

Joint Multi-View Collaborative Clustering

Yasser KHALAFAOUI*

R&D department

ALTECA

Massy, France

mykhalafaoui@alteca.fr

Basarab MATEI

LIPN - CNRS UMR 7030

Sorbonne Paris Nord University

Villetaneuse, France

matei@lipn.univ-paris13.fr

Nistor GROZAVU

ETIS - CNRS UMR 8051

CY Cergy Paris University

Cergy, France

nistor.grozavu@cyu.fr

Martino LOVISETTO

R&D department

ALTECA

Lyon, France

mlovisetto@alteca.fr

Abstract—Data is increasingly being collected from multiple sources and described by multiple views. These multi-view data provide richer information than traditional single-view data. Fusing the former for specific tasks is an essential component of multi-view clustering. Since the goal of multi-view clustering algorithms is to discover the common latent structure shared by multiple views, the majority of proposed solutions overlook the advantages of incorporating knowledge derived from horizontal collaboration between multi-view data and the final consensus. To fill this gap, we propose the Joint Multi-View Collaborative Clustering (JMVCC) solution, which involves the generation of basic partitions using Non-negative Matrix Factorization (NMF) and the horizontal collaboration principle, followed by the fusion of these local partitions using ensemble clustering. Furthermore, we propose a weighting method to reduce the risk of negative collaboration (i.e., views with low quality) during the generation and fusion of local partitions. The experimental results, which were obtained using a variety of data sets, demonstrate that JMVCC outperforms other multi-view clustering algorithms and is robust to noisy views.

Index Terms—Multi-view clustering, collaborative clustering, ensemble clustering, non-negative matrix factorization

I. INTRODUCTION

Data collections became very diverse in the recent decade [1] as a result of the advent of multi-modal data sets, multi-view data sets (i.e. the same data sample described in different ways), and distributed data. For example, captions and textual tags are frequently used to describe images, and some news are covered by several media organisations and in multiple languages. Given that these different representations frequently provide compatible and complimentary information [2], it is essential to extract intrinsic information from these multi-view data sets rather than focusing on standard single-views. One of the most important aspects of learning from multi-view data sets is to use the extracted knowledge from each view to deal with outliers and noisy characteristics [3].

The task of finding intrinsic information across multiple views, known as multi-view clustering, has received a lot of attention [4]–[6]. The objective of multi-view clustering is to leverage heterogeneous information from different views to create a high quality clustering result. There are four types of multi-view clustering approaches in the literature. *Multi-view graph clustering* is a set of methods that finds a fusion graph

across all views and then applies clustering techniques to the fusion graph to produce the clustering result. Wang et al. [7] created a nearest neighbor graph to find an encoding for the associated manifold information. Then, to determine the best intrinsic manifold, they built a multiple graph ensemble regularization framework. Cross diffused matrix alignment based on feature selection is a technique that Wei et al. [8] proposed for choosing features for each view while doing alignment on a cross diffused matrix. The final clustering results were then obtained via co-regularized spectral clustering on these chosen features [9]. *Multi-kernel learning* uses predetermined kernels related to several views that are subsequently combined either linearly or non-linearly to enhance clustering performance [10]. Based on the minimizing-disagreement methodology, Su et al. [11] developed a novel kernel combination method. They created a multi-partite graph to infer a kernel, which was then applied to spectral clustering. *Multi-task multi-view clustering* enhance clustering performance by assigning one or more tasks to each view, distributing inter-task knowledge, and making use of connections between many tasks and views. Wang et al. [12] investigated multi-view spectral clustering using a multi-objective framework (viewed as a multi-task), which is addressed by Pareto optimization. *Collaborative clustering* is a co-training strategy used to handle multi-view data. It bootstraps the clustering of multiple views by leveraging the knowledge extracted from each view [13]. When this procedure is applied iteratively, the clustering results of all views tend to converge. Jiang et al. [14] established a co-regularized Probabilistic Latent Semantic Analysis (PLSA) model for multi-view co-clustering.

However, some views with irrelevant or noisy features may negatively impact the common space and decrease clustering quality. Furthermore, as the number of views grows, there is less common area shared by all views.

