1}^{c} \mu_{ik}^{h} (1 - (\mu_{ik}^{h})^{m-1}) - \sum_{i=1}^{n} \delta_{ik}^{h} \sum_{k=1}^{c} \mu_{ik}^{h} (1 - (\mu_{ik}^{h})^{m-1})$, υ_h is a view weight for the hth view with $\sum_{h=1}^{s} \upsilon_h = 1$, $\beta > 0$ is a parameter used to control the view weight,$\alpha_{ik}^{h} = \eta (d_{ik}^{h})^{2}$, δ_{ik}^{h} has four cases including $\delta_{ik}^{h} = \eta \frac{1}{s-1} \sum_{h'=1,h'\neq h}^{s} (d_{ik}^{h'})^{2}$, $\delta_{ik}^{h} = \frac{\eta}{s} \sum_{h=1}^{s} (d_{ik}^{h})^{2}$, $\delta_{ik}^{h} = \eta \min_{h' \neq h} \{(d_{ik}^{h'})^{2}\}$, and $\delta_{ik}^{h} = \eta \sqrt[s-1]{\prod_{h' \neq h} (d_{ik}^{h'})^{2}}$, $0 < \eta < 1$ is a parameter used to control the penalty related to the disagreement, $m > 1$ is the fuzziness degree.

Different from Co-FC, Co-FKM and WV-Co-FCM, the Minimax-FCM algorithm proposed by Wang and Chen [12] did not use an additional step to compute the disagreement between views. Minimax-FCM does not require a different membership for each view, but instead uses the same membership for all views. In other words, Wang and Chen [12] directly concatenated the memberships of views during clustering processing. Their initializations are started from randomly selected cluster centers in each view and then minimized the maximum intra-clustering error in each view. Moreover, the minimax-FCM has an exponent parameter to handle the distribution of view weights. The rules of minimax optimization in FCM are used to produce the minimum disagreement of different view weights. Wang and Chen [12] used a simple MVFCM objective function as follows:

$$J_{\text{minimax-FCM}}\left(V, U^{*}, A^{h}\right) = \sum_{h=1}^{s} (\upsilon_h)^{\beta} \sum_{i=1}^{n}\sum_{k=1}^{c} \left(\mu_{ik}^{*}\right)^{m} \sum_{j=1}^{d_h} \left(x_{ij}^{h} - a_{kj}^{h}\right)^{2} \tag{5}$$

where $U^{*} = [\mu_{ik}^{*}]_{n \times c}$ with $\mu_{ik}^{*} \in [0,1]$ and $\sum_{k=1}^{c} \mu_{ik}^{*} = 1 \forall i$, and μ_{ik}^{*} is the consensus membership of the data point i in the cluster centers k shared across different views, $A^h = \{a_1^h, \cdots, a_c^h\}$ is the set of the c cluster centers in the hth view with $a_k^h = \{a_{kj}^h\}$, and $V = [\upsilon_h]_{1\times s}$, where υ_h is the view weight vector v, and β is the exponent parameter to control the distribution of view weights, and $m > 1$ is the fuzziness degree. In the minimax-FCM, Wang and Chen [12] used the min-max optimization with $\min_{U^{*},\{A^h\}_{h=1}^{s}} \max_{\{\upsilon_h\}_{h=1}^{s}} \sum_{h=1}^{s} (\upsilon_h)^{\beta} \sum_{i=1}^{n}\sum_{k=1}^{c} (\mu_{ik}^{*})^{m} \sum_{j=1}^{d_h} (x_{ij}^{h} - a_{kj}^{h})^{2}$. In the minimax-FCM, the exponent parameter of view weights and the level of fuzziness need to be identified before clustering. Wang and Chen [12] suggested the fuzziness index m ranges from 1.1 to 2 with step 0.1, while the exponent parameter to handle the distribution of view weights ranges from 0.1 to 0.9 with step 0.1.

On the other hand, Liu et al. [21] developed a multi-view clustering based on a joint non-negative matrix factorization (NMF) technique, called MultiNMF. NMF technique in MultiNMF by Liu et al. [21] is used to solve multiple incomplete views by normalizing the coefficient matrix from each view into a common, and a consistent matrix was considered as a potential representation of the original data. Liu et al. [21] proposed the following objective function:

$$J_{MultiNMF} = \sum_{h=1}^{s} \left\| X^{h} - U^{h} \left(A^{h}\right)^{T} \right\|_{F}^{2} + \sum_{h=1}^{s} \lambda_h D\left(A^{h}, A^{*}\right) \tag{6}$$

in which $X^h = [X^1; X^2; \cdots; X^h] \in \mathbb{R}^{n \times d_h}$ denotes the nonnegative data matrix in the hth view, $U^h = [U^1; U^2; \cdots; U^h] \in \mathbb{R}^{n \times c}$ is the membership matrix of the hth view with $\sum_{i=1}^{n} \mu_{ik} = 1, \forall k$, and$A^h = [A^1; A^2; \cdots; A^h] \in \mathbb{R}^{c \times d_h}$ is the cluster indicator matrix of the hth view, where n is the number of data points, c is the number of clusters, d_h is the number of features in the hth view, λ_h is a parameter to control the relative weight among different views and between A^h and A^* that need to be defined by user, and $D(A^h, A^*) = \|A^h - A^*\|_F^2$ is the Frobenius norm and used as a measure of disagreement between the cluster indicator matrixA^h and the cluster consensus matrix A^* of the hth view.

3. The proposed collaborative feature-weighted MVFCM clustering algorithm

As mentioned before, there is less consideration of feature weights in multi-view FCM (MVFCM) in the literature. In this section, we propose a novel feature-weighted MVFCM (FW-MVFCM) clustering algorithm. Let $\mathbf{X} = \{x_1, \ldots, x_n\}$ be a multi-view data set in a D-dimensional Euclidean space $\mathbb{R}^d$with $x_i = \{x_i^h\}_{h=1}^s$, $x_i^h \in \mathbb{R}^{d_h}$, and $x_i^h = \{x_{ij}^h\}_{j=1}^{d_h}$ with $\sum_{h=1}^s d_h = d$. Let $U^h = [\mu_{ik}^h]_{n\times c}$, $h = 1, \cdots, s$ with $\mu_{ik}^h \in [0, 1]$ and $\sum_{k=1}^c \mu_{ik}^h = 1 \forall i, h$, where μ_{ik}^h is the degree of membership of the data point i belonging to the cluster k of the hth view. Let $A = \{a_1, \ldots, a_c\}$ be the c cluster centers with $a_k = \{a_{kj}^h\}$, $j = 1, \cdots, d_h$. Let $W = [w_j]_{1\times d}$, where $w_j = \{w_j^h\}_{h=1}^s$is thej jth feature weight in the hth view. Let $V = [v_h]_{1\times s}$, where v_h is a weight for the hth view. To address the FW-MVFCM, we propose the following simple objective function:

$$J_{FW-MVFCM} = \sum_{h=1}^{s} (v_h)^\beta \sum_{i=1}^{n}\sum_{k=1}^{c} (\mu_{ik}^h)^m \sum_{j=1}^{d_h} \left(w_j^h\right)^2 \left(x_{ij}^h - a_{kj}^h\right)^2 \tag{7}$$

subject to$\sum_{k=1}^c \mu_{ik}^h = 1, \mu_{ik}^h \in [0, 1]$, $\sum_{h=1}^s v_h = 1, v_h \in [0, 1]$, and $\sum_{j=1}^{d_h} w_j^h = 1, w_j^h \in [0, 1]$. By taking the derivative of Eq. (7) w.r.t. a_{kj}^h and setting them to be zero, the update equation for the cluster centers a_{kj}^h is with$a_{kj}^h = \sum_{i=1}^c (\mu_{ik}^h)^m x_{ij}^h / \sum_{i=1}^c (\mu_{ik}^h)^m$. For the update equations of view weights v_h, feature weights w_j^h, and membership μ_{ik}^h, we need to consider the following Lagrangian function $\tilde{J}_{FW-MVFCM} = J_{FW-MVFCM} - \lambda_1(\sum\limits_{k=1}^{c} \mu_{ik}^h - 1) - \lambda_2(\sum\limits_{h=1}^{s} v_h - 1) - \lambda_3(\sum\limits_{j=1}^{d_h} w_j^h - 1)$. By taking the partial derivative of the Lagrangiang w.r.t v_h and setting them to be zero, we obtain the equation $(\lambda_2)^{\frac{1}{\beta-1}} = (\sum\limits_{h=1}^{s} (\beta \sum\limits_{i=1}^{n}\sum\limits_{k=1}^{c} (\mu_{ik}^h)^m \sum\limits_{j=1}^{d_h} (w_j^h)^2 (x_{ij}^h - a_{kj}^h)^2)^{\frac{-1}{\beta-1}})^{-1}$. Thus, the updating equation for v_h can be obtained with $v_h = (\sum\limits_{h'=1}^{s} (\frac{\sum_{i=1}^{n}\sum_{k=1}^{c} (\mu_{ik}^h)^m (w_j^h)^2 (x_{ij}^h - a_{kj}^h)^2}{\sum_{i=1}^{n}\sum_{k=1}^{c} (\mu_{ik}^{h'})^m (w_j^{h'})^2 (x_{ij}^{h'} - a_{kj}^{h'})^2})^{\frac{1}{\beta-1}})^{-1}$. Similarly, by taking the partial derivative of the Lagrangian w.r.t w_j^h and setting them to be zero, we obtain the updating equation for w_j^h with $w_j^h = (\sum\limits_{j'=1}^{d_h} (\frac{(v_h)^\beta \sum_{i=1}^{n}\sum_{k=1}^{c} (\mu_{ik}^h)^m (x_{ij}^h - a_{kj}^h)^2}{(v_h)^\beta \sum_{i=1}^{n}\sum_{k=1}^{c} (\mu_{ik}^h)^m (x_{ij'}^h - a_{kj'}^h)^2}))^{-1}$. By taking the partial derivative of the Lagrangian w.r.t. μ_{ik}^h and setting them to be zero, we obtain the updating equation for μ_{ik}^h with $\mu_{ik}^h = (\sum\limits_{k'=1}^{c} (\frac{(v_h)^\beta \sum_{j=1}^{d_h} (w_j^h)^2 (x_{ij}^h - a_{kj}^h)^2}{(v_h)^\beta \sum_{j=1}^{d_h} (w_j^h)^2 (x_{ij}^h - a_{k'j}^h)^2}))^{\frac{-1}{m-1}}$. Thus, a simple feature-weighted MVFCM algorithm for clustering multi-view data can be constructed based on the above updating equations fora_{kj}^h, μ_{ik}^h,v_h, and w_j^h.

