

Federated Fuzzy C-Means with Schatten p -Norm Minimization

Paper ID: 4668

ABSTRACT

Federated multi-view clustering aims to provide a feasible and effective solution for handling unlabeled data owned by multiple clients. There are two main challenges: 1) The local data is always sensitive, thus preventing any inadvertent data leakage to the server or other clients. 2) Multi-view data contain both consistency and complementarity information, necessitating thorough exploration and utilization of these aspects to achieve enhanced clustering performance. Fully considering the above challenges, in this paper, we propose a novel federated multi-view method named Federated Fuzzy C-Means with Schatten- p Norm Minimization (FFCMSP) which is based on Fuzzy C-Means and Schatten p -norm. Specifically, we utilize the membership degrees to replace conventional hard clustering results in K-means, enabling improved uncertainty handling and less information loss. Moreover, we introduce a Schatten p -norm-based regularizer to fully explore the inter-view complementary information and global spatial structure. Correspondingly, we also proposed a federated optimization algorithm enabling clients to collaboratively learn the clustering results. Extensive experiments on several datasets demonstrate that our proposed method exhibits superior performance in federated multi-view clustering.

CCS CONCEPTS

• **Computing methodologies** → **Machine learning**; *Artificial intelligence*.

KEYWORDS

Fuzzy C-Means, Federated, multi-view clustering, tensor Schatten p -norm

1 INTRODUCTION

Multi-view data refers to data containing information from multiple perspectives (*e.g.*, modalities, sources, and viewpoints), and each view represents a distinct observation of the same object, thus, they generally have consistent and complementary information, which is beneficial for several applications. For example, by leveraging diverse types of medical data, such as CT scans and electrocardiograms, hospitals can obtain a more comprehensive assessment of a patient's physical condition. However, multi-view data learning confronts several challenges because the high cost of annotating complex multi-view data makes it difficult to acquire a substantial

number of high-quality labels. Thus, multi-view clustering has attracted considerable attention as an unsupervised learning approach [5, 35].

Existing multi-view clustering methods can be summarised into two categories [12]: neural network-based multi-view clustering and heuristic-based multi-view clustering. Furthermore, neural network-based multi-view clustering can be categorized into deep representation learning [40, 43] and deep graph learning [44, 45] while heuristic-based multi-view clustering can be categorized into non-negative matrix factorization [19, 30], graph learning [20, 28], latent representation learning [22, 46], and tensor learning [33].

Even though existing multi-view clustering methods achieve promising performance, they cannot be applied to scenarios where multi-view data is distributed across multiple clients, while clients are unwilling to share their data for the sake of privacy. Inspired by federated learning, which enables multiple clients to cooperatively train a global model without revealing any sensitive information [34], federated multi-view learning is developed. For example, Huang et al. extended multi-view clustering based on non-negative matrix factorization (NMF) to construct a federated multi-view clustering method named FedMVL [19] by dividing the global optimization problem into multiple sub-problems that can be solved locally to fully protect local data privacy.

However, FedMVL cannot obtain optimal performance because the underlying K-means enforces the clustering result to assign a concrete cluster label to each label and thus suffers from sensitivity to outliers and . Besides, it separates feature extraction and clustering into two steps and results in performance degradation. For another, FedMVL first learns each view's feature via NMF separately and then fuses multi-view features for clustering. Thus, it is unable to exploit the inter-view spatial structure of multi-view data.

To overcome these weaknesses, we propose a novel Federated Fuzzy C-Means with Schatten- p Norm Minimization (FFCMSP). Compared with FedMVL, FFCMSP replaces K-means with fuzzy C-Means. Since fuzzy C-Means directly learn soft assignment via membership degree matrix, it alleviates the sensitivity to outliers due to the hard assignment of K-means and caused by two-step clustering. Besides, inspired tensor-based methods [33, 48], we construct a tensor from the membership degree matrix of each view and validate Schatten p -norm based loss on it to learn the inter-view spatial structure. We further develop a distributed optimization algorithm, which enables all clients to optimize the model locally and collaboratively. Specifically, our contributions are as follows:

- We introduce fuzzy C-Means to federated multi-view clustering to overcome the weaknesses of K-means. It utilizes the membership degree instead of hard assignment to capture the overall structure of distributed data with heterogenic features.
- We construct a membership degree tensor from each view's membership degree matrix, and by minimizing the Schatten p -norm based regularizer, it can better explore the relationship between views and the complementary information of different views.

Permission to make digital or hard copies of all or part of this work for personal or

Unpublished working draft. Not for distribution. This work is distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ACM MM, 2024, Melbourne, Australia

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM

<https://doi.org/10.1145/nmmmmmmmmmmmm>

- To enable clients to cooperatively train a global model without any privacy data leakage, we developed a federated optimization algorithm to solve the tensor-based multi-view clustering objective.
- We conducted extensive experiments on multiple datasets to evaluate the model performance and compare it with several state-of-the-art methods, whose results demonstrate the superiority of our proposed approach.

2 RELATED WORK

2.1 Multi-View Clustering

Multi-view clustering aims to partition multi-view data into different clusters, which is widely applied in multi-view data labeling and pre-processing. Existing multi-view clustering methods include neural network-based multi-view clustering and heuristic multi-view clustering [12]. Although neural network based methods exhibit outstanding performance due to their ability to extract deep and nonlinear features of multi-view, their training cost is extremely high and is not suitable for lightweight applications. Differently, heuristic has lower computational complexity and better interpretability, which has attracted more attention recently.

There are mainly four categories of heuristic multi-view clustering methods: non-negative matrix factorization-based (NMF) methods [19, 30], graph learning-based methods [20, 28], latent representation learning methods [22, 46], and tensor learning-based methods [33]. The main idea of NMF-based methods is to factorize the data matrix of each view into a coefficient matrix and a basis matrix with lower dimensions for clustering. For example, Liu et al. [30] proposed joint multi-view NMF to learn a common consensus coefficient matrix; Liang et al. [29] introduced co-orthogonal constraints to capture the diversity within views and learn the orthogonal basis matrices. Multi-view graph clustering attempts to learn a consistent clustering structure from the correlation of samples, which is composed of graph fusion and graph partition. Wang et al. [41] firstly perceived the correlations between samples under the same view, and then all views adaptively collaborated to construct the consensus graph. Latent representation learning-based methods learn the shared latent representation of each view for clustering. [24] presents regularized and hybrid multiview coding (RHMC), which employs self-supervised learning to enhance the discriminative information of the shared feature.

The aforementioned methods all process multi-view data separately and then fuse the learned features or clustering results. To better exploit the inter-view spatial structure, tensor-based methods compose multi-view data into a three-order tensor. For instance, Lu et al. [33] constructed the third-order tensor from view-specific label matrices and minimized the divergence between the matrices to fully explore the complementary information between views. Li et al. [23] extended NMF into orthogonal non-negative tensor factorization and introduced tensor Schatten p -norm regularizer to fully utilize the complementary information provided by each view. Nevertheless, these methods are designed for centralized settings and ignore the possibility that multi-view data are held by different entities.

2.2 Federated Learning

Federated learning [34] is a distributed learning paradigm that enables multiple entities to collaboratively train a global model on data distributed across multiple devices without revealing any private information. The concept of federated learning and its specific algorithm, FedAVG, were proposed simultaneously [34]. FedAVG involves performing stochastic gradient descent on each client while aggregating the local trained models on the server by performing averaging. Afterward, many variants of federated learning were proposed: FedAMP [21], FedProx [27]. Federated learning can be categorized into three types: Horizontal Federated learning (HFL), vertical federated learning (VFL), and federated transfer learning (FTL).