In recent years, Non-negative Matrix Factorization has gained a lot of attention and has been applied in a range of fields such as feature selection, dimensionality reduction, text mining, and clustering [15], [16]. Paatero developed the NMF technique [17], an unsupervised clustering algorithm that factorizes data matrices into two matrices by imposing non-negativity constraints on the components, one indicating the data partitions and the other representing the data set's cluster prototypes. The absence of negative values simplifies the interpretation of the generated matrices. For example,

* Corresponding author

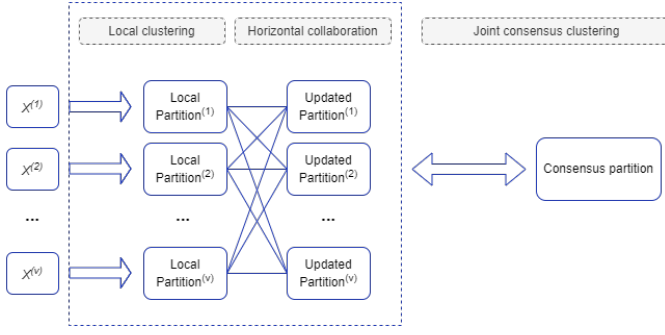


Fig. 1. Proposed pipeline for the joint multi-view collaborative clustering

Brunet et al. [18] used NMF to identify distinctive molecular patterns. They demonstrated that NMF is more effective than other methods in obtaining biologically important information from cancer-related microarray data and is robust to initial settings.

The main difficulty in using NMF to multi-view clustering is determining how to restrict the search for factorizations to those that provide meaningful and comparable clustering solutions across several views at the same time. Furthermore, the NMF technique is non-convex algorithm and lacks solution unicity, which implies that if we run NMF on the same data set many times, we get different factorizations each time [28].

Moreover, the existing studies on multi-view clustering either ignore the interaction between local partitions and the consensus partition or pay little attention to exchanging information between local partitions.

To address this gap, we set out to develop a novel approach, Joint Multi-View Collaborative Clustering (JMVCC), which allows information to be exchanged across multiple local partitions and seeks a consensus partition from these different views. Fig 1 shows the pipeline of the proposed solution. First, we generate local partitions with the NMF method and improve the clustering quality of each view during the horizontal collaboration phase. Then, we combine these improved local partitions to form a consensus partition. Finally, using an iterative approach, the consensus partition leads the update of local partitions.

JMVCC provides three advantages over other multi-view clustering techniques:

- The horizontal collaboration allows the local partitions to exchange information and enhance the quality of the clustering.
- The local and consensus partitions are mutually updated iteratively.
- During the fusion and horizontal collaboration phases, a weighting strategy is used to mitigate the risk of negative collaboration (i.e., collaboration between views with greater clustering quality and views with lower quality reduces the quality of the initial views), making the solution more robust to noisy views.

The rest of the paper is structured as follows: Section II goes through the preliminary setup as well as the formal

definition of our solution JMVCC. Section III presents the solution optimization under various conditions. Finally, in Section IV, we evaluate the proposed algorithm's performance using various datasets. The paper ends with a conclusion and a discussion of several future works.

II. JOINT MULTI-VIEW COLLABORATIVE CLUSTERING

In this part, we introduce the NMF algorithm, followed by the formulation of our solution JMVCC and the subsequent functions.

A. NMF algorithm

Given a set of non-negative data matrices $X = \{X^1, X^2, \dots, X^V\}$, such that $X^v = (x_1^v, x_2^v, \dots, x_N^v) \in \mathbb{R}_+^{M \times N}$ is the data matrix corresponding to the v -th view, where $x_n^v \in \mathbb{R}_+^{M \times 1}$ represents the n^{th} object of X^v , $v \in \mathbb{N}$.

The standard NMF algorithm factorizes the matrix X^v into two low rank matrices F^v and G^v , such that F^v is the matrix of cluster centroids and G^v is the matrix of data partitions defined respectively as $F^v = (f_1^v, f_2^v, \dots, f_k^v) \in \mathbb{R}_+^{M \times K}$, and $G^v = (g_1^v, g_2^v, \dots, g_N^v) \in \mathbb{R}_+^{K \times N}$, with K a parameter representing the number of components. One of the NMF algorithm's frequent reconstruction methods can be expressed as a Frobenius norm optimization problem, which is stated as:

$$\min_{F^v, G^v} \mathcal{L}_v = \|X^v - F^v G^v\|_F^2, \text{ s.t. } F^v \geq 0, G^v \geq 0 \quad (1)$$

where $\|\cdot\|_F$ represents the Frobenius norm and $F^v \geq 0, G^v \geq 0$ indicate the constraints that all matrices elements must be positive.