Although we previously proposed the simplest one of feature-weighted MVFCM clustering for multi-view data. However, the correct classification rate is not enough for most of multi-view data, and so it can be improved. This is because the recognition rate of the feature-weighted MVFCM algorithm based on the objective function (7) can evaluate the error of misclassification contained in the final partitioning, but it provides only partial information about the structural quality of the obtaining partition. That is, it may improve the clustering performance but the resulting partition might be practically useless. In order to get a more objective judgement about the partitioning quality, the exchanging information about the local partition between each view during the learning process should be taken into considerations. Pedrycz [19] provided the collaborative learning with the objective function $J_{Co-FW-MVFCM} = \sum\limits_{h=1}^{s} (v_h)^\beta (\sum\limits_{i=1}^{n}\sum\limits_{k=1}^{c} (\mu_{ik}^h)^m \sum\limits_{j=1}^{d_h} (w_j^h)^2 (x_{ij}^h - a_{kj}^h)^2 + \varphi)$where$\varphi = \alpha \sum\limits_{h'=1, h'\neq h}^{s} \sum\limits_{i=1}^{n}\sum\limits_{k=1}^{c} \{\mu_{ik}^h - \mu_{ik}^{h'}\}^2 (d_{ik,w}^h)^2$. Thus, we propose a collaborative feature-weighted MVFCM (Co-FW-MVFCM) objective function as follows:

$$J_{Co-FW-MVFCM} = \sum_{h=1}^{s} (v_h)^\beta \left(\sum_{i=1}^{n}\sum_{k=1}^{c} (\mu_{ik}^h)^2 (d_{ik,w}^h)^2 + \alpha \sum_{h'=1, h'\neq h}^{s} \sum_{i=1}^{n}\sum_{k=1}^{c} \left\{\mu_{ik}^h - \mu_{ik}^{h'}\right\}^2 \left(d_{ik,w}^h\right)^2 \right) \tag{8}$$

subject to$\sum_{k=1}^c \mu_{ik}^h = 1, \mu_{ik}^h \in [0, 1]$, $\sum_{h=1}^s v_h = 1, v_h \in [0, 1]$, and $\sum_{j=1}^{d_h} w_j^h = 1, w_j^h \in [0, 1]$, where$d_{ik,w}^h = \sqrt{\sum\limits_{j=1}^{d_h} (w_j^h)^2 (x_{ij}^h - a_{kj}^h)^2}$ is the weighted distance between the data point i and the cluster k in the hth view. The first terms of Eq. (8) is a single view partition process to produce a local clustering partition in each view. The second terms of Eq. (8) corresponds to the collaborative learning between views in multi-view data with a collaborative parameter α. As it can be seen, the fuzzier m in the objective function (7) is suppressed with a fixed *m=2* in Co-FW-MVFCM. That is, the proposed Co-FW-MVFCM does not use the parameter of fuzzier m. We suppose the collaborative parameter α will give more effects than the fuzzier m for Co-FW-MVFCM. In this sense, we give a fixed *m=2* and then focus on an estimation of the parameter α. We note that the fixed fuzzier *m=2* in Co-FW-MVFCM is enough to control the extents for sharing among fuzzy clusters by decreasing complexity in determining the fuzzier parameter m. However, the parameter α has the main role in controlling the strengthens of collaboration between views that we need to estimate. In order to merge the partition matrices in each view and obtain the global clustering result, we adopt the summation of each weighted fuzzy partition matrix for each view as:

$$\widehat{U} = \sum_{h=1}^{s} v_h U^h \tag{9}$$

Theorem 1. *The updating equations for necessary conditions to minimize the Co-FW-MVFCM objective function* $J_{Co-FW-MVFCM}$ *of* Eq. (8) *are*

$$\mu_{ik}^{h}=\left(1-\sum_{l=1}^{c}\left(\frac{\varphi_{il}^{h}}{1+\alpha}\right)\right)\left(\sum_{l=1}^{c}\frac{\left(d_{ik,w}^{h}\right)^{2}}{\left(d_{il,w}^{h}\right)^{2}}\right)^{-1}+\frac{\varphi_{ik}^{h}}{1+\alpha} \tag{10}$$

$$v_{h}=\left(\sum_{r=1}^{s}\left(\frac{\sum_{i=1}^{n}\sum_{k=1}^{c}\left(\mu_{ik}^{h}\right)^{2}\left(d_{ik,w}^{h}\right)^{2}+\alpha\sum_{h'=1,h'\neq h}^{s}\sum_{i=1}^{n}\sum_{k=1}^{c}\left\{\mu_{ik}^{h}-\mu_{ik}^{h'}\right\}^{2}\left(d_{ik,w}^{h}\right)^{2}}{\sum_{i=1}^{n}\sum_{k=1}^{c}\left(\mu_{ik}^{h}\right)^{2}\left(d_{ik,w}^{r}\right)^{2}+\alpha\sum_{h'=1,h'\neq r}^{s}\sum_{i=1}^{n}\sum_{k=1}^{c}\left\{\mu_{ik}^{r}-\mu_{ik}^{h'}\right\}^{2}\left(d_{ik,w}^{r}\right)^{2}}\right)^{\frac{1}{\beta-1}}\right)^{-1} \tag{11}$$

$$w_{j}^{h}=\left(\sum_{j'=1}^{d_h}\left(\frac{(v_h)^{\beta}\left(\sum_{i=1}^{n}\sum_{k=1}^{c}\left(\mu_{ik}^{h}\right)^{2}\left(x_{ij}^{h}-a_{kj}^{h}\right)^{2}+\alpha\sum_{h'=1,h'\neq h}^{s}\sum_{i=1}^{n}\sum_{k=1}^{c}\left\{\mu_{ik}^{h}-\mu_{ik}^{h'}\right\}^{2}\left(x_{ij}^{h}-a_{kj}^{h}\right)^{2}\right)}{(v_h)^{\beta}\left(\sum_{i=1}^{n}\sum_{k=1}^{c}\left(\mu_{ik}^{h}\right)^{2}\left(x_{ij'}^{h}-a_{kj'}^{h}\right)^{2}+\alpha\sum_{h'=1,h'\neq h}^{s}\sum_{i=1}^{n}\sum_{k=1}^{c}\left\{\mu_{ik}^{h}-\mu_{ik}^{h'}\right\}^{2}\left(x_{ij'}^{h}-a_{kj'}^{h}\right)^{2}\right)}\right)\right)^{-1} \tag{12}$$

$$a_{kj}^{h}=\frac{(v_h)^{\beta}\sum_{i=1}^{n}\left(\mu_{ik}^{h}\right)^{2}\left(w_{j}^{h}\right)^{2}x_{ij}^{h}+\alpha\sum_{h'=1,h'\neq h}^{s}(v_h)^{\beta}\sum_{i=1}^{n}\left\{\mu_{ik}^{h}-\mu_{ik}^{h'}\right\}^{2}\left(w_{j}^{h}\right)^{2}x_{ij}^{h}}{(v_h)^{\beta}\sum_{i=1}^{n}\left(\mu_{ik}^{h}\right)^{2}\left(w_{j}^{h}\right)^{2}+\alpha\sum_{h'=1,h'\neq h}^{s}(v_h)^{\beta}\sum_{i=1}^{n}\left\{\mu_{ik}^{h}-\mu_{ik}^{h'}\right\}^{2}\left(w_{j}^{h}\right)^{2}} \tag{13}$$

Proof: We prove it by using Langrange multipliers. The Lagrangian of $J_{Co-FW-MVFCM}$ subject to $\sum_{k=1}^{c}\mu_{ik}^{h}=1,\mu_{ik}^{h}\in[0,1]$ is $\tilde{J}=J_{Co-FW-MVFCM}-\lambda_1(\sum_{k=1}^{c}\mu_{ik}^{h}-1)$. By taking the partial derivative of the Lagrangian w.r.t. μ_{ik}^{h} and setting them to be zero, we obtain the equation $\frac{\partial\tilde{J}}{\partial\mu_{ik}^{h}}=(v_h)^{\beta}(2\mu_{ik}^{h}(d_{ik,w}^{h})^{2}+2\alpha\sum_{h'=1,h'\neq h}^{s}(\mu_{ik}^{h}-\mu_{ik}^{h'})(d_{ik,w}^{h})^{2})-\lambda_1=0$. Thus, we have $\mu_{ik}^{h}=\frac{\lambda_1+(v_h)^{\beta}\alpha\sum_{h'=1,h'\neq h}^{s}\mu_{ik}^{h'}(d_{ik,w}^{h})^{2}}{(v_h)^{\beta}(1+\alpha)(d_{ik,w}^{h})^{2}}$. Let $\varphi_{ik}^{h}=\alpha\sum_{h'=1,h'\neq h}^{s}\mu_{ik}^{h'}$. We can rewrite $\mu_{ik}^{h}=\frac{\lambda_1}{(v_h)^{\beta}(1+\alpha)}\frac{1}{(d_{ik,w}^{h})^{2}}+\frac{\varphi_{ik}^{h}}{1+\alpha}$. Since $\sum_{l=1}^{c}\mu_{il}^{h}=1$, we get $\frac{\lambda_1}{(v_h)^{\beta}(1+\alpha)}=1-\sum_{l=1}^{c}(\frac{\varphi_{il}^{h}}{1+\alpha})/\sum_{l=1}^{c}\frac{1}{(d_{il,w}^{h})^{2}}$. Thus, the updating equation for μ_{ik}^{h} can be obtained as $\mu_{ik}^{h}=(1-\sum_{l=1}^{c}(\frac{\varphi_{il}^{h}}{1+\alpha}))(\sum_{l=1}^{c}\frac{(d_{ik,w}^{h})^{2}}{(d_{il,w}^{h})^{2}})^{-1}+\frac{\varphi_{ik}^{h}}{1+\alpha}$. We consider the Lagrangian of $J_{Co-FW-MVFCM}$ subject to $\sum_{h=1}^{s}v_h=1,v_h\in[0,1]$ with $\tilde{J}=J_{Co-FW-MVFCM}-\lambda_2(\sum_{h=1}^{s}v_h-1)$. By taking the partial derivative of the Lagrangian w.r.t v_h and setting them to be zero, we obtain the equation $\frac{\partial\tilde{J}}{\partial v_h}=\beta(v_h)^{\beta-1}(\sum_{i=1}^{n}\sum_{k=1}^{c}(\mu_{ik}^{h})^{2}(d_{ik,w}^{h})^{2}+\alpha\sum_{h'=1,h'\neq h}^{s}\sum_{i=1}^{n}\sum_{k=1}^{c}\{\mu_{ik}^{h}-\mu_{ik}^{h'}\}^{2}(d_{ik,w}^{h})^{2})-\lambda_2=0$. Thus, we have $v_h=(\lambda_2)^{\frac{1}{\beta-1}}(\sum_{i=1}^{n}\sum_{k=1}^{c}(\mu_{ik}^{h})^{2}(d_{ik,w}^{h})^{2}+\alpha\sum_{h'=1,h'\neq h}^{s}\sum_{i=1}^{n}\sum_{k=1}^{c}\{\mu_{ik}^{h}-\mu_{ik}^{h'}\}^{2}(d_{ik,w}^{h})^{2})^{\frac{-1}{\beta-1}}$. Since $\sum_{h=1}^{s}v_h=1,v_h\in[0,1]$, we get $(\lambda_2)^{\frac{1}{\beta-1}}=(\sum_{r=1}^{s}(\sum_{i=1}^{n}\sum_{k=1}^{c}(\mu_{ik}^{r})^{2}(d_{ik,w}^{r})^{2}+\alpha\sum_{h'=1,h'\neq r}^{s}\sum_{i=1}^{n}\sum_{k=1}^{c}\{\mu_{ik}^{r}-\mu_{ik}^{h'}\}^{2}(d_{ik,w}^{r})^{2})^{\frac{-1}{\beta-1}})^{-1}$. Thus, the updating Eq. (11) for v_h is obtained. We finally consider the Lagrangian of $J_{Co-FW-MVFCM}$ subject to $\sum_{j=1}^{d_h}w_{j}^{h}=1,w_{j}^{h}\in[0,1]$ with $\tilde{J}=J_{Co-FW-MVFCM}-\lambda_3(\sum_{j=1}^{d_h}w_{j}^{h}-1)$. By taking the partial derivative of the Lagrangian w.r.t w_{j}^{h} and setting them to be zero, we obtain the equation $\frac{\partial\tilde{J}}{\partial w_{j}^{h}}=(v_h)^{\beta}(2\sum_{i=1}^{n}\sum_{k=1}^{c}(\mu_{ik}^{h})^{2}w_{j}^{h}(x_{ij}^{h}-a_{kj}^{h})^{2}+2\alpha\sum_{h'=1,h'\neq h}^{s}\sum_{i=1}^{n}\sum_{k=1}^{c}\{\mu_{ik}^{h}-\mu_{ik}^{h'}\}^{2}w_{j}^{h}(x_{ij}^{h}-a_{kj}^{h})^{2})-\lambda_3=0$. Thus, we have $w_{j}^{h}=\lambda_3((v_h)^{\beta}(\sum_{i=1}^{n}\sum_{k=1}^{c}(\mu_{ik}^{h})^{2}(x_{ij}^{h}-a_{kj}^{h})^{2}+\alpha\sum_{h'=1,h'\neq h}^{s}\sum_{i=1}^{n}\sum_{k=1}^{c}\{\mu_{ik}^{h}-\mu_{ik}^{h'}\}^{2}(x_{ij}^{h}-a_{kj}^{h})^{2}))^{-1}$, and then we get $\lambda_3=(\sum_{j'=1}^{d_h}((v_h)^{\beta}(\sum_{i=1}^{n}\sum_{k=1}^{c}(\mu_{ik}^{h})^{2}(x_{ij'}^{h}-a_{kj'}^{h})^{2}+\alpha\sum_{h'=1,h'\neq h}^{s}\sum_{i=1}^{n}\sum_{k=1}^{c}\{\mu_{ik}^{h}-\mu_{ik}^{h'}\}^{2}(x_{ij'}^{h}-a_{kj'}^{h})^{2}))^{-1})^{-1}$. Thus, we can obtain the updating Eq. (12) for w_{j}^{h}. By taking the derivative of the Co-FW-MVFCM objective function $J_{Co-FW-MVFCM}$ w.r.t a_{kj}^{h} and setting them to be zero, we obtain the equation $\frac{\partial J}{\partial a_{kj}^{h}}=(v_h)^{\beta}(\{-2\sum_{i=1}^{n}(\mu_{ik}^{h})^{2}(w_{j}^{h})^{2}(x_{ij}^{h}-a_{kj}^{h})\}+\{-2\alpha\sum_{h'=1,h'\neq h}^{s}\sum_{i=1}^{n}\{\mu_{ik}^{h}-\mu_{ik}^{h'}\}^{2}(w_{j}^{h})^{2}(x_{ij}^{h}-a_{kj}^{h})\})=0$, and then $a_{kj}^{h}((v_h)^{\beta}\sum_{i=1}^{n}(\mu_{ik}^{h})^{2}(w_{j}^{h})^{2}+\alpha\sum_{h'=1,h'\neq h}^{s}(v_h)^{\beta}\sum_{i=1}^{n}\{\mu_{ik}^{h}-\mu_{ik}^{h'}\}^{2}(w_{j}^{h})^{2})=(v_h)^{\beta}\sum_{i=1}^{n}(\mu_{ik}^{h})^{2}(w_{j}^{h})^{2}x_{ij}^{h}+\alpha\sum_{h'=1,h'\neq h}^{s}(v_h)^{\beta}\sum_{i=1}^{n}\{\mu_{ik}^{h}-\mu_{ik}^{h'}\}^{2}(w_{j}^{h})^{2}x_{ij}^{h}$. Thus, the update Eq. (13) for the cluster centers a_{kj}^{h} can be obtained. ■