In HFL, several clients share the same feature spaces but different sample spaces. [14] introduced HFL for Electroencephalography classification techniques to meet the requirements of privacy protection and data heterogeneity. Moreover, [47] focused on measuring the contribution to the model of all clients and the global server evaluates the contribution through reinforcement learning techniques. For VFL, several clients share the same sample space but different feature spaces. [6] proposed a novel method that enables each client to execute stochastic gradient algorithms independently, and a new technique of perturbed local embedding is proposed. Traditional VFL identifies the shared sample by Private Set Intersection (PSI), which may lead to some data leakage. Thus, [39] proposed a Private Set Union (PSU) based VFL framework to avoid such kind of data leakage. For FTL, clients share both different sample spaces and feature spaces. [31] first proposed the concept of FTL, enabling a target-domain party to build flexible and effective models by leveraging rich labels from a source domain. All the previously mentioned methods are designed for single-view data and fail to fully exploit the latent information inherent in multi-view data for learning.

2.3 Federated Multi-view Clustering

Traditional multi-view clustering methods generally assume that all data are held by a single party and never consider the application in which multi-view data are distributed across different clients and privacy-preservation. To tackle this problem, federated multi-view learning was proposed, and there are several federated multi-view learning methods developed for different settings. Multi-view learning was introduced to the personalized recommendation system [10], cancer subtype identification [7], and other domains.

However, similarly, labeling multi-view data is costly, several times more expensive than labeling single-view data. This led to the emergence of federated multi-view clustering (FMVC). [4] introduced FMVC to the medical field. The proposed method applies to both VFL and HFL and keeps sensitive data private. [8] is a federated multi-view clustering method based on neural networks and guarantees data security by transmitting non-sensitive parameters between clients and the global server. Neural network-based FMVC methods are always time-consuming. Thus, most FMVC methods are based on heuristic clustering methods. [19] combines NMF and K-means to construct a time-efficient FMVC framework and the final clustering result comes from the coefficient matrix. However, because of using K-means on the coefficient matrix, we cannot effectively capture the relationship between samples and clusters, leading

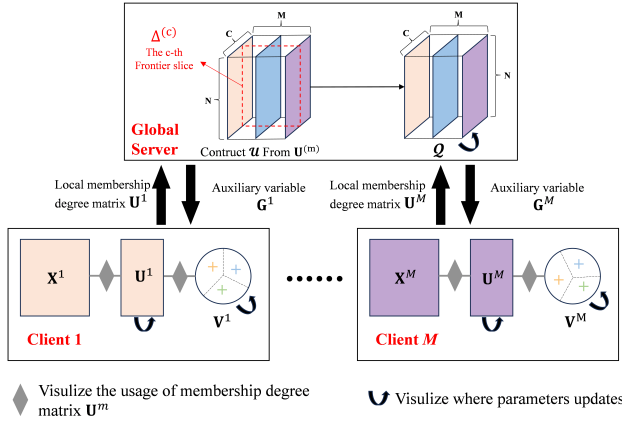


Figure 1: The framework of the proposed FFCMSP. The federated framework contains one global server and M local clients. X^m , V^m denotes the data matrix and center cluster centroid matrix. The grey \diamond represents that the membership matrix U is used to describe the membership relationship between samples and cluster centroid. Moreover, black \cup visually represents whether the parameters are updated on the client or the server. The construction of tensor \mathcal{U} is also visualized. Finally, the $\Delta^{(c)}$ denotes the c -th frontier slice of \mathcal{U} and it well characterize the complementary information embedded in inter-views. [15]

to some information loss and performance degradation. Therefore, [17] proposed fuzzy C-Means-based FMVC which uses membership degree instead of hard clustering assignment. Nevertheless, it fails to capture the consistent information between views.

3 METHOD

3.1 Preliminaries

For easier presentation, we introduce Schatten p -norm [32] as follows.

DEFINITION 1. The matrix Schatten p -norm for matrix $M \in \mathbb{R}^{n_1 \times n_2}$ is defined as below:

$$\|\mathbf{M}\|_{sp} = \left(\sum_{i=1}^{\min(n_1, n_2)} \sigma_i^p(\mathbf{M}) \right)^{\frac{1}{p}} \quad (1)$$

where $\sigma_i(\mathbf{M})$ is the i -th singular value of matrix \mathbf{M} .

DEFINITION 2 (TENSOR SINGULAR VALUE DECOMPOSITION, T-SVD). Given a tensor $\mathcal{M} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, then we have its t -SVD:

$$\mathcal{M} = \mathbf{U} * \mathcal{S} * \mathbf{V}^T \quad (2)$$

where $\mathbf{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$, $\mathbf{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ are both orthogonal tensors, and $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is diagonal tensor. Moreover, all frontal slices of \mathcal{S} are diagonal matrixes, and the i -th singular value of k -th frontal slices of \mathcal{S} is denoted as $\mathcal{S}^k(i)$.

DEFINITION 3 (TENSOR SCHATTEN p -NORM). Based on Definition 1 and Definition 2, the tensor Schatten p -norm for tensor \mathcal{M}

is defined as below:

$$\|\mathcal{M}\|_{sp} = \left(\sum_{i=1}^{\min(n_1, n_2)} \sum_{k=1}^{n_3} (\mathcal{S}^k(i))^p \right)^{\frac{1}{p}} \quad (3)$$

where $0 \leq p \leq 1$, tensor Schatten p -norm has proven to be a better rank approximation method than the tensor nuclear norm (TNN), which motivates us to apply it in our work to learn more robust and clustering results. Given a typical Tensor Schatten p -norm problem, it can be solved with the following Lamma.

LAMMA 1. For $\mathcal{B} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{C} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the function

$$\min_{\mathcal{B}} \tau \|\mathcal{B}\|_{sp}^p + \frac{1}{2} \|\mathcal{B} - \mathcal{C}\|_F^2 \quad (4)$$

has the optimal solution given by

$$\mathcal{B}^* = \text{iffi}(\mathbf{U} * \mathcal{D}_{\tau, p}(\mathcal{C}) * \mathbf{V}^T) \quad (5)$$

where \mathbf{U} and \mathbf{V} are obtained via t -SVD of \mathcal{C} , i.e. $\mathcal{C} = \mathbf{U} * \mathcal{S} * \mathbf{V}^T$, and $\mathcal{D}_{\tau, p}(\mathcal{C}^{(m)}) = \text{diag}(\zeta(\mathcal{C}^{(m)}))$, $\zeta(\mathcal{C}^{(m)}) = \text{GST}(\sigma(\mathcal{S}^{(m)}, \tau, p))$. The GST algorithm is introduced in [15].

3.2 Objective Function

Problem Statement: We consider the same federated multi-view settings as that in [19], which includes a centralized server \mathcal{S} and M clients. Multi-view data, denoted as $\mathbf{X} = \{\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \dots, \mathbf{X}^{(M)}\}$, is held by different clients. Specifically, the m -th view $\mathbf{X}^{(m)} \in \mathbb{R}^{N \times d^{(m)}}$ ($m = 1, 2, \dots, M$) is distributed in the m -th client, C_m , where $d^{(m)}$ is the feature dimension, N is sample number, and M is the number of views as well as client number. Our goal is enable the server and clients to cooperatively train a global clustering model without exchanging any raw data.