B. Objective Function

As previously stated, we want to use the horizontal collaboration method and the fusion of local partitions to enhance the clustering quality in a multi-view setting. To do so, we introduce the objective function of JMVCC, which is defined as follows:

$$\mathcal{J}(F^v, G^v, G^*) = \sum_{v=1}^V \left(\mathcal{L}_v(F^v, G^v) + \sum_{v' \neq v}^V \alpha_{v, v'} \mathcal{H}(v, v') \right) + \sum_v^V \beta_v \mathcal{U}_v(G^v, G^*).$$

The optimization problem can be written as:

$$(F^v, G^v, G^*) = \arg \min_{F^v, G^v, G^*} \mathcal{J}(F^v, G^v, G^*) \quad (2)$$

where $\mathcal{L}_v(F^v, G^v)$ is the standard NMF algorithm introduced in (1), $\mathcal{H}(v, v')$ is the multi-view collaboration term, $\mathcal{U}_v(G^v, G^*)$ is the category utility function [19] between the consensus partition G^* and the v -th local partition G^v , and $\alpha_{v, v'}$, β_v are the weights of the horizontal collaboration and fusion respectively.

The objective function combines consensus clustering with the generation and horizontal collaboration of local partitions to provide a one-step solution. In what follows, we define the horizontal collaboration term, the category utility function and the respective weights.

C. Horizontal collaboration term

A common collaborative strategy is to add information that has been retrieved from different views $v' \neq v$ to a view v [19], [20]. In order to incorporate the information from several views, Grozavu et al. [20] introduced a multi-view collaboration technique that reduces the distance between a data point and its corresponding centroid.

Similarly, we want to include the information extracted from different views during the generation of the local partitions. In order to achieve this, we define the multi-view collaboration term as:

$$\mathcal{H}_{v,v'}(F^v, G^v) = \|F^v(G^v - G^{v'})\|_F^2. \quad (3)$$

Our goal is to reduce the multi-view collaboration term using two data partition matrices of various views $G^{v'}$ and G^v . Notice that in (3) when G^v and $G^{v'}$ agree, the collaborative term equals zero and we consider only the v^{th} local NMF.

D. Category Utility Function

The category utility function is closely related to the contingency matrix, which measures the co-occurrence of two discrete random variables.

Given a consensus partition matrix $G^* \in \mathbb{R}_+^{K \times N}$ and a set of local partitions $G = \{G^{(1)}, G^{(2)}, \dots, G^{(V)}\}$, we define the category utility function between G^* and each G^v with $1 \leq v \leq V$ as follows:

$$\mathcal{U}(G^v, G^*) = \|G^v - G^*\|_F^2. \quad (4)$$

Equation (4) is simply the sum of the euclidean distances squared and is used to compute the discrepancy between the set of local partitions G and the consensus partition matrix G^* .

E. Weighting Strategy

In order to minimize the reconstruction error in the solution proposed, we introduce weights during the horizontal collaboration and fusion phases: we investigate how optimizing the weights can lead to an optimal value of the global objective function (2) and minimize the likelihood of negative collaboration by adjusting the weight factors during each iteration.

In (2), $\alpha_{v,v'}$ is the weight of the horizontal collaboration between the v -th view and the other views $v' \neq v$ and β_v is the weight used during the fusion phase. Additionally, the weights α and β are learnable parameters such that $\sum_{v' \neq v} \alpha_{v,v'}^\gamma = 1$ and $\sum_v \beta_v^\gamma = 1$, with $\gamma > 1$ a parameter that controls the weights distribution. Hence, important views (i.e., views with better clustering quality) will receive higher weights during the update.

III. OPTIMIZATION ALGORITHM

Given that the Frobenius norm is continuous and differentiable, it follows that the cost function given in (2) is also differentiable, and its derivative exists at every point in its domain. As a result, there is always a minimum. However, the NMF algorithm is non-convex in both F^v and G^v , therefore it is not possible to find the global minimum.

To overcome the problem of non-uniqueness of the NMF algorithm, we proposed the use of the aforementioned weighting strategy.

Moreover, to minimize the objective function defined in (2), we propose a convergent iterative update approach based on the gradient descent optimization technique. Specifically, we consider Lee and Seung' technique [21] and employ an adaptive learning rate to ensure that the non-negativity constraint is satisfied, as well as the multiplicative update rule for the centroid, partition and consensus matrices F , G and G^* during optimization. This implies that for $\Theta \in \{F^v, G^v, G^*\}$ s.t. $\Theta \geq 0$, the adaptive learning rate η_Θ is written as:

$$\eta_\Theta = \frac{\Theta}{[\nabla_\Theta \mathcal{J}]_+} \quad (5)$$

and the multiplicative update rule is defined as:

$$\Theta = \Theta \circ \frac{[\nabla_\Theta \mathcal{J}]_-}{[\nabla_\Theta \mathcal{J}]_+}. \quad (6)$$

Here, \mathcal{J} is the functional defined in (2), the fraction line and \circ represent the element-wise division and the Hadamard product respectively. $[\cdot]_+$ and $[\cdot]_-$ are the positive and negative terms of the gradient respectively.