In practical applications, irrelevant features may reduce the performance of clustering algorithms. Thus, a feature-reduction behavior in clustering algorithms should be an important issue. To create a feature-reduction schema, the process for selecting these important features in each view is a crucial step. The importance of features can be identified based on its score. The higher the score is, the more importance/relevance the feature is. In general, data points are represented by feature vectors. Therefore, we first learn the sensitivity of α and β in the effectiveness of the feature-reduction behavior in each view. We then evaluate the quality of two parameters in obtaining the feature reduction based on the performance of the proposed Co-FW-MVFCM algorithm after discarding these irrelevant/redundant features.

However, it is still not clear how many features should be excluded during clustering processes. To cope with this problem, we introduce a threshold to eliminate a sequence of unimportant features in each view. To estimate the threshold, we need to set the value not very high, but not very low. If a threshold is very low (near zero), then the proposed algorithm will add too many features to the feature space. On the other hand, if a threshold is very high (near one), the algorithm will not add a single feature to the feature space. To achieve this goal, we consider the number n of data points and the dimension number d_h in each view to estimate the threshold. The threshold should fit all views, and we set the threshold as

$$W^{h^{(t)}} < 1/\sqrt{nd_h} \tag{14}$$

We note that the feature component should exist during clustering processes if the feature weight is larger than the threshold $1/\sqrt{nd_h}$. That is, the feature component should be reduced and do not continuously used in clustering processes if the feature weight is smaller than the threshold.

The ideal value of the exponent parameterβ is the next essential question in Co-FW-MVFCM because this parameter handles the distribution of view weights with its feature components. There are three core aspects necessary to be considered when we estimate this parameter, such as estimating time cost, must be adjusted to obtain an appropriate feature selection in one view, and fit into collaborative learning. An appropriate value of β is also important for improving clustering performance. We reach an optimal parameter when such of the parameter can identify the view and its feature components. For the given of the actual number of iterations t and the total number of the data view s, we give the calculation expression of the exponent parameter as

$$\beta=t/s \tag{15}$$

where the value of t guarantees the value of β to be changed after each iteration, i.e. a greater number of β will produce a sharp distinction between the important views and unimportant views. Hence, this behavior can automatically identify feature components in each view.

The next key parameter is the collaborative parameter α. The original idea of collaborative learning in Co-FC (Pedrycz [19]) is the coefficient of agreement $\alpha_{hh'}$ to measure the degree of agreement among databases h and h'. It is suppose that bigger of the agreement coefficient $\alpha_{hh'}$ should be better. However, the different values of the parameter $\alpha_{hh'}$ heavily affect final clustering results with a very unstable situation. It is difficult to give a good estimation. We now propose an estimation of α with

$$\alpha = t/n \tag{16}$$

A combination of these actual number of iterations *(t)* and the total number of data points *(n)* are taken into consideration when learning the collaboration between views. The actual number of iterations t during the clustering processes will change the value of α after each iteration. The coefficient parameter α will start from a small value and frequently increase until the convergence. We next demonstrate these behaviors by using a numerical example. The example is also used to explain the selected parameter in handling the distribution of view weights and learning the effect of each parameter in performing final results.

Example 1 In this example, a two-view numerical data set with 2 clusters and 2 feature components is considered. The data points in each view are generated from a 2-component 2-variate Gaussian mixture model (GMM) where their mixing proportions $\alpha_1^{(1)} = \alpha_1^{(2)} = 0.55$ and $\alpha_2^{(1)} = \alpha_2^{(2)} = 0.45$. The means $\mu_k^{(1)}$ for the view 1 are $(\ 15.5 \quad 8\ )$ and $(\ 8.5 \quad 8.5\ )$. The means $\mu_k^{(2)}$ for the view 2 are $(\ 8.5 \quad 8.5\ )$and $(\ 8 \quad 15.5\ )$. The covariance matrices for the two views are $\sum_1^{(1)} = \sum_1^{(2)} = \begin{pmatrix} 2.5 & 0 \\ 0 & 2.5 \end{pmatrix}$ and $\sum_2^{(1)} = \sum_2^{(2)} = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}$. These $x_2^{(1)}$ and $x_3^{(1)}$ are the coordinates for the view 1, and $x_2^{(2)}$ and $x_3^{(2)}$ are the coordinates for the view 2, as shown in Fig. 1(a) and (b), respectively. To study the feature-reduction behavior of the proposed Co-FW-MVFCM clustering algorithm, an additional feature generated from a uniform distribution over the interval [0, 0.2] is added in each view that are coordinated by $x_1^{(1)}$ for the view 1, and coordinated by $x_1^{(2)}$ for the view 2, as shown in Fig. 1(c) and (d), respectively. If we implement the proposed Co-FW-MVFCM algorithm for the multi-view data set with n=800, c=2, s=2 and $d_h = 3, h = 1, \cdots, s$, as shown Fig. 1(c) and (d), we get different weighted proportion histogram of view 1 and view 2 as shown in Fig. 2(a) and (b). Fig. 2(a), (b) presents the feature components of one view in each iteration. As can be seen, the proposed Co-FW-MVFCM algorithm immediately weighted the first features in both views with the smallest values at the first iteration and discarded it at the second iteration. Finally, by examining the important features of one view into Co-FW-MVFCM algorithm, we obtain the highest AR's, as shown in Fig. 3(a). This result also presented the effect of $\alpha$and β on the proposed Co-FW-MVFCM algorithm in terms of combining the dimensionality reduction behavior and collaborative learning to increase the clustering performance. Note that the value for α and β are automatically computed based on the defining Eqs. (15) and (16). Fig. 3(a) shows that these values of $\alpha$and β are significantly increased as the number of actual iteration increased and, accordingly, the estimated values of $\alpha$and β are worth in deciding how many features to be discarded without hurting the performances of clustering. For these feature reduction behaviors, we assume that features with the lower weights tend to be identified as the set of irrelevant features in one view and not required to be continuously used in clustering processes. Fig. 3(a) also presents the performance of AR's and RI's are enhancing after discarding the irrelevant feature in one view. The performances of AR's and RI's increase significantly from *AR=0.89* at the first iteration to *AR=0.99* at the fifth iteration; *RI=0.95* at the first iteration to *RI=0.99* at the fifth iteration. Fig. 3(b) presents the importance of one view during the clustering processes. Note that the larger the values, the more important the view is. From Fig. 3(a), (b), we can observe the parameter estimation for $\alpha$and β is proportional to gain the feature reduction behavior and simultaneously improved the performance of Co-FW-MVFCM. In general, we can say that the estimated parameter in Eqs. (15), (16) are highly effective in order to get the feature reduction behavior and implemented for collaborative learning. Furthermore, these $\alpha$and β automatically can demonstrate the importance of one view included its feature components. For the two-view data set with 2 clusters and 2 feature components as shown in Fig. 1, we also give Fig. 4 to demonstrate its road map with feature-view weights and memberships. As shown in Fig. 4, the cluster memberships, cluster centers, feature weights, and view weights are generated within each view. Different cluster memberships are updating during processing by sharing their information with each other. In general, the collaboration step can be used to detect data structure more accessible. This

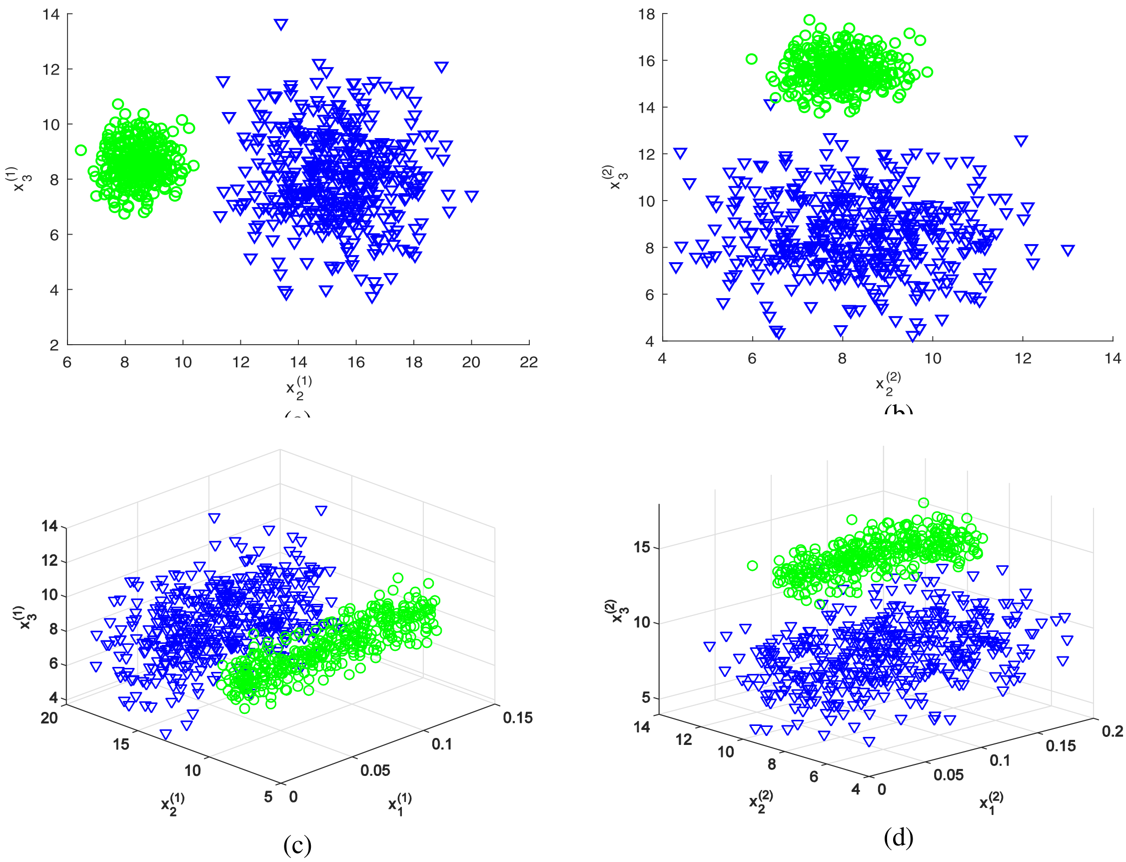

Fig. 1. The 2-cluster data set for (a) view 1; (b) view 2; The 2-cluster data set with a noise feature for (c) view 1; (d) view 2.