Federated Multi-View Fuzzy C-Means: Our work is derived from fuzzy C-Means (FCM) [1], which assigns data points to clusters with membership degree matrix. Suppose there are N samples of dimension D denoted by $\mathbf{X} \in \mathbb{R}^{N \times D}$, FCM divides them into C clusters with the following objective:

$$\begin{aligned} \min_{U, V} \sum_{i=1}^N \sum_{j=1}^C u_{ij}^t \|x_i - v_j\|_F^2 \\ \text{s.t.} \quad \sum_{j=1}^C u_{ij} \geq 0 = 1, u_{ij} \geq 0, t \geq 1 \end{aligned} \quad (6)$$

where u_{ij} is the membership degree, representing the probability that the i -th sample belongs to the j -th cluster; $\mathbf{V} \in \mathbb{R}^{C \times D}$ is the clustering centroids matrix; t is the fuzzification coefficient, and bigger t tends to defocus membership towards the fuzziest state [1], whose suggested range is $1 \leq t \leq 3$.

Compared with K-means, FCM can better resist noises or outliers, motivated by which, we adopt multi-view FCM as our underlying multi-view clustering method, which can be expressed as follows:

$$\begin{aligned} \min_{U^{(m)}, V^{(m)}} \sum_{m=1}^M \sum_{i=1}^N \sum_{j=1}^C (u_{ij}^{(m)})^t \|x_i^{(m)} - v_j^{(m)}\|_F^2 \\ \text{s.t.} \quad \sum_{j=1}^C u_{ij}^{(m)} = 1, u_{ij}^{(m)} \geq 0, t \geq 1 \end{aligned} \quad (7)$$

where $\mathbf{X}^{(m)}$ is the data matrix located on the m -th clients and $\mathbf{V}^{(m)}$ is the corresponding cluster centroids. t is also the fuzzification coefficient, similarly. To simplify the optimization process, this paper sets the fuzzification coefficient t to 1. Thus, based on Eq. (7), the objective function is:

$$\begin{aligned} \min_{\mathbf{U}^{(m)}, \mathbf{V}^{(m)}} & \sum_{m=1}^M \sum_{i=1}^N \sum_{j=1}^C u_{ij}^{(m)} \|\mathbf{x}_i^{(m)} - \mathbf{v}_j^{(m)}\|_F^2 \\ \text{s.t.} & \sum_{j=1}^C u_{ij}^{(m)} = 1, u_{ij}^{(m)} \geq 0 \end{aligned} \quad (8)$$

Tensor Schatten p -Norm Based Regularizer: By observing Eq. (8), we find that in each view, each sample corresponds to a membership degree u_{ij} for each cluster, and for N samples and C clusters, the membership degrees form a matrix $\mathbf{U}^{(m)} \in \mathbb{R}^{N \times C}$, thus latent information is hidden between different view-specific membership degree matrix. Inspired by [33], we introduce a regularizer based on the tensor Schatten p -norm to enforce the model to learn the global spatial structure and inter-view complementary information, the final objective function becomes:

$$\begin{aligned} \min_{\mathbf{U}^{(m)}, \mathbf{V}^{(m)}} & \sum_{m=1}^M \sum_{i=1}^N \sum_{j=1}^C u_{ij}^{(m)} \|\mathbf{x}_i^{(m)} - \mathbf{v}_j^{(m)}\|_F^2 + \lambda \|\mathbf{U}\|_{sp}^p \\ \text{s.t.} & \sum_{j=1}^C u_{ij}^{(m)} = 1, u_{ij}^{(m)} \geq 0 \end{aligned} \quad (9)$$

where λ is the trade-off parameter and the membership degree tensor \mathbf{U} is constructed by stacking all views' local membership degree matrix and taking a simple rotation as shown in Fig. 1.

4 OPTIMIZATION ALGORITHM

In centralized scenarios, Eq. (9) can be easily solved. However, with the multi-view data distributed to local clients, our design goal is to employ federated learning to collaboratively learn a clustering model while considering the requirement of data privacy. This requires that the raw models or parameters transmitted between clients and the global server should be non-sensitive, which poses extra challenges of optimization. To address this issue, we further develop a federated optimization algorithm in this scenario.

Firstly, we introduced augmented Lagrange multiplier (ALM) to Eq. (9):

$$\begin{aligned} \min \mathcal{L}(\mathbf{U}, \mathbf{G}, \mathbf{V}) &= \sum_{m=1}^M \sum_{i=1}^N \sum_{j=1}^C u_{ij}^{(m)} \|\mathbf{x}_i^{(m)} - \mathbf{v}_j^{(m)}\|_F^2 \\ &+ \lambda \|\mathbf{G}\|_{sp}^p + \langle \mathbf{Q}, \mathbf{U} - \mathbf{G} \rangle + \frac{\mu}{2} \|\mathbf{U} - \mathbf{G}\|_F^2 \\ \text{s.t.} & \sum_{j=1}^C u_{ij}^{(m)} = 1, u_{ij}^{(m)} \geq 0 \end{aligned} \quad (10)$$

where μ is the penalty parameter, \mathbf{Q} is the Lagrange multiplier, and $\langle \cdot \rangle$ denotes the inner product operation. The optimization algorithm is divided into three parts, as described below.

Solving $\mathbf{U}^{(m)}$ with fixed \mathbf{G} and \mathbf{V} : with fixed \mathbf{G} and \mathbf{V} , the objective function becomes:

$$\begin{aligned} \min & \sum_{m=1}^M \sum_{i=1}^N \sum_{j=1}^C u_{ij}^{(m)} \|\mathbf{x}_i^{(m)} - \mathbf{v}_j^{(m)}\|_F^2 + \langle \mathbf{Q}, \mathbf{U} - \mathbf{G} \rangle + \frac{\mu}{2} \|\mathbf{U} - \mathbf{G}\|_F^2 \\ \text{s.t.} & \sum_{j=1}^C u_{ij}^{(m)} = 1, u_{ij}^{(m)} \geq 0 \end{aligned} \quad (11)$$

Considering the independence of \mathbf{U} , \mathbf{Q} , and \mathbf{G} , Eq. (10) could be decomposed into M independent subproblems, where the m -th subproblem is:

$$\begin{aligned} \min & \sum_{i=1}^N \sum_{j=1}^C u_{ij}^{(m)} \|\mathbf{x}_i^{(m)} - \mathbf{v}_j^{(m)}\|_F^2 + \langle \mathbf{Q}^{(m)}, \mathbf{U}^{(m)} - \mathbf{G}^{(m)} \rangle \\ &+ \frac{\mu}{2} \|\mathbf{U}^{(m)} - \mathbf{G}^{(m)}\|_F^2 \\ \text{s.t.} & \sum_{j=1}^C u_{ij}^{(m)} = 1, u_{ij}^{(m)} \geq 0 \end{aligned} \quad (12)$$

Through simple algebra, Eq. (12) can be transformed into:

$$\begin{aligned} \min & \sum_{i=1}^N \sum_{j=1}^C u_{ij}^{(m)} \mathbf{D}_{ij}^{(m)} + \frac{\mu}{2} \|\mathbf{U}^{(m)} - \mathbf{T}^{(m)}\|_F^2 \\ \text{s.t.} & \sum_{j=1}^C u_{ij}^{(m)} = 1, u_{ij}^{(m)} \geq 0 \end{aligned} \quad (13)$$

where $\mathbf{D}_{ij}^{(m)} = \|\mathbf{x}_i^{(m)} - \mathbf{v}_j^{(m)}\|_F^2$, and $\mathbf{T}^{(m)} = \mathbf{G}^{(m)} - \frac{\mathbf{Q}^{(m)}}{\mu}$. By introducing the Karush-Kuhn-Tucker (KKT) condition, we have:

$$u_{ij}^{(m)} = \frac{C\mathbf{T}_{ij}^{(m)} - \sum_{j=1}^C \mathbf{T}_{ij}^{(m)}}{C} + \frac{C\mathbf{D}_{ij}^{(m)} - \sum_{j=1}^C \mathbf{D}_{ij}^{(m)}}{C\mu} + \frac{1}{C} \quad (14)$$