A. Iterative Update Approach

The proposed iterative update technique is described in detail in the paragraphs that follow.

1) **Fixing G^* and F^v , minimize (2) over G^v** : When G^* and F^v are fixed, the optimization leads to the following update rule:

$$G^v = G^v \circ \frac{F^{vT} X^v + \sum_{v' \neq v} \alpha_{v,v'} (F^{vT} F^v G^{v'}) + \beta_v G^*}{F^{vT} F^v G^v + \sum_{v' \neq v} \alpha_{v,v'} (F^{vT} F^v G^v) + \beta_v G^v}. \quad (7)$$

2) **Fixing G^* and G^v , minimize (2) over F^v** : When G^* and G^v are fixed, we only consider the standard NMF and horizontal collaboration functions. The following updating rule results from the minimization:

$$F^v = F^v \circ \frac{X^v G^{vT} + \sum_{v' \neq v} \alpha_{v,v'} (F^v G^v G^{v'T} + G^{v'} G^{vT})}{F^v G^v G^{vT} + \sum_{v' \neq v} \alpha_{v,v'} (F^v G^v G^{vT} + G^{v'} G^{vT})}. \quad (8)$$

3) **Fixing G^v and F^v , minimize (2) over G^*** : Here, we only consider the consensus term in our objective function. The update rule for the consensus partition matrix G^* is defined as:

$$G^* = G^* \circ \frac{\sum_v \beta_v G^v}{\sum_v \beta_v G^*}. \quad (9)$$

Note that G^* is also positive by definition.

B. Update of the horizontal collaboration weight

Our goal is to minimize the objective function defined in (2) with respect to $\alpha_{v,v'}$. We follow the strategy proposed by Cai et al. [22] and only consider the horizontal collaboration term

$$\min_{\alpha_{v,v'}} \sum_{v' \neq v} \alpha_{v,v'} \mathcal{H}_{v,v'}, \quad \text{s.t.}, \quad \sum_{v' \neq v} \alpha_{v,v'}^\gamma = 1, \quad \alpha_{v,v'} \geq 0, \quad (10)$$

where $\mathcal{H}_{v,v'}$ is the collaboration function defined in (3). Taking into account the aforementioned constraints, we define the Lagrangian function of (10) as:

$$\sum_{v' \neq v} \alpha_{v,v'} \mathcal{H}_{v,v'} - \lambda_{v,v'} \left(\sum_{v' \neq v} \alpha_{v,v'}^\gamma - 1 \right) - \theta_{v,v'} \alpha_{v,v'} \quad (11)$$

where $\lambda_{v,v'}$ and $\theta_{v,v'}$ are Lagrange multipliers, with respect to v', v . By setting the derivative of (11) with respect to $\alpha_{v,v'}$ to zero, then substituting the resulting $\alpha_{v,v'}$ into the constraint $\sum_{v' \neq v} \alpha_{v,v'}^\gamma = 1$, we get:

$$\alpha_{v,v'} = \frac{(\mathcal{H}_{v,v'})^{\frac{1}{\gamma-1}}}{\sum_{v' \neq v} (\mathcal{H}_{v,v'})^{\frac{1}{\gamma-1}}}. \quad (12)$$

C. Update of the consensus weight

We present another method for optimizing the consensus weights based on the Karush-Kuhn-Tucker (KKT) conditions [23].

Since $\beta_v \geq 0$, our goal is to find the positive weights that will determine the contribution strength of each local partition G^v to the consensus partition G^* . Additionally, and following the update of the matrices F^v , G^v and G^* , we only consider the consensus function \mathcal{U}_v defined in (4) in order to optimize the consensus weights β_v . In what follows, we define ϕ as the consensus functional:

$$\begin{cases} \beta_v = \arg \min_{\beta} \phi = \arg \min_{\beta} \sum_v \beta_v \mathcal{U}_v(G^v, G^*) \\ \forall v \in V \quad \sum_v \beta_v^\gamma = 1 \text{ and } \beta_v \geq 0. \end{cases} \quad (13)$$

Using Lagrange multipliers λ_v and θ_v for the above system, we get the following KKT conditions, for $v \in V$:

$$\begin{cases} \beta_v \geq 0 \\ \sum_v \beta_v^\gamma = 1 \\ \lambda \cdot \beta_v = 0 \\ \nabla \phi_v - \lambda_v \nabla \beta_v + \theta_v \nabla (\beta_v^\gamma - 1) = 0. \end{cases} \quad (14)$$

Below are the optimization's results under the Karush-Kuhn-Tucker conditions:

$$\beta_v = \frac{(\mathcal{U}_v(G^v, G^*))^{\frac{1}{\gamma-1}}}{\sum_v (\mathcal{U}_v(G^v, G^*))^{\frac{1}{\gamma-1}}}. \quad (15)$$

Notice that all the horizontal collaboration and consensus terms receive equal weight factors when $\gamma \rightarrow \infty$. Moreover, we avoid the trivial solution of $\gamma \rightarrow 1$. Finally, the formulation in (12) and (15) enables us to regulate every weight factor with just one parameter γ .