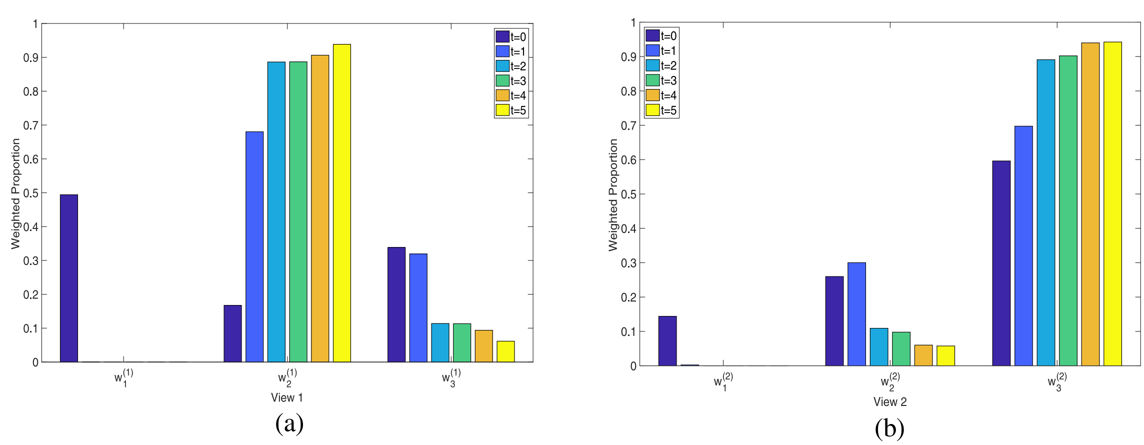

Fig. 2. Weighted proportion histogram of (a) view 1; (b) view 2.

is due to the learning paradigm that different views/sources can be gotten more insights according to the complementary information from multiple sources.

For clear demonstration, we next use a flowchart, as shown in Fig. 5, to illustrate the main procedures of the proposed Co-FW-MVFCM. In Fig. 5, it comprises three main parts: the local step, the feature reduction scheme, and the collaborative step. The initialization is a set of base partitions, and the output is the global solution for clustering. The local step is generated by feature-weighted MVFCM. The collaborative step is used to create an agreement between local memberships. The collaborative step uses these newly generated data after the feature reduction scheme is implemented. In this sense, some features that are recognized as redundant features are excluding during the collaborative step. This feature reduction scheme is used to discard unimportant features to get a global solution. We define a certain condition to measure important and unimportant features connected to the proposed Co-FW-MVFCM by simply taking the data number n and the dimension d_h in each view as a threshold.

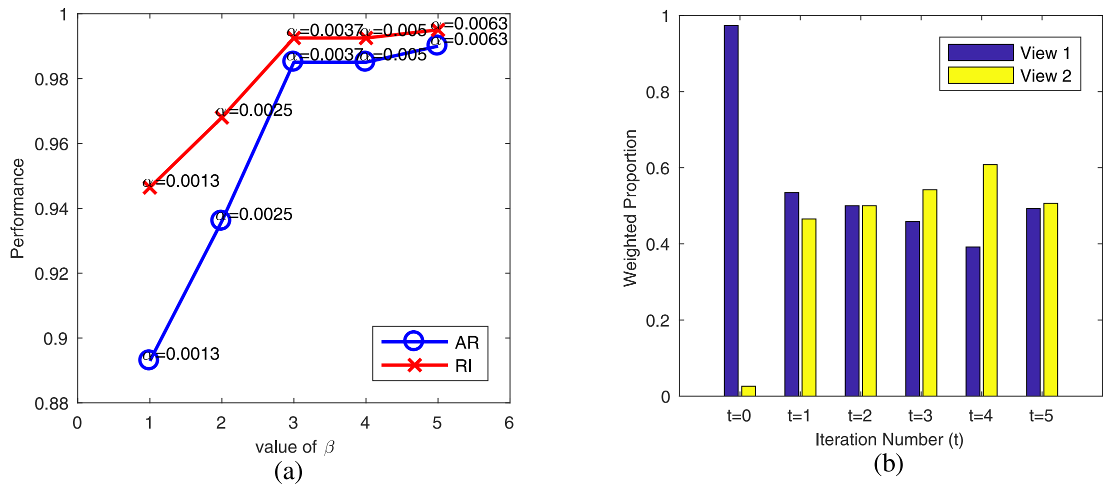

Fig. 3. (a) Relations of α and β to AR and RI performance of the proposed Co-FW-MVFCM; (b) Variation of the view-weight proportion in each iteration with the given α and β.

Thus, the proposed Co-FW-MVFCM clustering algorithm can be summarized as follows: *The Co-FW-MVFCM Algorithm*

Input: Dataset $\mathbf{X} = \{x_1, \ldots, x_n\}$ with $x_i = \{x_i^h\}_{h=1}^{s}$ and $x_i^h = \{x_{ij}^h\}_{j=1}^{d_h}$, number $c$of cluster, and $\varepsilon > 0$.

Output: $w_j^h, \upsilon_h, a_{kj}^h, \mu_{ik}^h$,and $\widehat{\mu}_{ik}$.

Initialization: Randomly generate initial U^h, Cluster centers A^h, initialize view weight $V^{(t)} = [\upsilon_h]_{1\times s}$ (user may define $\upsilon_h = 1/s\forall h$), and set $t = 1$.

Step 1: Calculate α and β by Eqs. (16) and (15).

Step 2: Compute the feature weight $W^{h^{(t)}}$ using $A^{h^{(t-1)}}$, $U^{h^{(t-1)}}$and $V^{(t-1)}$ by Eq. (12).

Step 3: Discard total d_r number of these j feature components for $W^{h^{(t)}}$ if $W^{h^{(t)}} < 1/\sqrt{nd_h}$, and set $d = d^{new} = D - d_r$.

Step 4: Adjust $W^{h^{(t)}}$ by $(w_j^h)' = w_j^h / \sum_{p=1}^{d_h(new)} w_p^h$.

Step 5: Update the view weight $V^{h^{(t)}}$ using $A^{h^{(t-1)}}$, $U^{h^{(t-1)}}$and $W^{h^{(t)}}$ by Eq. (11).

Step 6: Update the cluster centers $A^{h^{(t)}}$ using $W^{h^{(t)}}$ and $U^{h^{(t-1)}}$ by Eq. (13).

Step 7: Update membership $U^{h^{(t)}}$ using $W^{h^{(t)}}$, $A^{h^{(t)}}$and α by Eq. (10).

Step 8: If $(|\|U^{h^{(t)}}\| - \|U^{h^{(t-1)}}\||)/nc < \varepsilon$, then stop;

Else set $t = t + 1$ and go back to Step 1.

Step 9: Compute the global fuzzy partition matrix U by Eq. (9).

We finally analyze the computational complexity of the Co-FW-MVFCM algorithm. Suppose that the input n data points, c clusters, s views, d dimensions, and t training iterations for the Co-FW-MVFCM are given. The computational complexity of the Co-FW-MVFCM can be divided into four parts. The first part is to compute the feature weight $w_j^h$of each view with$O(n^2cdst)$. The second part is to update the view weight υ_h with $O(n^2cst)$. The third part is to compute the cluster center a_{kj}^h with $O(nst)$. The fourth part is to update the membership μ_{ik}^h for each view with$O(ncdst)$. Thus, the overall computational complexity of the Co-FW-MVFCM algorithm becomes as $O(n^2cdst)$.

4. Experimental results and comparisons

To evaluate the performance of the proposed Co-FW-MVFCM algorithm, we make the comparisons of Co-FW-MVFCM with these existing MVFCM algorithms of Co-FKM [9], MultiNMF [21], WV-Co-FCM [10], and minimax-FCM [12] in this section. For measuring clustering performance, we use an accuracy rate (AR), Rand index (RI) [24], Fowlkes-Mallows-Index (FMI) [25], Normalized mutual information (NMI) [26], and Jacard index (JI) [27]. AR is the percentage of data points that are correctly identified in clustering results. Let$C = \{C_1, C_2, \cdots, C_c\}$ be the set of c clusters for the given data set and $C' = \{C'_1, C'_2, \cdots, C'_c\}$ be the set of c clusters generated by the clustering algorithm. Let (X_i, X_j) be a given pair of points in the data set. Let a be the number of pairs of points if both points belong to the same cluster in C and the same cluster in C', b is the number of points if the two points belong to the same cluster in C and to two different clusters in C', and d be the number of pairs of points if the two points belong to two different clusters in $C$and to the same cluster in C'. RI is defined as $\mathrm{RI} = (a + d)/(n(n-1)/2)$ where n is the number of data points. NMI can defined as $NMI = 2I(X:Y)/[H(X) + H(Y)]$ where $I(X:Y)$is the mutual information between the class labels $H(X)$ and $H(Y)$. JI is commonly used to compares members for two sets X and Y to see which members are shared and which are distinct. It is defined as the size of the intersection divided by the size of the union of X and $Y$with $J(X, Y) = |X \cap Y|/|X \cup Y|$. These AR, RI, FMI, NMI, and JI ranges from 0 to 1, where 1 indicates a higher similarity between cluster solutions. The larger these AR, RI, FMI, NMI, and JI is, the better the clustering performance is.