Solving \mathbf{G} with fixed \mathbf{U} and \mathbf{V} : With fixed \mathbf{U} and \mathbf{V} , the objective function concerning \mathbf{G} becomes:

$$\begin{aligned} \min & \lambda \|\mathbf{G}\|_{sp}^p + \langle \mathbf{Q}, \mathbf{U} - \mathbf{G} \rangle + \frac{\mu}{2} \|\mathbf{U} - \mathbf{G}\|_F^2 \\ &= \min \lambda \|\mathbf{G}\|_{sp}^p + \frac{\mu}{2} \|\mathbf{G} - \mathcal{P}\|_F^2 \end{aligned} \quad (15)$$

where $\mathcal{P} = \mathbf{U} - \frac{\mathbf{Q}}{\mu}$. From Lemma. 1, we get the optimal solution of Eq. (15):

$$\mathbf{G}^* = \text{ifft} \left(\mathbf{U} * \mathcal{D}_{\frac{\lambda}{\mu}, p}(\mathcal{P}) * \mathbf{V}^T \right) \quad (16)$$

where \mathbf{U} and \mathbf{V}^* are obtained via t-SVD of \mathcal{P} , i.e., $\mathcal{P} = \mathbf{U} * \mathbf{S} * \mathbf{V}_*^T$.

Solving $\mathbf{V}^{(m)}$ with fixed \mathbf{U} and \mathbf{G} : With fixed \mathbf{U} and \mathbf{V} , the objective function concerning \mathbf{V} becomes:

$$\min \sum_{m=1}^M \sum_{i=1}^N \sum_{j=1}^C u_{ij}^{(m)} \|\mathbf{x}_i^{(m)} - \mathbf{v}_j^{(m)}\|_F^2 \quad (17)$$

Similarly, we decompose Eq. (17) into M independent subproblems owing to the independence of \mathbf{U} , and the m -th subproblem is defined as:

$$\min \mathcal{L}_{\mathbf{V}^{(m)}} = \min \sum_{i=1}^N \sum_{j=1}^C u_{ij}^{(m)} \|\mathbf{x}_i^{(m)} - \mathbf{v}_j^{(m)}\|_F^2 \quad (18)$$

We take the partial derivative of Eq. (18) concerning $\mathbf{v}_j^{(m)}$ and set it to zero:

$$\begin{aligned} \sum_{i=1}^N u_{ij}^{(m)} \mathbf{x}_i^{(m)} &= \sum_{i=1}^N u_{ij}^{(m)} \mathbf{v}_j^{(m)} \\ \mathbf{v}_j^{(m)} &= \frac{\sum_{i=1}^N u_{ij}^{(m)} \mathbf{x}_i^{(m)}}{\sum_{i=1}^N u_{ij}^{(m)}} \end{aligned} \quad (19)$$

4.1 Federated workflow

Considering local client data security, the data or parameters transmitted between the client and the server must be insensitive. Therefore, based on the above optimization algorithm, we propose the workflow in the federated setting. By observing Eq. (14), Eq. (16), and Eq. (19), we draw the following conclusions:

- The update of $u_{ij}^{(m)}$ is related to $d_{ij}^{(m)}$, $\mathbf{G}^{(m)}$, and $\mathbf{Q}^{(m)}$, where $d_{ij}^{(m)}$ describes the distance between i -th sample and j -th cluster centroid and it can be calculated locally. Therefore, each client can update $u_{ij}^{(m)}$ locally as long as they have a backup of $\mathbf{G}^{(m)}$ and $\mathbf{Q}^{(m)}$.
- The update of \mathcal{G} is related to \mathcal{U} and \mathcal{Q} , thus, it can only be performed globally and $\mathbf{U}^{(m)}$ should be transmitted to the global server in advance.
- The update of $\mathbf{V}^{(m)}$ is related to $\mathbf{U}^{(m)}$ and $\mathbf{X}^{(m)}$, and it can be performed locally because the relevant data is stored locally.
- The optimal solution of \mathcal{Q} is $\mathcal{Q} \leftarrow \mathcal{Q} + \mu(\mathcal{U} - \mathcal{G})$, which can only be performed globally because it requires tensor \mathcal{U} .

After the above analysis, we described the detailed workflow in the federated setting.

- (1) Firstly, each client C_m initializes $u_{ij}^{(m)} = \frac{1}{C}$ and $\mathbf{V}^{(m)}$ randomly, and the global server initializes $\mathcal{G} = \mathcal{U}$ and $\mathcal{Q} = \mathbf{0}$.
- (2) In each communication round, each client C_m updates $\mathbf{U}^{(m)}$ with Eq. (14) and $\mathbf{V}^{(m)}$ with Eq. (19) locally and then sends updated $\mathbf{U}^{(m)}$ to the global server \mathcal{S} .
- (3) After receiving them, \mathcal{S} stacks all $\mathbf{U}^{(m)}$ into a tensor \mathcal{U} , and then update \mathcal{G} with Eq. (16) and $\mathcal{Q} \leftarrow \mathcal{Q} + \mu(\mathcal{U} - \mathcal{G})$. The global server then transmits the m -th slice of updated \mathcal{G} and \mathcal{G} , $\mathbf{G}^{(m)}$ and $\mathbf{Q}^{(m)}$, to the corresponding client.
- (4) The step (2) and (3) repeat until convergence, and then, the final cluster assignment matrix \mathbf{H} is calculated globally by:

$$h_{ij} = \begin{cases} 1 & j = \arg \min_p \left(\sum_{m=1}^M u_i^{(m)} \right)_p \\ 0 & \text{otherwise} \end{cases} \quad (20)$$

The detailed workflow is summarized in *Algorithm 1*.

4.2 Complexity Analysis

4.2.1 Computational Complexity. The complexity of FFCSP should be analyzed from two aspects, i.e., local client-side and global server-side.

Local client-side: The computation load on each client includes the update of $\mathbf{U}^{(m)}$ and $\mathbf{V}^{(m)}$. In each local iteration, the computational complexity of the update of $\mathbf{U}^{(m)}$ is $\mathcal{O}\left(NC\left(d_i^{(m)} + C\right)\right)$ while the computational complexity of the update of $\mathbf{V}^{(m)}$ is $\mathcal{O}(NC)$. Thus,

Algorithm 1 Federated Fuzzy C-Means with Schatten p -Norm Minimization(FFCMSP)

input: The data $\mathbf{X} = \{\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \dots, \mathbf{X}^{(M)}\}$ on M local clients, cluster number C ; parameters: μ, p, λ ;

output: Output: Cluster assignment \mathbf{H} ;

- 1: Client initialization: Initialize $u_{ij}^{(m)} = \frac{1}{C}$ and $\mathbf{V}^{(m)}$ randomly.
- 2: Server initialization: Initialize $\mathcal{Q} = \mathcal{U}$ and \mathcal{G} to be all zero.
- 3: **while** not converged **do**
- 4: **for** $m = 1$ to M **do**
- 5: **▷ On m -th client C_m**
- 6: Update membership matrix $\mathbf{U}^{(m)}$ according to (14);
- 7: Update cluster centroids matrix $\mathbf{V}^{(m)}$ according to (19);
- 8: Send updated $\mathbf{U}^{(m)}$ to the global server;
- 9: **end for**
- 10: **▷ On Server \mathcal{S}**
- 11: Stack all $\mathbf{U}^{(m)}$ to construct tensor \mathcal{U} as shown in Fig. 1;
- 12: Update \mathcal{G} with new \mathcal{U} according to (16);
- 13: Update \mathcal{Q} according to $\mathcal{Q} = \mathcal{Q} + \mu(\mathcal{U} - \mathcal{G})$;
- 14: Send $\mathbf{G}^{(m)}$ and $\mathbf{Q}^{(m)}$ to C_m ;
- 15: **end while**
- 16: \mathcal{S} aggregates $\mathbf{U}^{(m)}$ into cluster assignment matrix \mathbf{H} according to (20);
- 17: **return** \mathbf{H}

the overall computational complexity of updating local model $ITER_L$ iterations is $\mathcal{O}\left(ITER_L NC\left(d_i^{(m)} + C\right)\right)$.