The proposed solution is summarised in Alg. III-C.

Algorithm 1 JMVCC Algorithm

Input: Multi-view data set $X = \{X^1, X^2, \dots, X^v\}$, with $X^v \in \mathbb{R}_+^{M \times N}$ and v number of views
The number of components K
The parameter γ
Initialization: Randomly initialize the consensus matrix G^* and the matrices F^v and G^v
For all realizations
forall views v **do**
 Compute the optimized $\alpha_{v,v'}$ with (12).
 Compute the optimized β_v with (15).
 Update the partition matrices of all views (7).
 Update the centroid matrices of all views (8).
 Update the consensus matrix with (9)
end

IV. EXPERIMENTS

In this part, we evaluate the performance of our suggested joint multi-view collaborative clustering solution using three multi-view data sets, namely NUS-WIDE, Caltech101 and Handwritten. To better explain the premise of the proposed approach, more information about the data sets is provided. Moreover, Since the labels are made available for these data sets, we use two metrics to assess the results of our solution: the purity score and the Normalized Mutual Information (NMI).

A. Data Sets Description

- **NUS-WIDE**¹ - it includes 5,018 distinct tags and 269,648 images from Flickr, which correspond to 81 classes. Additionally, six views are provided: the color histogram, color correlogram, edge direction histogram, wavelet texture, block-wise color moments, and a bag of visual words on SIFT descriptions. In order to conduct our experiment, we extracted two subsets (NUS-2B and NUS-CDF); (see Tab.I).
- **Caltech101**² - it contains 101 categories represented by 8677 photos. We selected the seven most popular categories: Faces, Motorbikes, Dolla-Bill, Garfield, Snoopy, Stop-Sign, and Windsor Chair, which is equivalent to 1474 images. Following [24], we extract six views, namely Gabor features, Wavelet Moments (WM), CENTRIST, HOG, GIST, LBP. Tab. I gives more details about these views.
- **Handwritten**³ - it consists of 2000 data samples for ten digits (0 to 9) with 200 data samples per class/digit. We test our solution on two views. Specifically, we use the pixel averages in 2×3 windows (PIX) and the Fourier coefficients of the digit shapes (FOU).

¹<https://lms.comp.nus.edu.sg/wp-content/uploads/2019/research/nuswide/NUS-WIDE.html>

²<https://github.com/yeqinglee/mvdata>

³<https://mvlearn.github.io/references/datasets.html>

TABLE I
EXPERIMENTAL DATA SETS

Data sets	# Instances	# Clusters	# Features
NUS-CDF	3930	3	64, 144, 75, 128, 225, 500
NUS-2B	4701	2	64, 144, 75, 128, 225, 500
Caltech101-7	1474	7	48, 40, 254, 1984, 512, 928
Handwritten	2000	10	240, 74

B. Evaluation Metrics

Since the ground truth of the data sets is readily available, we evaluate our approach with the commonly-used external measures Purity score and NMI for cluster validity.

1) **Purity score**: it assesses the degree to which a cluster belongs to a class. Let $P = \{p_1, p_2, \dots, p_n\}, n \in \mathbb{N}$ and $K = \{k_1, k_2, \dots, k_m\}, m \in \mathbb{N}$ be the known data labels and centroids respectively. The purity score is defined as:

$$purity = \frac{1}{N} \sum_{m=1}^{|K|} \max_{i=1}^{|P|} |k_m \cap p_n|. \quad (16)$$

The clustering result's purity is equal to the expected purity of all clusters. A purity of 1 implies good clustering, while a purity of 0 implies poor clustering.

2) **Normalized Mutual Information**: NMI is a version of Mutual Information, a common measure in information theory. It is frequently regarded because of its ability to compare two partitions with varied numbers of clusters. It is defined as:

$$NMI(P, K) = \frac{2 \times I(P; K)}{H(P) + H(K)}, \quad (17)$$

where $I(P; K)$ is the mutual information between the ground truth P and the clustered set K . $H(P)$ and $H(K)$ represent the entropy.

C. Baseline Algorithms

We compare the proposed Joint Multi-View Collaborative Clustering approach against a number of baseline algorithms for multi-view and ensemble clustering to demonstrate its utility.