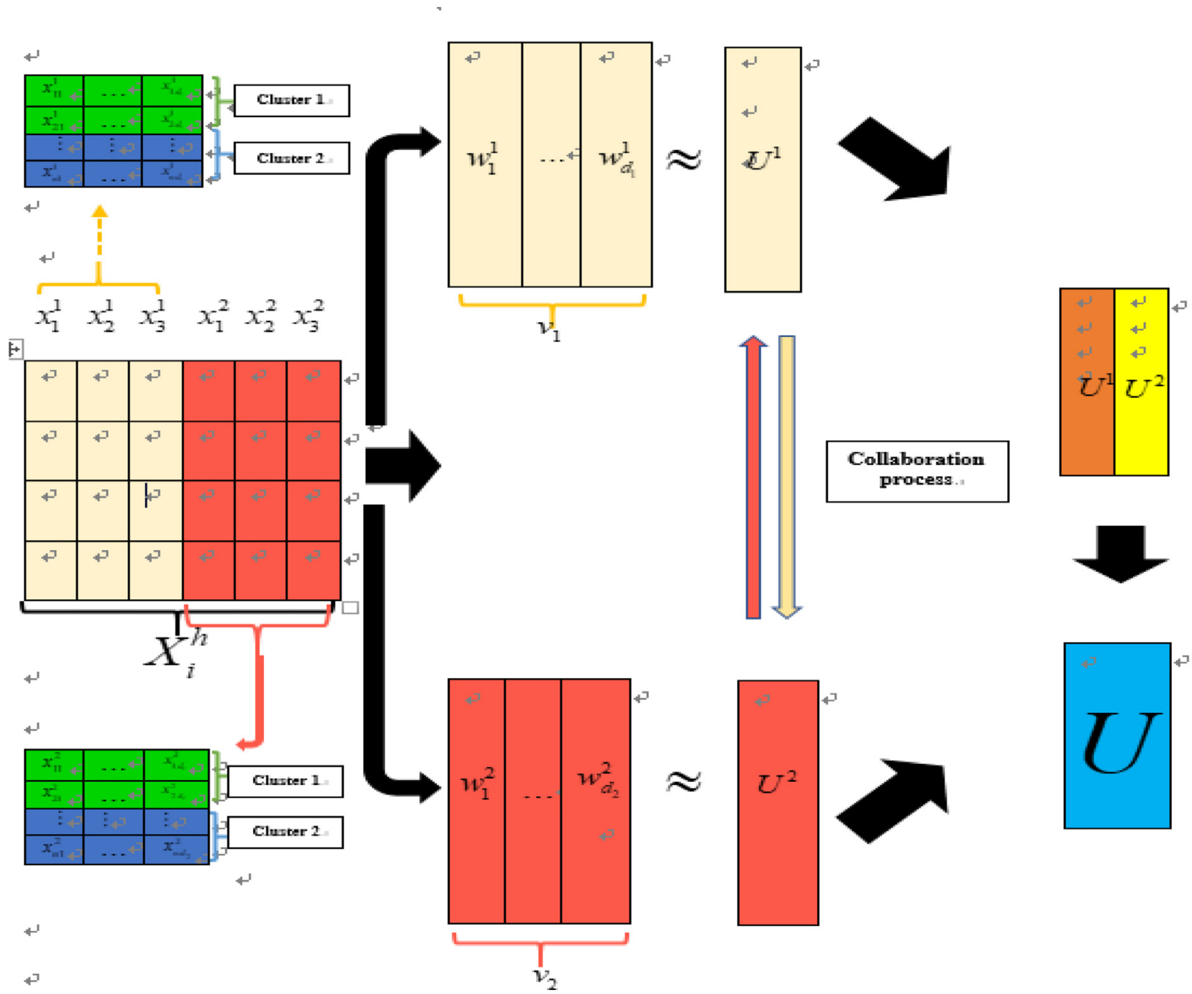

Fig. 4. Road map with feature-view weights and memberships for the multi-view data.

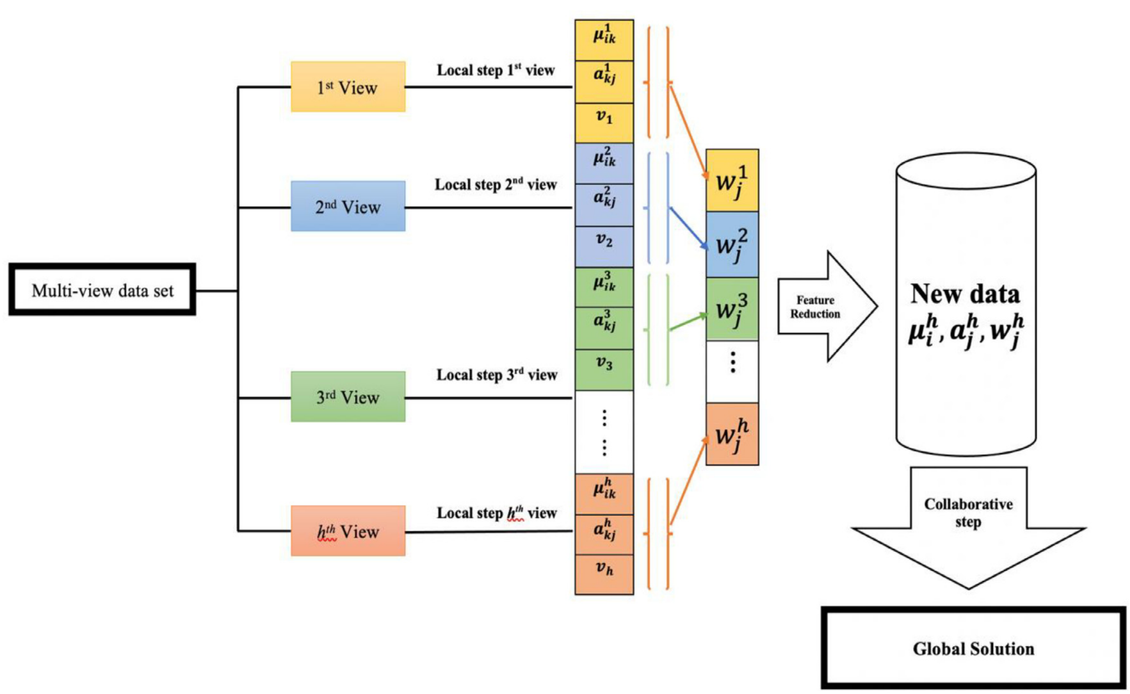

Fig. 5. The framework of Co-FW-MVFCM for multi-view data.

Table 1
A brief description of the six real multi-view data sets.

Data set	C	n	s	View name	d_h
Prokaryotic	4	551	3	Gene repertoire	393
				Proteome composition	3
				Textual	438
MSRC-V1	7	210	5	Color moment	24
				HOG	576
				GIST	512
				LBP	256
				CENTRIST	254
Caltech 101–7	7	1474	6	Gabor	48
				Wavelet moments	40
				CENTRIST	254
				HOG	1984
				GIST	512
				LBP	928
Image Segmentation	7	2310	2	Shape information	9
				RGB color model	10
IA	2	2359	6	Geometry of the image	3
				Base URL	457
				Image URL	495
				Target URL	472
				Anchor text	111
				Alt text	19
Wikipedia Articles	10	693	2	Text	128
				Image	10

Six real multi-view data sets are used to evaluate the performance of these MVFCM algorithms. These are Prokaryotic phyla [28] in Example 2, Microsoft Research Cambridge Volume 1 (MSRC-V1) data set [29] in Example 3, Wikipedia articles data set [30-31] in Example 4, Caltech 101–7 data set [32] in Example 5, image segmentation [33] in Example 6, and internet advertisement (IA) multi-view data [33] in Example 7. The characteristics of these six real data sets are summarized in Table 1 in terms of data types, the cluster number c, the data number n, the view number s, view names, and the feature dimension d_h. We compare the proposed Co-FW-MVFCM algorithm with Co-FKM, MultiNMF, WV-Co-FCM, and minimax-FCM. The parameter settings of the Co-FKM, WV-Co-FCM, Minimax FCM algorithms are presented in Table 2. The experiments are implemented in the Matlab 2017b with the same initializations. Experiments are repeated 40 times using different random initializations of the Co-FKM, WV-Co-FCM, Minimax FCM, Co-FW-MVFCM algorithms.

Example 2 In this example, we implement the proposed Co-FW-MVFCM, Co-FKM, MultiNMF, WV-Co-FCM, and Minimax FCM to a real biological data set, called Prokaryotic phyla [28]. It involves 551 types of Prokaryotic species where each species is described by three views including one textual data view and two genomic data views. The textual view is a document representation of prokaryotic species. The 2 genomic data views are known as gene repertoire and proteome composition. The proteome composition is encoded by 393 values representing relative frequencies of standard amino acids, while the gene repertoire is encoded by genome sequences, gene annotations, and gene families. These Prokaryotic genome sequences and gene annotations are from NCBI Entrez Genomes, while COG/NOG gene families are from eggNOG 3. Experiments are repeated 40 times on the data set, and the worst, average, and best of AR, RI, FMI, NMI, and JI are used to measure the performance of the Co-FKM, MultiNMF, WV-Co-FCM, Minimax FCM, and proposed Co-FW-MVFCM algorithms, as shown in Table 3, where the best result is highlighted in boldface. It is seen that the proposed Co-FW-MVFCM actually gives a comparative better performance than the other clustering algorithms, while the only Co-FKM algorithm gives the best NMI among them. The success of Co-FW-MVFCM in this example also show that the feature-reduction behavior in the proposed Co-FW-MVFCM can improve the performance of the clustering algorithm.

Example 3 In this example, the proposed Co-FW-MVFCM, Co-FKM, MultiNMF, WV-Co-FCM, and Minimax FCM algorithms are implemented on a real Microsoft Research Cambridge Volume 1 (MSRC-V1) data set [29]. This data set has eight classes, and each class has 240 images. In the experiment, 210 images and seven classes are selected. They are trees, buildings, airplanes, cow, face, car, bicycle, and each class have 30 images. Each image is represented by five different views: 24 features of the color moment, 254 features of CENTRIST, 512 features of GIST, 576 features of the histogram of oriented gradients (HOG), and 256 features of local binary patterns (LBP). Since different views of the data set have drastically different scales, to avoid overfitting issue, we perform min-max normalization for the data set. We normalize each view by$(X^h - \min X^h)/(\max X^h - \min X^h)$. Experiments are repeated 40 times on the data set over 40 different initializations. The worst, average, and best of AR, RI, FMI, NMI, and JI obtained by the Co-FKM, MultiNMF, WV-Co-FCM, Minimax FCM, and proposed Co-FW-MVFCM algorithms, as shown in Table 4, where the best result is highlighted in boldface. It is seen that the proposed Co-FW-MVFCM algorithm consistently outperform Co-FKM, MultiNMF, WV-Co-FCM, and Minimax FCM in terms of AR and RI. The Co-FKM algorithm gives the best in FMI, NMI, and JI. However, the clustering performance of the proposed Co-FW-MVFCM is superior to MultiNMF, WV-Co-FCM, and MinimaxFCM in terms of the average FMI's, NMI's, and JI's.

Example 4 To demonstrate the effectiveness of the proposed Co-FW-MVFCM clustering algorithm in structuring the multimedia data set, we implemented these Co-FKM, MultiNMF, WV-Co-FCM, Minimax FCM, and the proposed Co-FW-MVFCM into Wikipedia ar-

Table 2
Parameter setting for Co-FKM, MultiNMF, WV-Co-FCM, and minimax-FCM.