Global client-side: The computational load on the global server includes the update of \mathcal{G} and \mathcal{Q} . In each global iteration, the computational complexity of the update of \mathcal{G} is $\mathcal{O}(2NCM \log(CM) + NCM^2)$, while the computational complexity of the update of \mathcal{Q} is $\mathcal{O}(NCM)$. Thus, the overall computational complexity of updated global model $ITER_G$ iterations is $\mathcal{O}\left(ITER_G(2NCM \log(CM) + NCM^2)\right)$.

4.2.2 Communication Complexity. Moreover, we analyze the transmission load between each client and the global server.

Local Client \rightarrow Global Server: Each client transmits its local $\mathbf{U}^{(m)}$ to the global server before aggregating. For all clients, the amount of data transmitted in each communication round is $\mathcal{O}(NCM)$.

Global Server \rightarrow Local Client: The global server sends the m -th slice of \mathcal{G} and \mathcal{Q} to the m -th client after aggregating. Thus, the total amount of data transmitted in each communication round is $\mathcal{O}(2NCM)$.

5 EXPERIMENT

5.1 Experimental Settings

We validate our model on eight multi-view datasets and compare it with several state-of-the-art methods. We implement all the methods with MATLAB2023b on a Desktop running Windows 11 equipped with Intel Core i5-13400CPU(2.50GHz) and 32GB DDR4 RAM. For the federated setting, our experiment consists of the same number of clients as the number of views and one server, while each client stores the data with one view. We use three widely used metrics, i.e., Accuracy (ACC), Normalized Mutual Information (NMI), and Purity (PUR) to evaluate our method.

Datasets	3-sources			BBCSport			Sonar			Caltech-5v		
Metrics	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR
DiMSC	70.81	63.81	<u>76.13</u>	82.17	64.07	82.17	56.41	1.63	56.41	57.57	39.76	61.43
MvLRSSC	54.67	44.92	63.31	64.07	40.92	65.07	50.48	0.01	53.37	46.15	34.81	46.79
RMSL	34.91	14.43	42.60	76.63	72.36	76.63	50.48	1.76	53.37	55.00	52.18	59.07
GMC	69.23	62.16	74.56	80.70	76.00	79.43	50.48	4.50	53.37	34.07	48.40	36.07
MvDGNMF	66.27	48.77	70.41	<u>85.11</u>	70.07	<u>85.11</u>	63.94	6.00	63.94	49.57	38.24	53.86
UDBGL	34.91	5.60	35.50	36.40	2.43	36.58	57.21	1.61	57.21	31.80	23.54	19.28
FastMICE	52.07	45.57	66.27	50.00	25.49	55.33	58.65	3.63	58.65	<u>77.58</u>	69.6	<u>79.57</u>
FedMVL	56.21	45.88	68.05	62.13	42.28	71.14	<u>64.90</u>	8.71	<u>64.90</u>	<u>29.71</u>	11.20	<u>30.36</u>
FedFuzzy	80.47	66.09	80.47	87.15	82.91	87.15	72.60	17.05	72.60	80.83	<u>64.51</u>	81.00

Table 1: Clustering performance comparison in terms of ACC(%), NMI(%), and PUR(%) on 3-sources, BBCSport, Sonar and Caltech-5V datasets.

Datasets	Yale			Vehicle Sensor			HAR			RGBD		
Metrics	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR
DiMSC	48.28	51.85	49.09	<u>76.06</u>	<u>29.47</u>	<u>76.06</u>	51.79	32.14	25.69	40.72	32.57	50.10
MvLRSSC	45.85	50.16	46.97	56.78	6.12	56.78	49.38	53.56	53.40	<u>43.95</u>	37.29	43.29
RMSL	<u>67.27</u>	74.02	68.48	68.07	12.34	68.07	48.64	52.99	55.38	13.80	3.06	26.43
GMC	54.55	62.44	54.55	64.68	19.55	64.68	48.04	57.40	48.60	40.23	33.06	46.51
MvDGNMF	47.27	52.24	50.91	52.63	0.20	52.63	46.36	35.21	46.36	26.57	0.78	26.98
UDBGL	52.73	65.94	54.55	51.69	0.08	51.69	47.78	46.20	50.45	43.89	<u>35.96</u>	53.55
FastMICE	66.68	67.94	67.88	51.76	0.09	51.76	<u>56.66</u>	50.17	<u>56.71</u>	41.81	32.61	49.53
FedMVL	46.67	51.50	47.27	74.03	17.39	74.03	<u>53.68</u>	<u>54.70</u>	43.71	32.51	23.65	45.89
FFCMSP	67.64	<u>73.67</u>	73.91	99.94	99.31	99.94	59.54	<u>53.22</u>	59.54	46.71	27.13	<u>51.84</u>

Table 2: Clustering performance comparison in terms of ACC(%), NMI(%), and PUR(%) on Yale, Vehicle Sensor, HAR, and RGBD datasets.

Datasets: We evaluate our method on eight public multi-view datasets, Concretely: (1)**3-sources** is a three-view text dataset and the three views are sourced from three reputable news outlets: BBC, Reuters, and The Guardian while their dimensions are 3056, 3631, and 3068. 169 samples are selected to form the datasets. (2)**BBCSport** [16] is composed of 544 sports news articles sourced from the BBC Sport website spanning the years 2004-2005. The dataset is categorized into five distinct topical areas with the following class labels: athletics, cricket, football, rugby, and tennis. It has two views and the dimensions are 3283 and 3183 respectively (3)**Sonar** [37] extracts its multi-view features from the 111 patterns obtained by bouncing sonar signals off a metal cylinder at various angles and under various conditions and 97 patterns obtained from rocks under similar conditions. Then the 60 features are divided into three views equally. (4)**Caltech-5v** [13] is the five-view version of Caltech-7, which has 7 classes and 1400 sample. The dimensions of Caltech-5v are [40, 254, 1984, 512, 1400]. (5)**Yale** is a two-view dataset of 165 facial images of 11 people, and the first view is a 32×32 image, while the second view is a 64×64 image. (6)**Vehicle Sensor** [9] is a four-view dataset and the four features are gathered from distributed sensors while their dimensions are 5, 5, 7, and 5. (7)**Human Activity Recognition(HAR)** [36] is a dataset documenting 30 individuals performing six daily activities(walking, walking upstairs, walking downstairs,

sitting, standing, lying down), and the features are gathered from Samsung Galaxy S II on the waist. The four-view dataset consists of 10299 samples. The feature dimensions of the four views are 20, 65, 237, and 59. (8)**SentencesNYU v2(RGB-D)** [38] is about images of indoor scenes and corresponding descriptions. We preprocess the dataset by following [40] and obtained a two-view dataset of 1449 samples divided into 13 classes, and the feature dimensions are 2048, and 300.