- **ColNMF** - it uses a shared coefficient matrix with various basis matrices across views to handle multi-view data [25], as seen below:

$$\sum_v \alpha_v \|X^v - F^v G^{*T}\|_F^2. \quad (18)$$

- **C-NMF** - introduced by Benlamine et al. [26], it is divided into two steps. The approach first learns a NMF for a fixed number of components K for each view, then computes a consensus solution for the various NMF models in the second step.
- **MultiNMF** - it seeks a factorization that produces compatible clustering solutions across many views [2]. It is done by developing a collaborative matrix factorization process with the constraint of pushing each clustering

solution towards a consensus rather than directly fixing it.

- **CMVC** - it generates several basic partitions from each view, followed by ensemble clustering to obtain the consensus partition; the later updates the basic partitions jointly and iteratively [27].

D. Results

The clustering performance of various methods across all four data sets is presented in Tab. II. 10 test runs with various random initializations were performed, and the results are provided along with the average performance and standard deviation.

Since the consensus matrix G^* is a continuous matrix, we discretize it by computing the predominant basis component. The row index for which the item is the maximum within the column is used to determine the dominating basis component.

In terms of purity/normalized mutual information, JMVCC exceeds the second best algorithm with a gap of 7.2%/5.8% on Caltech101, 1.6%/1.1% on Handwritten, 28.3%/24.8% on NUS-CDF and 12.9%/11.9% on NUS-2B.

On the NUS-CDF and NUS-2B data sets, JMVCC significantly outperforms the baseline approaches with an average of 20.6%/18.3%. This can be explained by the presence of noisy views in these data sets, which are assigned less weight than views with higher clustering quality, thus minimizing the impact of these unfavorable views. The proposed JMVCC can effectively find high-quality consensus clustering because it reduces the impact of noisy views by implementing a weighting strategy during both the horizontal collaboration and fusion steps. As a result, it confirms our claim that the proposed approach JMVCC is quite useful in clustering multi-view data sets with noisy views.

Both JMVCC and CMVC show promising clustering results on the Caltech101-7 and Handwritten digit data sets. JMVCC performs slightly better than CMVC and gets a 14.5%/19.8% performance advantage over the other algorithms. JMVCC improves its clustering solution by exchanging extracted knowledge from diverse views $v \in V$ during the horizontal collaboration phase.

E. Parameter study

If each individual view gives higher weights to views that have the most similar solutions to the local one in the setting of horizontal collaboration and fusion, the global results are expected to be better.

Further investigation reveals, in Fig. 3, that the degree to which one view v should collaborate with other views $v' \neq v$ with dissimilar partitions is determined by the parameter γ . If multiple views have similar clustering partitions, they will be given equal weight. The views with the most similar partitions would still be preferred to optimize the cost function of the proposed solution even while using a larger γ . However, views with less identical partitions would still be taken into account. In fact, the weight of the partitions from different views would increase as γ increases, eventually equaling the weight of all

TABLE II
CLUSTERING PERFORMANCE ON FOUR DATA SETS

Algorithms	Purity score (%)				NMI score (%)			
	Caltech101-7	Handwritten	NUS-CDF	NUS-2B	Caltech101-7	Handwritten	NUS-CDF	NUS-2B
ColNMF	60.7 \pm 2.0	51.9 \pm 1.3	34.1 \pm .0	39.2 \pm .05	49.1 \pm 2.6	50.3 \pm 1.6	21.9 \pm 0.8	38.1 \pm .2
C-NMF	70.8 \pm .05	64.8 \pm .01	51.9 \pm .08	67.6 \pm .02	70.7 \pm .0	50.7 \pm .02	43.2 \pm .1	64.3 \pm .02
Multi-NMF	58.9 \pm .01	74.2 \pm 2.1	46.3 \pm .0	52.7 \pm .0	56.1 \pm .1	71.6 \pm .01	44.1 \pm .0	45.0 \pm .0
CMVC	67.6 \pm .0	81.8 \pm .0	61.5 \pm .0	63.6 \pm .0	64.3 \pm .0	77.8 \pm .0	58.2 \pm .0	62.3 \pm .0
JMVCC	78.0 \pm .02	83.4 \pm .04	89.8 \pm .01	80.5 \pm .2	76.5 \pm .03	77.9 \pm .04	83.0 \pm .0	76.2 \pm .03