Algorithms	Data set	Parameter setup			
		m	β	η	λ_h
Co-FKM	Prokaryotic MSRC-V1	$m = \frac{\min(n,d-1)}{\max(n,0.5(d-s)-s^2C(s+c,c-1))}$	–	$\eta = \frac{s-1}{s}$	–
	Wikipedia Articles	$m = [1.05\ 1.10\ 1.15\ 1.25\ 1.35\ 1.5\ 1.75\ 2\ 2.2\ 2.4\ 2.6\ 2.8\ 3.0]$			
	Caltech 101–7	$m = [1.05\ 1.10\ 1.25\ 1.35\ 1.50\ 1.80\ 1.95\ 2.1\ 2.25\ 2.5\ 2.65\ 2.8]$		$\eta = 0.3$	
	IS IA	$m = [1.05\ 1.10\ 1.15\ 1.20\ 1.25\ 1.30\ 1.35\ 1.40\ 1.45\ 1.50]$		$\eta = \frac{s-1}{s}$	
MultiNMF	Prokaryotic MSRC-V1 Wikipedia Articles Caltech 101–7 IS IA	–	–	–	0.01
WV-Co-FCM	Prokaryotic MSRC-V1	$m = \frac{\min(n,d-1)}{\max(n,0.5(d-s)-s^2C(s+c,c-1))}$	$\beta = 4$	$\eta = \frac{s-1}{s}$	–
	Wikipedia Articles	$m = [1.05\ 1.10\ 1.15\ 1.25\ 1.35\ 1.5\ 1.75\ 2\ 2.2\ 2.4\ 2.6\ 2.8\ 3.0]$			
	Caltech 101–7	$m = [1.05\ 1.10\ 1.25\ 1.35\ 1.50\ 1.80\ 1.95\ 2.1\ 2.25\ 2.5\ 2.65\ 2.8]$			
	IS IA	$m = [1.05\ 1.10\ 1.15\ 1.20\ 1.25\ 1.30\ 1.35\ 1.40\ 1.45\ 1.50]$			
Minimax-FCM	Prokaryotic MSRC-V1	$m = \frac{\min(n,d-1)}{\max(n,0.5(d-s)-s^2C(s+c,c-1))}$	$\beta = \frac{s-1}{s}$	–	–
	Wikipedia Articles	$m = [1.05\ 1.10\ 1.15\ 1.25\ 1.35\ 1.5\ 1.75\ 2\ 2.2\ 2.4\ 2.6\ 2.8\ 3.0]$			
	Caltech 101–7	$m = [1.05\ 1.10\ 1.25\ 1.35\ 1.50\ 1.80\ 1.95\ 2.1\ 2.25\ 2.5\ 2.65\ 2.8]$			
	IS IA	$m = [1.05\ 1.10\ 1.15\ 1.20\ 1.25\ 1.30\ 1.35\ 1.40\ 1.45\ 1.50]$			

Table 3
Classification results of Co-FKM, MultiNMF, WV-Co-FCM, Minimax FCM, and the proposed Co-FW-MVFCM in Prokaryotic phyla multi-view data set.

	Co-FKM	MultiNMF	WV-Co-FCM	Minimax-FCM	Co-FW-MVFCM
AR	0.44/0.54/0.54	0.28/0.43/0.48	0.50/0.52/0.52	0..45/0.45/0.45	0.45/**0.63**/0.68
RI	0.61/0.61/0.64	0.57/0.59/0.61	0.56/0.56/0.57	0.60/0.60/0.60	0.60/**0.66**/0.71
FMI	0.45/0.45/0.46	0.37/0.40/0.41	0.50/0.52/0.52	0.39/0.39/0.39	0.45/**0.54**/0.59
NMI	0.36/**0.39**/0.39	0.19/0.22/0.23	0.24/0.24/0.25	0.21/0.21/0.21	0.24/0.31/0.35
JI	0.29/0.29/0.29	0.22/0.24/0.25	0.33/0.34/0.35	0.23/0.23/0.23	0.29/**0.36**/0.42

Table 4
Classification results of Co-FKM, MultiNMF, WV-Co-FCM, Minimax FCM, proposed Co-FW-MVFCM in MSRC-V1 multi-view data set.

	Co-FKM	MultiNMF	WV-Co-FCM	Minimax-FCM	Co-FW-MVFCM
AR	0.60/0.65/0.74	0.42/0.56/0.68	0.17/0.25/0.31	0.53/0.58/0.60	0.61/**0.66**/0.76
RI	0.85/0.87/0.89	0.81/0.84/0.87	0.56/0.67/0.75	0.85/0.86/0.87	0.85/**0.87**/0.89
FMI	0.54/**0.58**/0.61	0.39/0.47/0.56	0.24/0.27/0.30	0.50/0.53/0.55	0.46/0.54/0.62
NMI	0.56/**0.59**/0.64	0.42/0.50/0.59	0.11/0.25/0.31	0.51/0.54/0.55	0.48/0.55/0.63
JI	0.37/**0.40**/0.44	0.24/0.30/0.38	0.13/0.15/0.16	0.33/0.36/0.38	0.30/0.37/0.45

Table 5
Classification results of MultiNMF and proposed Co-FW-MVFCM in Wikipedia articles multi-view data set.

	AR	RI	FMI	NMI	JI
MultiNMF	0.48/0.53/0.56	0.87/0.88/0.89	0.43/0.46/0.49	0.49/0.52/0.54	0.28/0.30/0.32
Co-FW-MVFCM	0.59/**0.59**/0.60	0.89/**0.90**/0.90	0.51/**0.52**/0.53	0.55/**0.55**/0.56	0.34/**0.35**/0.36

Table 6
Classification results of Co-FKM, MultiNMF, WV-Co-FCM, Minimax FCM, and proposed Co-FW-MVFCM in Caltech 101–7 multi-view data set.

	Co-FKM	MultiNMF	WV-Co-FCM	Minimax-FCM	Co-FW-MVFCM
AR	0.37/0.63/0.76	0.34/0.39/0.48	0.27/0.49/0.75	0.33/0.41/0.46	0.64/**0.68**/0.72
RI	0.62/0.72/0.79	0.66/0.69/0.73	0.39/0.57/0.75	0.60/0.63/0.66	0.69/**0.75**/0.78
FMI	0.44/0.63/0.73	0.44/0.48/0.56	0.40/0.53/0.69	0.36/0.39/0.43	0.61/**0.66**/0.70
NMI	0.33/0.41/0.59	0.38/**0.43**/0.50	0.01/0.19/0.45	0.14/0.19/0.27	0.31/0.35/0.40
JI	0.28/0.45/0.57	0.25/0.28/0.35	0.25/0.35/0.53	0.20/0.23/0.27	0.44/**0.49**/0.54

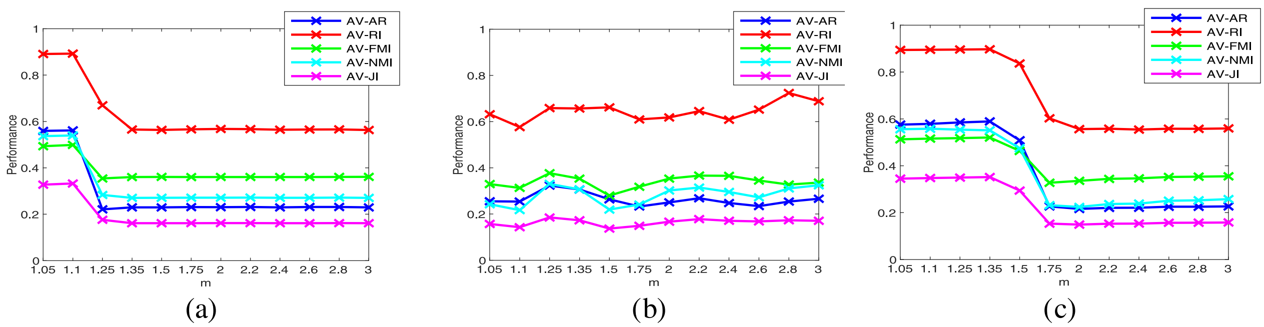

Fig. 6. Performance comparison of m ranges from 1.05 to 3 for (a) Co-FKM; (b) WV-Co-MVFCM; and (c) Minimax-FCM on Wikipedia Articles data set.

ticles data set [30,31]. We ran these five existing algorithms over 40 random initializations, and we set the t_{max}=20. Fig. 6(a)–(c) shows the effects of various m on Wikipedia data set for Co-FKM, WV-Co-FCM, and minimax FCM, respectively. We show the average AR, RI, FMI, NMI, and JI after 40 runs. From Fig. 6(a), we can observe that the clustering performance for Co-FKM starts to decrease when m=1.25 and reaches the best performance when m=1.1. The average AR's, RI's, FMI's, NMI's, and JI's of Co-FKM with m=1.1 are 0.56, 0.89, 0.50, 0.54, and 0.33, respectively. From Fig. 6(b), we can observe that the clustering performance for WV-Co-FCM tends to unstable and achieves the best performance of AR's, FMI's, NMI's, and JI's when m=1.25 with AV-AR=0.32, AV-FMI=0.38, AV-NMI=0.33, AV-JI=0.19; The average RI's of WV-Co-FCM achieved better performance when m=2.8 with AV-RI=0.72. While the lowest average of AR's is achieved when m=1.75 with AV-AR=0.23; The lowest average of RI's and NMI's is achieved when m=1.1 with AV-RI=0.58, AV-NMI=0.22; The lowest average of FMI's and JI's is achieved when m=1.5 with AV-FMI=0.28, AV-JI=0.14. Fig. 6(c) shows that the Minimax FCM starts to decrease when m=1.5 and reaches stable performances when m=[1.05 1.1 1.25 1.35]. The average AR's, RI's, FMI's, and JI's of Minimax FCM achieved better performance when m=1.35 with AV-AR=0.59, AV-RI=0.90, AV-FMI=0.52, AV-JI=0.35; The average NMI's of Minimax FCM achieved better performance when m=1.1 with AV-NMI=0.56. While the lowest average of AR's, NMI's, and JI's is achieved when m=2 with AV-AR=0.22, AV-NMI=0.22, and AV-JI=0.15; The lowest average of RI's is achieved when m=2.4 with AV-RI=0.56; The lowest average of FMI's is achieved when m=1.75 with AV-FMI=0.33. The clustering performance of the proposed Co-FW-MVFCM and MultiNMF clustering algorithms are reported in Table 5. According to the results, the best performances of RI's, FMI's, NMI's, and JI's of Minimax FCM and Co-FW-MVFCM are quite competitive. As such, Minimax FCM outperforms the proposed Co-FW-MVFCM by at least 0.11–1.13% in terms of the average RI's, FMI's, NMI's, and JI's. However, the gap between the best and the worst performances of Minimax FCM in terms of AR's, RI's, FMI's, NMI's, and JI's are too large, that is 37.4%, 34.3%, 19.3%, 33.4%, and 20.3%, respectively. Based on these observations, we conclude that our proposed Co-FW-MVFCM outperforms the comparison algorithms in terms of the stability and efficiency compare to five clustering measurements on the Wikipedia data set.

Example 5 To further confirm our conclusion, we advance our experiment in this example. We implement Co-FKM, MultiNMF, WV-Co-FCM, minimax-FCM, and the proposed Co-FW-MVFCM on a real Caltech 101–7 data set [32]. The focus of this experiment is to evaluate the clustering performance of Co-FKM, WV-Co-FCM, and minimax-FCM under various fuzzier parameter m in handling a data set with more data points, data views, and feature components in one view. This Caltech 101–7 consists of a total 9146 images belonging to 101 categories. We select seven classes (Face, motorbikes, Dolla bill, Garfield, Snoopy, stop-sign, Windsor chair) of 101 categories with 1474 images. Each image is described by six views (48 features of Gabor, 254 features of CENTRIST, 1984 features of the histogram of oriented gradients (HOG), 40 features of wavelet moments, 928 features of local binary patterns (LBP), 512 features of GIST). Since each view of the Caltech-101–7 contains a higher feature dimension, we set the t_{max}=10 in our experiment. Here, we use the normalization technique of Li et al. [34] in this experiment. It is shown that a set of fuzzier parameters m in Co-FKM, WV-Co-FCM, and minimax-FCM may affect the performance of clustering algorithms by fixing the coefficient parameter of collaborative learning. In this experiment, we assign the collaborative parameter 0.3 for Co-FKM and $\eta = (s-1)/s$ for WV-Co-FCM. While the fuzzier parameters for Co-FKM, WV-Co-FCM, and minimax-FCM are [1.05 1.10 1.25 1.35 1.50 1.80 1.95 2.10 2.25 2.50 2.65 2.80]. Another problem in WV-Co-FCM is to choose the optimal parameter to control the view weights. Since there is no clear explanation in how the users select the optimal value of the exponent parameter of view weights, we set the fixed exponent parameterβ=4 for WV-Co-FCM. Hereinafter, we only estimate the first case of δ_{ik}^{h} with $\delta_{ik}^{h} = (\eta/(s-1))\sum_{h'=1,h'\neq h}^{s}\left(d_{ik}^{h'}\right)^{2}$in the WV-Co-FCM clustering algorithm. In this example, we randomly run each algorithm ten times on each of the twelve combinations of the fuzziness level m and then calculate the worst, average, and best of AR, RI, FMI, NMI, and JI, as shown in Table 6. It is seen that the proposed Co-FW-MVFCM algorithm gets the best AR, RI, FMI, and JI, and the MultiNMF algorithm gets the best NMI. The Co-FKM algorithm is still competitive, but both WV-Co-FCM and minimax-FCM algorithms are not competitive for the Caltech 101–7 data set.