Compared Methods: We compared our methods with five centralized multi-view clustering methods and two federated multi-view clustering to verify the superiority of our proposed model. Concretely, (1)**DiMSC** [3] boosts the multi-view clustering by mining the complementary information among multi-view features. (2)**MvLRSSC** [2] constructs an affinity matrix shared among all views to learn a joint subspace representation. (3)**RMSL** [26] is composed of Hierarchical Self-Representative Layers and Backward Encoding Networks, which recover the subspace structure of data and explore the complex relationships among different views respectively. (4)**GMC** [42] is a graph-based multi-view clustering method that learns a unified graph from the affinity graph of each view for clustering. (5)**MvDGNMF** [25] is an NMF-based method that can extract more abstract representation by constructing a multi-layer NMF model with graph Laplacian regularizer. (6) **UDBGL** [11]

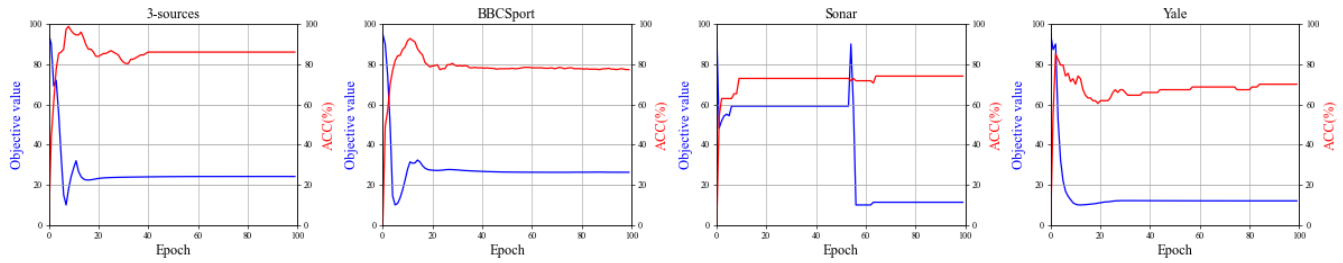


Figure 2: The convergence curves of FFCMSP on 3-sources, BBCSport, Sonar, and Yale.

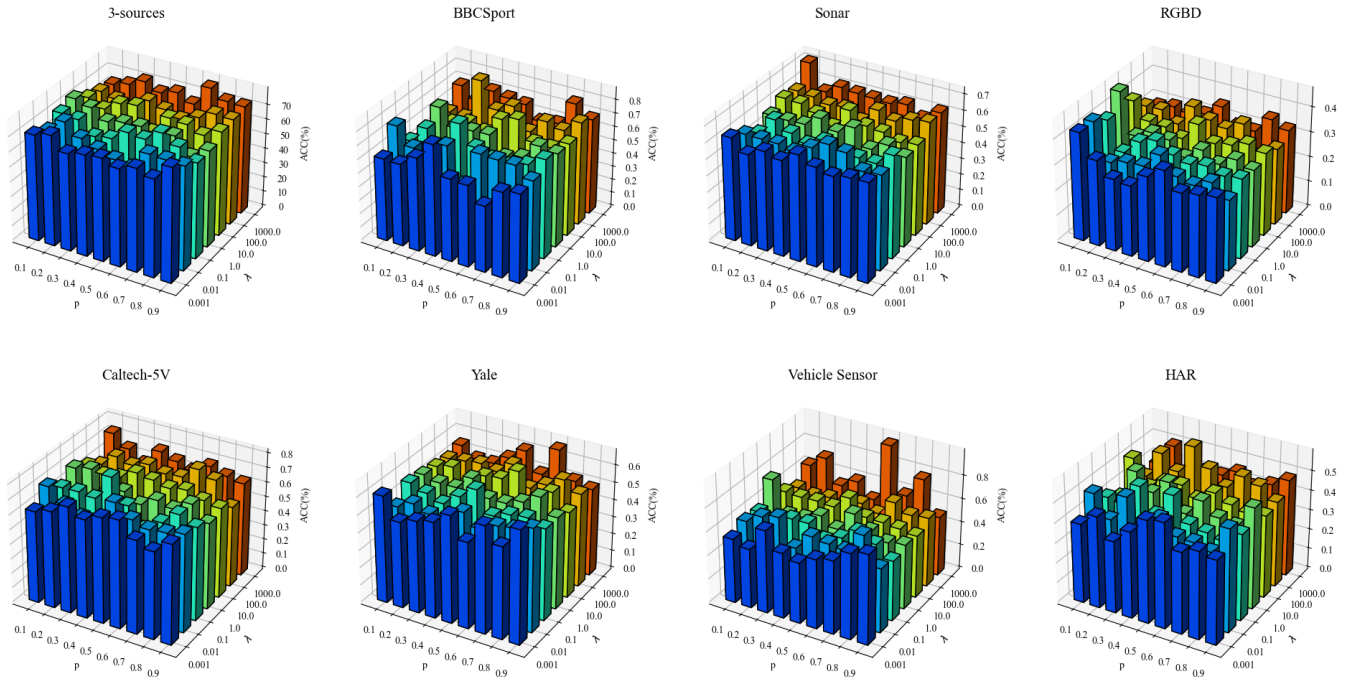


Figure 3: Parameter sensitivity analysis concerning p and λ .

jointly learns both single-view and consensus graphs and combines it with k-means to construct a unified framework. (7) **FastMICE** [18] presents the concept of a random view group to capture the versatile view-wise relationships and design a hybrid early-late fusion strategy. The method has linear time and space complexity. (8) **FedMVL** [19] is a federated multi-view clustering based on NMF and K-means.

5.2 Experiment Results and Analysis

Table. 1 and **Table. 2** present the experiment results of the proposed method and eight other multi-view clustering methods on eight real-world datasets. From the results we can observe that on ACC, NMI, and PUR, FFCMSP demonstrates superior performance across almost all datasets, demonstrating its robust clustering capability. Additionally, on the sonar dataset, it outperforms the second-best model by 7.70%, 8.34%, and 7.70%, respectively. Particularly, its performance on the Vehicle Sensor dataset, where its ACC, NMI, and PUR metrics nearly reach 100%, far surpassing other algorithms.

This could be attributed to the proposed FFCMSP effectively capturing the consistency and complementary information between views using tensor Schatten p -norm while leveraging FCM to better preserve information between data points. Besides, in the eight chosen compared method, FedMVL is the only federated multi-view clustering method, thus, the experiment results further demonstrate the superiority of FFCMSP over centralized ones. Overall, these experimental results validate the superiority of FFCMSP in multi-view clustering tasks.