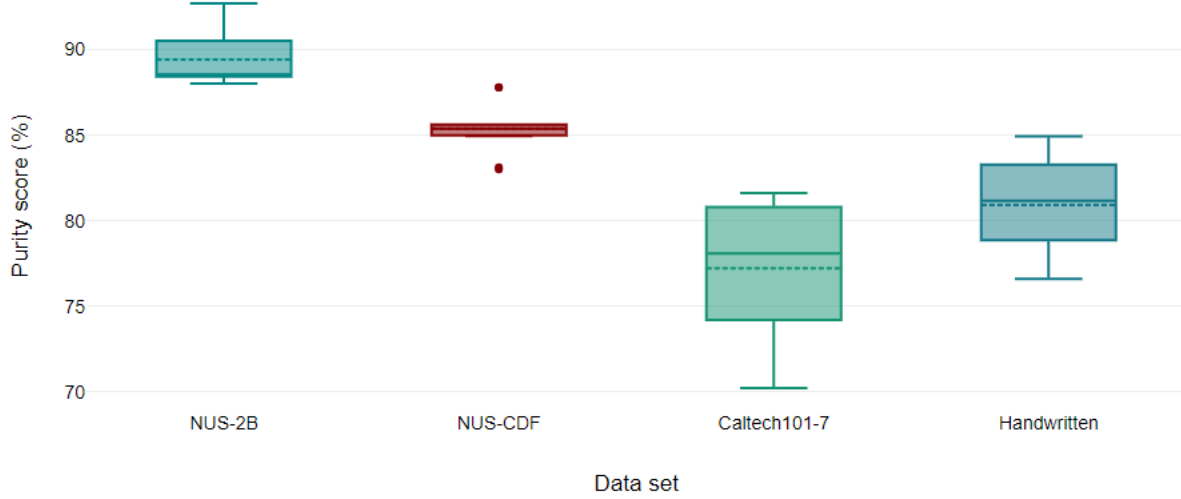


Fig. 2. Comparative results on different data sets.

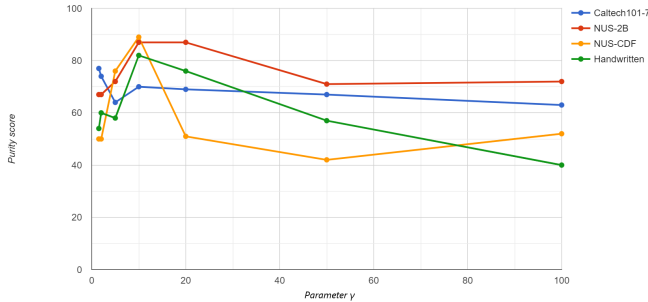


Fig. 3. Impact of the integer parameter $\gamma \geq 2$ on the purity score.

other partitions. When the value of γ is high enough in this later scenario, it would be comparable to giving each view the same weight.

V. CONCLUSION

In this paper, we presented a novel technique for multi-view consensus clustering between different local Non-negative Matrix Factorizations. The suggested approach uses the horizontal collaboration technique to update the local clusterings associated with each local view, fusing these updated clusterings

into a consensus clustering which then iteratively and alternately updates local clusterings. Additionally, we introduced a weighting strategy to mitigate the risk of negative information exchange during the horizontal collaboration and consensus steps.

The results of the experiments demonstrate that the suggested method, which has been verified against a number of data sets, produces better results with respect to the standard multi-view consensus clustering methods in terms of clustering quality. Furthermore, we show that the proposed technique is only barely influenced by random initialization.

As part of our future work, we plan to develop a strategy in the case of $\gamma \rightarrow 1$ where each view v would only collaborate with the views $v' \neq v$ that have the most similar solution. We also plan to implement the solution proposed in a deep learning context using the Multi-Layer NMF. Finally, we will integrate white noise and conduct extensive comparisons with other methods to properly assess the robustness of the proposed solution.

REFERENCES

- [1] Yan, A., Wang, W., Ren, Y. and Geng, H., 2021. A Clustering Algorithm for Multi-Modal Heterogeneous Big Data With Abnormal Data. *Frontiers in Neurorobotics*, 15, p.64.