Next, we study the detailed effect of the different fuzzier parameter m (ranges from 1.05 to 2.80) on Co-FKM, WV-Co-FCM,

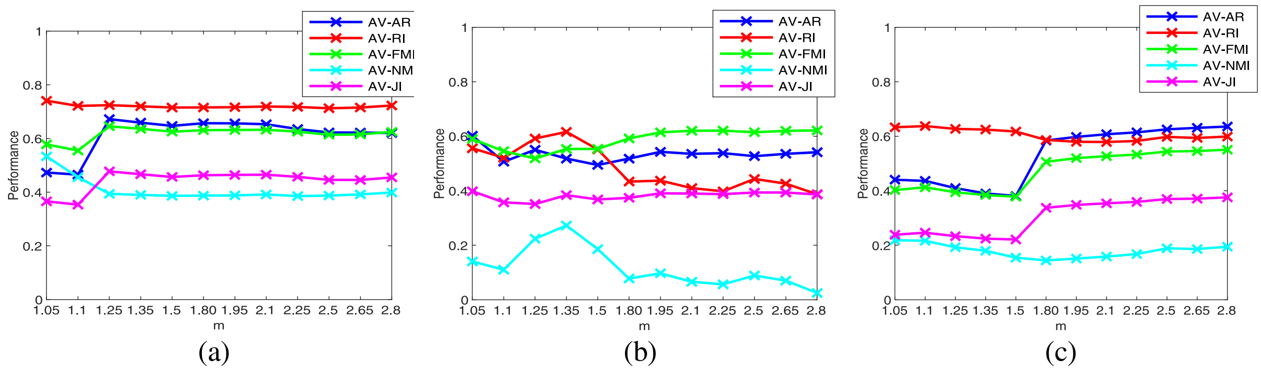

Fig. 7. Performance comparison of m ranges from 1.05 to 2.8 (a) Co-FKM; (b) WV-Co-MVFCM; and (c) Minimax-FCM on Caltech 101 data set.

and minimax-FCM w.r.t. the averages AV-AR, AV-RI, AV-FMI, AV-NMI and AV-JI. Fig. 7(a)–(c) shows the clustering performance of the Co-FKM, WV-Co-FCM, and minimax FCM under different values of m and a fixed η. The results indicate that different values of the fuzzier parameter m affect clustering performances. Fig. 7(a) shows that the average AR, FMI, and JI values of Co-FKM achieve better performance when m=*1.25* with AV-AR=0.67, AV-FMI=0.65, AV-JI=0.48. The average RI, and NMI values of Co-FKM achieve better performance when m=*1.05* with AV-RI=0.74, AV-NMI=0.53. The lowest average AR, FMI, and JI values achieve when m=*1.1* with AV-AR=0.47, AV-FMI=0.56, and AV-JI=0.35. Fig. 7(b) shows that the average AR value of WV-Co-FCM achieves better performance when m=*1.05* with AV-AR=0.60, the average RI value of WV-Co-FCM achieves better performance when m=*1.35* with AV-RI=0.62, the average FMI value of WV-Co-FCM achieves better performance when m=*2.8* with AV-FMI=0.62, the average NMI value of WV-Co-FCM achieves better performance when m=*1.35* with AV-NMI=0.27, and the average JI value of WV-Co-FCM achieves better performance when m=*2.65* with AV-JI=0.39. However, the lowest average AR value achieves when m=*1.5* with AV-AR=0.49, the lowest average RI value achieves when m=*2.8* with AV-RI=0.39, the lowest average FMI and JI values achieve when m=*1.25* with AV-FMI=0.52, AV-JI=0.35. The average AR, FMI, and JI values in minimax-FCM achieve better performance when m=2.80 with AV-AR=0.64, AV-FMI=0.55, and AV-JI=0.38. Fig. 7(c) shows that the average RI value of minimax-FCM achieves better performance when m=1.15 with AV-RI=0.64, while the high average NMI value achieves when m=1.05 with AV-NMI=0.22. From above analysis, we can conclude that the Co-FKM, WV-Co-FCM, and minimax-FCM algorithms are sensitive to the parameter of fuzzier m.

Example 6 The Image segmentation (IS) data set contains 2310 instances of seven outdoor images. These seven outdoor images are Brickface, Sky, Foliage, Cement, Window, Path, and Grass. Each instance is represented by two different views with 9 features of the shape information and 10 features of the RGB color. This data set is available at the UCI machine learning repository [33]. In this experiment, we use all features with normalization for these proposed Co-FW-MVFCM, Co-FKM, MultiNMF, WV-Co-FCM, and minimax-FCM algorithms. In the experiment, we repeat these five existing clustering algorithms for 40 times and report the worst, average, and best performances by using the parameter settings as presented in Table 2. For Co-FKM, WV-Co-FCM, and MinimaxFCM, the worst clustering performance is chosen by taking the small value among different m, and the best clustering performance is chosen by taking the highest value among different m, and the average is calculated by taking the mean values of all different m. The clustering results in terms of AR, RI, FMI, NMI, and JI are reported in Table 7. The better results are highlighted in boldface. As it can be seen, the proposed Co-FW-MVFCM and MultiNMF clustering performances are highly competitive to each other. From the results, the proposed Co-FW-MVFCM has the best average of AR's and RI's while MultiNMF has the best average of FMI's, NMI's, and JI's.

Example 7 In most clustering algorithms, the size of the data affects the clustering quality. To quantify this effect, in this example, we consider an internet advertisement (IA) [33] multi-view data with a high number of instances. This data set consists of 1559 columns and 3279 number of instances. Each instance in the IA data set represents one image that is tagged as an advertisement ("ad") and non-advertisement ("non-ad") in the last column. Each instance is described by six different views (Geometry of the image, base URL, image URL, alt text, anchor text, target URL). Three features represent the geometry of the image, 457 features represent base URL, 495 features represent image URL, 19 features represent alt text, anchor text is described by 111 features, and 472 features represent the target URL. As the number of missing values in the data set is very high (± 28%), we reduce the number of instances by dropping all the data points with missing entries (NaN value). This data set is available at the UCI machine learning repository [30]. We implement the Co-FKM, MultiNMF, WV-Co-FCM, Minimax FCM, and the proposed Co-FW-MVFCM algorithms and compared their performance based on the averages of AR, RI, FMI, NMI, and JI, as shown in Table 8, where the better results are highlighted in boldface. We find that the size of the data set seems not to influence the accuracy of the five algorithms. The results shown in Table 8 indicate that the proposed Co-FW-MVFCM algorithm outperform the other algorithms by at least 4%. Both Co-FKM and Minimax-FCM algorithms are competitive for clustering this IA multi-view data set. The WV-Co-FCM algorithm attains lower performance in terms of AR. The MultiNMF algorithm attains lower performance in terms of RI and NMI. Overall, the proposed Co-FW-MVFCM has the best performance with free of parameter setting and also better scalability. Furthermore, we present the final d and also total running times (in seconds) obtained by the proposed Co-FW-MVFCM algorithm. Table 9 shows their final dimensions and total running times for the data sets in Examples 2, 3, 4, 5, 6, and 7, respectively. From Table 9, it is seen that the proposed Co-FW-MVFCM reduces all feature dimensions in all views for these data sets in Examples 2, 6, and 7, while the proposed Co-FW-MVFCM identify that the feature components in view one of Example 3, that is the color moment view of MSRC-V1 data, are all important during clustering processes. Moreover, the proposed Co-FW-MVFCM identify that the feature components in view one and view two for the data set in Example 5, that are Gabor and wavelet moments views of Caltech-101–7 data, are all important during clustering processes. These are because these color moment views of MSRC-V1, Gabor and wavelet moment views of Caltech-101–7 are mutually independent (not correlated to each other in any way). In other words, these feature components are the most predic-

Table 7
Classification results of Co-FKM, MultiNMF, WV-Co-FCM, Minimax FCM, and proposed Co-FW-MVFCM in Image Segmentation multi-view data set.

	Co-FKM	MultiNMF	WV-Co-FCM	Minimax-FCM	Co-FW-MVFCM
AR	0.36/0.51/0.64	0.48/0.59/0.62	0.12/0.25/0.46	0.28/0.50/0.52	0.55/**0.58**/0.65
RI	0.79/0.84/0.86	0.83/**0.86**/0.88	0.14/0.61/0.79	0.78/0.84/0.85	0.84/**0.86**/0.87
FMI	0.36/0.47/0.53	0.46/**0.53**/0.60	0.20/0.31/0.53	0.39/0.51/0.53	0.48/**0.52**/0.55
NMI	0.41/0.53/0.60	0.53/**0.57**/0.63	0.00/0.16/0.57	0.44/**0.56**/0.57	0.48/0.53/0.57
JI	0.22/0.31/0.36	0.30/**0.36**/0.42	0.11/0.16/0.31	0.24/0.34/0.35	0.31/**0.35**/0.38

Table 8
Classification results of Co-FKM, MultiNMF, WV-Co-FCM, Minimax FCM, and the proposed Co-FW-MVFCM in Internet advertisement multi-view data set.

	Co-FKM	MultiNMF	WV-Co-FCM	Minimax-FCM	Co-FW-MVFCM
AR	0.71/**0.85**/0.90	0.84/0.84/0.84	0.80/0.83/0.84	0.83/0.84/0.84	0.86/**0.89**/0.90
RI	0.59/**0.75**/0.81	0.73/0.73/0.74	0.71/0.75/0.75	0.72/0.75/0.75	0.76/**0.81**/0.81
FMI	0.68/0.83/0.88	0.86/0.86/0.86	0.80/0.86/0.86	0.84/**0.86**/0.86	0.83/**0.87**/0.88
NMI	0.11/**0.24**/0.31	0.05/0.05/0.75	0.03/0.11/0.13	0.00/0.12/0.12	0.25/**0.30**/0.30
JI	0.50/0.71/0.78	0.73/0.73/0.73	0.69/0.74/0.74	0.72/**0.74**/0.74	0.71/**0.77**/0.78

Table 9
Results of final feature dimensions and total running times (in seconds) of the proposed Co-FW-MVFCM.

	Original *d*						Final *d* by Co-FW-MVFCM						
	v_1	V_2	V_3	V_4	V_5	V_6	V_1	V_2	V_3	V_4	V_5	V_6	Total Running Time (in seconds)
Ex. 2	393	3	438	–	–	–	48	2	139	–	–	–	716.88
Ex. 3	24	576	512	256	254	–	24	27	32	64	68	–	5515.21
Ex. 4	128	10	–	–	–	–	85	10	–	–	–	–	3472.11
Ex. 5	48	40	254	1984	512	928	48	40	244	461	377	466	9748.34
Ex. 6	9	10	–	–	–	–	2	8		–	–	–	8562.76
Ex. 7	3	457	495	472	111	19	2	270	288	270	100	12	31,388.44

Table 10
Comparison of total running times (in seconds) of the five algorithms for data in Examples 3, 5, and 6.