5.3 Convergence analysis

We record the values of $\sum_{m=1}^M \sum_{i=1}^N \sum_{j=1}^C u_{ij}^{(m)} \|x_i^{(m)} - v_j^{(m)}\|_2^2 + \lambda \|U\|_{sp}^p$ and the ACC at each iteration on four datasets. The results are also displayed in **Fig 2**, from which we observed that FFCMSP converges quickly on these four datasets, typically reaching convergence within 20 iterations. Moreover, if the early-stop trick is

697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754

755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812

Variants	<i>case 1</i>			<i>case 2</i>			FFCMSP		
Dataset	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR
3-sources	61.54	34.81	61.54	<u>69.82</u>	<u>53.38</u>	<u>69.82</u>	80.47	66.09	80.47
BBCSport	41.73	9.45	41.73	<u>66.54</u>	<u>37.90</u>	<u>66.54</u>	87.15	82.91	87.15
Sonar	54.33	0.58	54.33	<u>66.35</u>	<u>7.75</u>	<u>66.35</u>	72.60	17.05	72.60
Caltech-5v	46.36	30.07	51.71	<u>60.07</u>	<u>49.78</u>	<u>60.07</u>	80.83	64.51	81.00
Yale	38.79	<u>48.83</u>	38.79	<u>39.39</u>	40.14	<u>39.39</u>	67.64	73.67	73.91
Vehicle	52.07	16.50	52.07	<u>74.53</u>	<u>20.56</u>	<u>74.53</u>	99.94	99.31	99.94
HAR	52.40	<u>59.70</u>	52.40	52.40	35.21	<u>54.16</u>	59.54	<u>53.22</u>	59.54
RGBD	36.71	<u>22.54</u>	39.82	<u>38.16</u>	<u>22.84</u>	<u>42.93</u>	46.71	27.13	51.84

Table 3: Results of ablation studies.

employed, the clustering accuracy of FFCMSP would further improve (The experimental results documented in this paper are all based on converged results). We can still observe that neither the convergence curves of the objective function nor the ACC exhibit oscillations, which typically arise from federated aggregation. This indicates the robustness of our model to federated settings.

5.4 Parameter Sensitivity Analysis

The objective function Eq. (9) has two main parameters: p and λ . We analyze their impact on the clustering results, which are depicted in Fig. 3. Firstly, we analyze the influence of two parameters. p is the most significant parameter in the tensor Schatten- p norm. When $0 \leq p \leq 1$, the tensor Schatten p -norm has better rank approximation. λ is the trade-off parameter to balance the effects of the two items. When λ is large, the tensor Schatten p -norm dominates, leading the model to explore more spatial structural information between views, while when λ is small, the model focuses on local training. From Fig. 3, we conclude that the chosen for p is closely related to the dataset, but when λ is set to a larger value, the model performs better. Above all, the suggested range of λ is [100, 1000].

5.5 Ablation Experiments

Our proposed FFCMSP consists of two main modules: The fuzzy C-Means module and the tensor Schatten p -norm. As the tensor \mathcal{U} is constructed from the all view-specific membership degree matrix \mathbf{U} , thus the tensor Schatten p -norm cannot exist independently. Consequently, we conduct the ablation study in the following two cases: (1) FFCMSP w/o tensor Schatten- p regularizer and fuzzy C-Means (*case 1*); (2) FFCMSP w/o tensor Schatten p -norm regularizer (*case 2*); It is obvious that *case 1* is equal to multi-view K-means, and the results are shown in Table. 3. From the results, we observe that on 3-sources, the ACC, NMI, and PUR of *case 2* outperform that in *case 1* by 7.28%, 18.57%, and 7.28 %, demonstrating that introducing membership degree to clustering surely enhances clustering performance. Moreover, the ACC, NMI, and PUR of FFCMSP on 3-sources also increase by 10.65%, 12.71%, and 10.65% compared with *case 2*, indicating that introducing the tensor Schatten p -norm can better explore spatial structural information between views. In summary, the ablation experiments strongly support the significant role of the two key modules in enhancing clustering performance.

6 CONCLUSION

In this paper, we proposed a novel fuzzy C-Means and Schatten p -norm based federated multi-view clustering method named FFCMSP. This method integrates fuzzy C-Means clustering and Schatten p -norm regularizer to enhance the clustering performance of multi-view data in a federated learning setting. Specifically, by employing fuzzy C-Means, we effectively alleviate the information loss caused by K-means, while the Schatten p -norm helps to exploit the inter-view complementary information and spatial structure, thereby improving the accuracy and robustness of clustering. Besides, the proposed federated optimization algorithm enables clients to train the global clustering model collaboratively. Furthermore, we evaluated the FFCMSP method on multiple real-world datasets and compared it with several state-of-the-art multi-view clustering methods. The experimental results demonstrate that our method generally achieves superior clustering performance.

REFERENCES

- [1] James C Bezdek, Robert Ehrlich, and William Full. 1984. FCM: The fuzzy c-means clustering algorithm. *Computers & geosciences* 10, 2-3 (1984), 191–203.
- [2] Maria Brbić and Ivica Kopriva. 2018. Multi-view low-rank sparse subspace clustering. *Pattern Recognition* 73 (2018), 247–258.
- [3] Xiaochun Cao, Changqing Zhang, Huazhu Fu, Si Liu, and Hua Zhang. 2015. Diversity-induced multi-view subspace clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 586–594.
- [4] Sicong Che, Zhaoming Kong, Hao Peng, Lichao Sun, Alex Leow, Yong Chen, and Lifang He. 2022. Federated multi-view learning for private medical data integration and analysis. *ACM Transactions on Intelligent Systems and Technology (TIST)* 13, 4 (2022), 1–23.
- [5] Man-Sheng Chen, Ling Huang, Chang-Dong Wang, and Dong Huang. 2020. Multi-view clustering in latent embedding space. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 34. 3513–3520.
- [6] Tianyi Chen, Xiao Jin, Yuejiao Sun, and Wotao Yin. 2020. Vaf: a method of vertical asynchronous federated learning. *arXiv preprint arXiv:2007.06081* (2020).
- [7] Wenlan Chen, Hong Wang, and Cheng Liang. 2023. Deep multi-view contrastive learning for cancer subtype identification. *Briefings in Bioinformatics* 24, 5 (2023), bbad282.
- [8] Xinyue Chen, Jie Xu, Yazhou Ren, Xiaorong Pu, Ce Zhu, Xiaofeng Zhu, Zhifeng Hao, and Lifang He. 2023. Federated Deep Multi-View Clustering with Global Self-Supervision. In *Proceedings of the 31st ACM International Conference on Multimedia*. 3498–3506.
- [9] Marco F Duarte and Yu Hen Hu. 2004. Vehicle classification in distributed sensor networks. *J. Parallel and Distrib. Comput.* 64, 7 (2004), 826–838.
- [10] Ali Mamdouh Elkahky, Yang Song, and Xiaodong He. 2015. A multi-view deep learning approach for cross domain user modeling in recommendation systems. In *Proceedings of the 24th international conference on world wide web*. 278–288.
- [11] Si-Guo Fang, Dong Huang, Xiao-Sha Cai, Chang-Dong Wang, Chaobo He, and Yong Tang. 2023. Efficient multi-view clustering via unified and discrete bipartite graph learning. *IEEE Transactions on Neural Networks and Learning Systems*