- [2] Liu, J., Wang, C., Gao, J. and Han, J., 2013, May. Multi-view clustering via joint nonnegative matrix factorization. In Proceedings of the 2013 SIAM international conference on data mining (pp. 252-260). Society for Industrial and Applied Mathematics.
- [3] Sun, S., 2013. A survey of multi-view machine learning. *Neural computing and applications*, 23(7), pp.2031-2038.
- [4] Eaton, E., Desjardins, M. and Jacob, S., 2010, October. Multi-view clustering with constraint propagation for learning with an incomplete mapping between views. In Proceedings of the 19th ACM international conference on Information and knowledge management (pp. 389-398).
- [5] Kim, Y.M., Amini, M.R., Goutte, C. and Gallinari, P., 2010, July. Multi-view clustering of multilingual documents. In Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval (pp. 821-822).
- [6] Bickel, S. and Scheffer, T., 2004, November. Multi-view clustering. In *ICDM (Vol. 4, No. 2004)*, pp. 19-26.
- [7] Wang, S., Ye, Y. and Lau, R.Y., 2015, August. A Generative Model with Ensemble Manifold Regularization for Multi-view Clustering. In *International Conf. on Intelligent Computing* (pp. 109-114). Springer.
- [8] Wei, X., Cao, B. and Philip, S.Y., 2017, May. Multi-view unsupervised feature selection by cross-diffused matrix alignment. In *2017 International Joint Conference on Neural Networks (IJCNN)* (pp. 494-501). IEEE.
- [9] Kumar, A., Rai, P. and Daume, H., 2011. Co-regularized multi-view spectral clustering. *Advances in neural information processing systems*, 24.
- [10] Yang, Y. and Wang, H., 2018. Multi-view clustering: A survey. *Big Data Mining and Analytics*, 1(2), pp.83-107.
- [11] De Sa, V.R., Gallagher, P.W., Lewis, J.M. and Malave, V.L., 2010. Multi-view kernel construction. *Machine learning*, 79(1), pp.47-71.
- [12] Wang, X., Qian, B., Ye, J. and Davidson, I., 2013, May. Multi-objective multi-view spectral clustering via pareto optimization. In Proceedings of the 2013 SIAM international conference on data mining (pp. 234-242). Society for Industrial and Applied Mathematics.
- [13] Khalafaoui, Y., Grozavu, N., Matei, B. and Goix, L.-W., 2022. Multi-modal Multi-view Clustering Based on Non-negative Matrix Factorization. In press: 2022 IEEE Symposium Series on Computational Intelligence (IEEE SSCI), pp.1386–1391.
- [14] Jiang, Y., Liu, J., Li, Z., Li, P. and Lu, H., 2013. Co-regularized PLSA for multi-view clustering. In *Asian Conference on Computer Vision* (pp. 202-213). Springer, Berlin, Heidelberg.
- [15] Cichocki, A., Zdunek, R., Phan, A.H. and Amari, S.I., 2009. *Non-negative matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*. John Wiley & Sons.
- [16] Kim, J. and Park, H., 2008. *Sparse Nonnegative Matrix Factorization for Clustering*. Georgia Institute of Technology.
- [17] Paatero, P. and Tapper, U., 1994. Positive Matrix Factorization: A Non-negative Factor Model with Optimal Utilization of Error Estimates of Data Values. *Environmetrics*, 5(2), pp.111-126.
- [18] Brunet, J.P., Tamayo, P., Golub, T.R. and Mesirov, J.P., 2004. Metagenes and molecular pattern discovery using matrix factorization. *Proceedings of the national academy of sciences*, 101(12), pp.4164-4169.
- [19] Mirkin, B., 2001. Reinterpreting the category utility function. *Machine Learning*, 45(2), pp.219-228.
- [20] Grozavu, N., Matei, B., Bennani, Y. and Benlamine, K., 2022. Multi-view Clustering Based on Non-negative Matrix Factorization. In *Recent Advancements in Multi-View Data Analytics* (pp. 177-200). Springer.
- [21] Seung, D. and Lee, L., 2001. Algorithms for Non-negative Matrix Factorization. *Advances in neural information processing systems*, 13, pp.556-562.
- [22] Cai, X., Nie, F. and Huang, H., 2013, June. Multi-view k-means clustering on big data. In *Twenty-Third International Joint conference on artificial intelligence*.
- [23] H. W. Kuhn and A. W. Tucker. *Nonlinear Programming*. In *Berkeley University of California Press, editor, Proceedings of 2nd Berkeley Symposium*, pages 481–492, 1951.
- [24] Li, Y., Nie, F., Huang, H. and Huang, J., 2015, February. Large-scale multi-view spectral clustering via bipartite graph. In *Twenty-ninth AAAI conference on artificial intelligence*.
- [25] Singh, A.P. and Gordon, G.J., 2008, August. Relational learning via collective matrix factorization. In Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 650-658).
- [26] Benlamine, K., Bennani, Y., Matei, B., Grozavu, N. and Falih, I., 2022. Collaborative Learning to Improve the Non-uniqueness of NMF. *International Journal of Computational Intelligence and Applications*, 21(01), p.2250001.
- [27] Liu, H. and Fu, Y., 2018. Consensus guided multi-view clustering. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 12(4), pp.1-21.
- [28] Wang, Y.X. and Zhang, Y.J., 2012. Nonnegative matrix factorization: A comprehensive review. *IEEE Transactions on knowledge and data engineering*, 25(6), pp.1336-1353.