	Co-FKM	MultiNMF	WV-Co-FCM	Minimax-FCM	Co-FW-MVFCM
Ex. 3	5387.09	430.81	1935.01	77.47	5515.21
Ex. 5	645.53	510.60	2278.89	70.86	9748.34
Ex. 6	3818.15	288.69	6739.38	141.43	8562.76

tive power during clustering processing. For the data set in Example 4, Co-FW-MVFCM considers that feature-images are significantly important to reveal the article quality in Wikipedia. When human users want to get more relevant and precise answers, it could be a welcomed result because images-view can help human users more naturally to understand the article content. Furthermore, we also investigate the running times for these algorithms of Co-FKM, MultiNMF, WV-Co-FCM, minimax-FCM and Co-FW-MVFCM for the data in Examples 3, 5, and 6. The results are shown in Table 10. From Table 10, we find that the minimax-FCM is faster compared to others. The proposed Co-FW-MVFCM actually cost the most running time. However, we mention that these running times for the Co-FKM, MultiNMF, WV-Co-FCM, minimax-FCM and Co-FW-MVFCM algorithms reported in Table 10 are only for one user-defined parameter. In fact, if we consider the running times based on a variety of the fuzziness index m for Co-FKM, MultiNMF, WV-Co-FCM, and minimax-FCM, they even cost more running times, and if we run these algorithms to find better clustering results based on the good combination of user-defined parameters, they may cost more expensive than the proposed Co-FW-MVFCM. This is because the proposed WV-Co-FCM is free of parameter setting.

5. Conclusions

Discovering the structure of multi-view data based on a collaborative manner with view weights and feature weights simultaneously is challenging. In this paper, we propose a novel feature-weighted MVFCM clustering algorithm based on a collaborative idea, namely as the collaborative feature-weighted MVFCM (Co-FW-MVFCM). To find a strong correlation within views in multi-view data, the proposed Co-FW-MVFCM generates the fuzzy membership matrix of each view first to search for a global membership solution. The Co-FW-MVFCM then utilizes a collaborative manner to discover redundant features with a feature-reduction behavior. Our collaborative framework contains two step procedures that are a local step and a collaborative step. The local step is a single-view partition process to produce local partition clustering in each view. The collaborative step is sharing information between views about their individual membership to achieve better clustering results. These two steps are followed by an aggregation way which aims at reaching a global result after collaboration. The proposed Co-FW-MVFCM algorithm can exclude redundant features in each view if those features are smaller than a threshold. In experiments, we use one numerical dataset and six real datasets to evaluate the performances of the proposed Co-FW-MVFCM clustering algorithm in comparisons with the existing algorithms of Co-FKM, MultiNMF, WV-Co-FCM, and Minimax FCM. Based on the experimental results, we conclude that collaborative learning and feature reduction behaviors in Co-FW-MVFCM are practical to get more accurate performance. Furthermore, the proposed Co-FW-MVFCM is free of parameter setting that reflects Co-FW-MVFCM to be more scalability. However, combining collaborative learning with a feature reduction scheme for multi-view data in the proposed Co-FW-MVFCM actually cost more running times. On the other hand, the proposed Co-FW-MVFCM algorithm still needs to give a number of clusters a priori. In our future work, we will have advance study to give a mechanism for decreasing the running times of the proposed Co-FW-MVFCM. We should also consider in proposing a method to boost Co-FW-MVFCM with automatically finding an optimal number of clusters.

Declaration of Competing Interest

The authors declare there is no conflict of interest.

Acknowledgments

The authors would like to thank the anonymous referees for their helpful comments in improving the presentation of this paper. This work was supported in part by the Ministry of Science and Technology, Taiwan, under Grant MOST-109–2118-M-033–001-.

References

[1] J.C. Dunn, A fuzzy relative of the ISODATA process and its use in detecting compact, well-separated clusters, J. Cybern. 3 (1973) 32–57.
[2] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.
[3] F. Hoppner, F. Klawonn, R. Kruse, T. Runkler, Fuzzy Cluster Analysis: Methods for Classification Data Analysis and Image Recognition, Wiley, New York, 1999.
[4] M.S. Yang, W.L. Hung, F.J. Cheng, Mixed-variable fuzzy clustering approach to part family and machine cell formation for GT applications, Int. J. Prod. Econ. 103 (2006) 185–198.
[5] H. Izakian, W. Pedrycz, I. Jamal, Clustering spatiotemporal data: an augmented fuzzy c-means, IEEE Trans. Fuzzy Syst. 21 (2013) 855–868.
[6] M. Gong, L. Su, M. Jia, W. Chen, Fuzzy clustering with a modified MRF energy function for change detection in synthetic aperture radar images, IEEE Trans. Fuzzy Syst. 22 (2014) 98–109.
[7] S.T. Chang, K.P. Lu, M.S. Yang, Fuzzy change-point algorithms for regression models, IEEE Trans. Fuzzy Syst. 23 (2015) 2343–2357.
[8] M.S. Yang, S.J. Chang-Chien, Y. Nataliani, Unsupervised fuzzy model-based Gaussian clustering, Inf. Sci. 481 (2019) 1–23 (Ny).
[9] G. Cleuziou, M. Exbrayat, L. Martin, J.H. Sublemontier, CoFKM: a centralized method for multiple-view clustering, in: Proceedings of the IEEE International Conference on Data Mining (ICDM'09), 2009, pp. 752–757.
[10] Y. Jiang, F.L. Chung, S. Wang, Z. Deng, J. Wang, P. Qian, Collaborative fuzzy clustering from multiple weighted views, IEEE Trans. Cybern. 45 (2015) 688–701.
[11] S. Zeng, X. Wang, H. Cui, C. Zheng, D. Feng, A unified collaborative multikernel fuzzy clustering for multiview data, IEEE Trans. Fuzzy Syst. (2018) 1671–1687.
[12] Y. Wang, L. Chen, Multi-view fuzzy clustering with minimax optimization for effective clustering of data from multiple sources, Expert Syst. Appl. 72 (2017) 457–466.
[13] S. Sun, A survey of multi-view machine learning, Neural Comput. Appl. 23 (2013) 2031–2038.
[14] S. Sun, X. Xie, M. Yang, Multiview uncorrelated discriminant analysis, IEEE Trans. Cybern. 46 (2016) 3272–3284.
[15] X. Chen, S. Chen, H. Xue, X. Zhou, A unified dimensionality reduction framework for semi-paired and semi-supervised multi-view data, Pattern Recognit. 45 (2012) 2005–2018.
[16] X. Xin, J. Wang, R. Xie, S. Zhou, W. Huang, N. Zheng, Semi-supervised person reidentification using multi-view clustering, Pattern Recognit. 88 (2019) 285–297.
[17] M. Brbić, I. Kopriva, Multi-view low-rank sparse subspace clustering, Pattern Recognit. 73 (2018) 247–258.
[18] 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.
[19] W. Pedrycz, Collaborative fuzzy clustering, Pattern Recognit. Lett. 23 (2002) 1675–1686.
[20] A. Cornuéjols, C. Wemmert, P. Gançarski, Y. Bennani, Collaborative clustering: why, when, what and how, Inf. Fusion. 39 (2018) 81–95.
[21] J. Liu, C. Wang, J. Gao, J. Han, Multi-view clustering via joint nonnegative matrix factorization, in: Proceedings of the SIAM International Conference on Data Mining, SDM, 2013, pp. 252–260. 2013.
[22] M.S. Yang, Y. Nataliani, A feature-reduction fuzzy clustering algorithm based on feature-weighting entropy, IEEE Trans. Fuzzy Syst. 26 (2018) 817–835.
[23] L. Zhu, F.L. Chung, S. Wang, Generalized fuzzy c-means clustering algorithm with improved fuzzy partitions, IEEE Trans. Syst. Man, Cybern. Part B 39 (2009) 578–591.
[24] W.M. Rand, Objective criteria for the evaluation of clustering methods, J. Am. Stat. Assoc. 66 (1971) 846–850.
[25] E.B. Fowlkes, C.L. Mallows, A Method for comparing two hierarchical clusterings, J. Am. Stat. Assoc. 78 (1983) 553–569.
[26] T.M. Cover, J.A. Thomas, Elements of Information Theory, Wiley, New York, 1991.
[27] P. Jaccard, Distribution de la flore alpine dans le basin des Dranses et dans quelques regions voisines, Bull. De la Soc. Vaud. des Sci. Nat. 18 (2016) 1008–1018.
[28] M. Brbić, M. Piškorec, V. Vidulin, A. Kriško, T. Šmuc, F. Supek, The landscape of microbial phenotypic traits and associated genes, Nucleic Acid. Res. 44 (2016) 10074–10090.
[29] J. Winn, N. Jojic, LOCUS: learning object classes with unsupervised segmentation, in: Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05), 2005, pp. 1–8.
[30] J.C. Pereira, E. Coviello, G. Doyle, N. Rasiwasia, G.R.G. Lanckriet, R. Levy, N. Vasconcelos, On the role of correlation and abstraction in cross-modal multimedia retrieval, IEEE Trans. Pattern Anal. Mach. Intell. 36 (2014) 521–535.
[31] H. Wang, Y. Yang, T. Li, Multi-view clustering via concept factorization with local manifold regularization, in: Proceedings of the IEEE International Conference on Data Mining, ICDM, 2017, pp. 1245–1250.
[32] L. Fei-Fei, R. Fergus, P. Perona, Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories, Comput. Vision Image Underst. 106 (2007) 59–70.
[33] C.L. Blake, C.J. Merz, UCI repository of machine learning databases, a huge collection of artificial and real-world data sets, 1998. https://archive.ics.uci.edu/ml/datasets.html
[34] C.-.G. Li, C. You, R. Vidal, Structured sparse subspace clustering: a joint affinity learning and subspace clustering framework, IEEE Trans. Image Process. 26 (2017) 2988–3001.

Miin-Shen Yang received the B.S. degree in mathematics from Chung Yuan Christian University, Taoyuan, Taiwan, in 1977, the M.S. degree in applied mathematics from National Chiao-Tung University, Hsinchu, Taiwan, in 1980, and the Ph.D. degree in statistics from the University of South Carolina, Columbia, USA, in 1989. In 1989, he joined the faculty of the Department of Mathematics, Chung Yuan Christian University (CYCU), as an Associate Professor, where he has been a Professor, since 1994. From 1997 to 1998, he was a Visiting Professor with the Department of Industrial Engineering, University of Washington, Seattle, USA. From 2001 to 2005, he was the Chairman of the Department of Applied Mathematics, CYCU. Since 2012, he has been a Distinguished Professor of the Department of Applied Mathematics, the Director of Chaplain's Office, and is currently the Dean of the College of Science, CYCU. His-research interests include applications of statistics, fuzzy clustering, soft computing, pattern recognition, and machine learning. He was an Associate Editor of the *IEEE Transactions on Fuzzy Systems* from 2005 to 2011, and is an Associate Editor of the *Applied Computational Intelligence and Soft Computing*, an Editorial Board Member of Computer Science and Engineering section in the journal *Electronics* (MDPI).

Kristina P. Sinaga received B.S. degree and M.S. degree in mathematics from University of Sumatera Utara, Indonesia, and received Ph.D. degree in Department of Applied Mathematics from Chung Yuan Christian University, Taiwan. In 2020, she joined the faculty of the Department of Master in Information System Management, BINUS Graduate Program, BINUS University, Indonesia as a Lecture. Her research interests include clustering and pattern recognition.