- (2023).
- [12] Uno Fang, Man Li, Jianxin Li, Longxiang Gao, Tao Jia, and Yanchun Zhang. 2023. A Comprehensive Survey on Multi-view Clustering. *IEEE Transactions on Knowledge and Data Engineering* (2023).
- [13] Li Fei-Fei, Rob Fergus, and Pietro Perona. 2004. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. In *2004 conference on computer vision and pattern recognition workshop*. IEEE, 178–178.
- [14] Dashan Gao, Ce Ju, Xiguang Wei, Yang Liu, Tianjian Chen, and Qiang Yang. 2019. Hhhl: Hierarchical heterogeneous horizontal federated learning for electroencephalography. *arXiv preprint arXiv:1909.05784* (2019).
- [15] Quanxue Gao, Pu Zhang, Wei Xia, Deyan Xie, Xinbo Gao, and Dacheng Tao. 2020. Enhanced tensor RPCA and its application. *IEEE transactions on pattern analysis and machine intelligence* 43, 6 (2020), 2133–2140.
- [16] Derek Greene and Pádraig Cunningham. 2006. Practical solutions to the problem of diagonal dominance in kernel document clustering. In *Proceedings of the 23rd international conference on Machine learning*. 377–384.
- [17] Xingchen Hu, Jindong Qin, Yinghua Shen, Witold Pedrycz, Xinwang Liu, and Jiyuan Liu. 2023. An Efficient Federated Multi-view Fuzzy C-Means Clustering Method. *IEEE Transactions on Fuzzy Systems* (2023).
- [18] Dong Huang, Chang-Dong Wang, and Jian-Huang Lai. 2023. Fast multi-view clustering via ensembles: Towards scalability, superiority, and simplicity. *IEEE Transactions on Knowledge and Data Engineering* (2023).
- [19] Shudong Huang, Wei Shi, Zenglin Xu, Ivor W Tsang, and Jiancheng Lv. 2022. Efficient federated multi-view learning. *Pattern Recognition* 131 (2022), 108817.
- [20] Shudong Huang, Ivor W Tsang, Zenglin Xu, and Jiancheng Lv. 2021. Measuring diversity in graph learning: A unified framework for structured multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering* 34, 12 (2021), 5869–5883.
- [21] Yutao Huang, Lingyang Chu, Zirui Zhou, Lanjun Wang, Jiangchuan Liu, Jian Pei, and Yong Zhang. 2021. Personalized cross-silo federated learning on non-iid data. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 35. 7865–7873.
- [22] Yi Jin, Chenning Li, Yidong Li, Peixi Peng, and George A Giannopoulos. 2021. Model latent views with multi-center metric learning for vehicle re-identification. *IEEE Transactions on Intelligent Transportation Systems* 22, 3 (2021), 1919–1931.
- [23] Jing Li, Quanxue Gao, Qianqian Wang, Ming Yang, and Wei Xia. 2024. Orthogonal non-negative tensor factorization based multi-view clustering. *Advances in Neural Information Processing Systems* 36 (2024).
- [24] Jiangmeng Li, Wenwen Qiang, Changwen Zheng, and Bing Su. 2022. RHMC: Modeling consistent information from deep multiple views via Regularized and Hybrid Multiview Coding. *Knowledge-Based Systems* 241 (2022), 108201.
- [25] Jianqiang Li, Guoxu Zhou, Yuning Qiu, Yanjiao Wang, Yu Zhang, and Shengli Xie. 2020. Deep graph regularized non-negative matrix factorization for multi-view clustering. *Neurocomputing* 390 (2020), 108–116.
- [26] Ruihuang Li, Changqing Zhang, Huazhu Fu, Xi Peng, Tianyi Zhou, and Qinghua Hu. 2019. Reciprocal multi-layer subspace learning for multi-view clustering. In *Proceedings of the IEEE/CVF international conference on computer vision*. 8172–8180.
- [27] Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. 2020. Federated optimization in heterogeneous networks. *Proceedings of Machine learning and systems* 2 (2020), 429–450.
- [28] Xuelong Li, Han Zhang, Rong Wang, and Feiping Nie. 2020. Multiview clustering: A scalable and parameter-free bipartite graph fusion method. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 1 (2020), 330–344.
- [29] Naiyao Liang, Zuyuan Yang, Zhenni Li, Weijun Sun, and Shengli Xie. 2020. Multi-view clustering by non-negative matrix factorization with co-orthogonal constraints. *Knowledge-Based Systems* 194 (2020), 105582.
- [30] Jialu Liu, Chi Wang, Jing Gao, and Jiawei Han. 2013. Multi-view clustering via joint nonnegative matrix factorization. In *Proceedings of the 2013 SIAM international conference on data mining*. SIAM, 252–260.
- [31] Yang Liu, Yan Kang, Chaoping Xing, Tianjian Chen, and Qiang Yang. 2020. A secure federated transfer learning framework. *IEEE Intelligent Systems* 35, 4 (2020), 70–82.
- [32] Yongli Liu, Xiaoqin Zhang, Guiying Tang, and Di Wang. 2019. Multi-view subspace clustering based on tensor Schatten-p norm. In *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 5048–5055.
- [33] Han Lu, Quanxue Gao, Qianqian Wang, Ming Yang, and Wei Xia. 2023. Centerless multi-view K-means based on the adjacency matrix. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37. 8949–8956.
- [34] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguerre y Arcas. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*. PMLR, 1273–1282.
- [35] Xi Peng, Zhenyu Huang, Jiancheng Lv, Hongyuan Zhu, and Joey Tianyi Zhou. 2019. COMIC: Multi-view clustering without parameter selection. In *International conference on machine learning*. PMLR, 5092–5101.
- [36] Anguita Davide Ghio Alessandro Oneto Luca Reyes-Ortiz, Jorge and Xavier Parra. 2012. Human Activity Recognition Using Smartphones. UCI Machine Learning Repository. DOI: <https://doi.org/10.24432/C54S4K>.
- [37] Terry Sejnowski and R. Gorman. [n.d.]. Connectionist Bench (Sonar, Mines vs. Rocks). UCI Machine Learning Repository. DOI: <https://doi.org/10.24432/C5T01Q>.
- [38] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. 2012. Indoor segmentation and support inference from rgb-d images. In *Computer Vision—ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7–13, 2012, Proceedings, Part V 12*. Springer, 746–760.
- [39] Jiankai Sun, Xin Yang, Yuanshun Yao, Anon Zhang, Weihao Gao, Junyuan Xie, and Chong Wang. 2021. Vertical federated learning without revealing intersection membership. *arXiv preprint arXiv:2106.05508* (2021).
- [40] Daniel J Trosten, Sigurd Lokse, Robert Jenssen, and Michael Kampffmeyer. 2021. Reconsidering representation alignment for multi-view clustering. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 1255–1265.
- [41] Huibing Wang, Guangqi Jiang, Jinjia Peng, Ruoxi Deng, and Xianping Fu. 2022. Towards adaptive consensus graph: multi-view clustering via graph collaboration. *IEEE Transactions on Multimedia* (2022).
- [42] Hao Wang, Yan Yang, and Bing Liu. 2019. GMC: Graph-based multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering* 32, 6 (2019), 1116–1129.
- [43] Jie Xu, Yazhou Ren, Guofeng Li, Lili Pan, Ce Zhu, and Zenglin Xu. 2021. Deep embedded multi-view clustering with collaborative training. *Information Sciences* 573 (2021), 279–290.
- [44] Yaming Yang, Ziyu Guan, Jianxin Li, Wei Zhao, Jiangtao Cui, and Quan Wang. 2021. Interpretable and efficient heterogeneous graph convolutional network. *IEEE Transactions on Knowledge and Data Engineering* 35, 2 (2021), 1637–1650.
- [45] Yaming Yang, Ziyu Guan, Wei Zhao, Weigang Lu, and Bo Zong. 2022. Graph substructure assembling network with soft sequence and context attention. *IEEE Transactions on Knowledge and Data Engineering* 35, 5 (2022), 4894–4907.
- [46] Changqing Zhang, Huazhu Fu, Qinghua Hu, Xiaochun Cao, Yuan Xie, Dacheng Tao, and Dong Xu. 2018. Generalized latent multi-view subspace clustering. *IEEE transactions on pattern analysis and machine intelligence* 42, 1 (2018), 86–99.
- [47] Jie Zhao, Xinghua Zhu, Jianzong Wang, and Jing Xiao. 2021. Efficient client contribution evaluation for horizontal federated learning. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 3060–3064.
- [48] Yujiao Zhao, Yu Yun, Xiangdong Zhang, Qin Li, and Quanxue Gao. 2022. Multi-view spectral clustering with adaptive graph learning and tensor Schatten p -norm. *Neurocomputing* 468 (2022), 257–264.