

Consistency-aware and Inconsistency-aware Graph-based Multi-view Clustering

Mitsuhiko Horie

Dept. of Communications and Computer Engineering
Waseda University
Tokyo, JAPAN
kidsgoldppp@akane.waseda.jp

Hiroyuki Kasai

Dept. of Communications and Computer Engineering
Waseda University
Tokyo, JAPAN
hiroyuki.kasai@waseda.jp

Abstract—Multi-view data analysis has gained increasing popularity because multi-view data are frequently encountered in machine learning applications. A simple but promising approach for clustering of multi-view data is multi-view clustering (MVC), which has been developed extensively to classify given subjects into some clustered groups by learning latent common features that are shared across multi-view data. Among existing approaches, graph-based multi-view clustering (GMVC) achieves state-of-the-art performance by leveraging a shared graph matrix called the unified matrix. However, existing methods including GMVC do not explicitly address inconsistent parts of input graph matrices. Consequently, they are adversely affected by unacceptable clustering performance. To this end, this paper proposes a new GMVC method that incorporates consistent and inconsistent parts lying across multiple views. This proposal is designated as CI-GMVC. Numerical evaluations of real-world datasets demonstrate the effectiveness of the proposed CI-GMVC.

I. INTRODUCTION

Many machine learning applications such as image classification, social networks, chemistry, signal processing, web analysis, item recommendation, and bioinformatics analysis usually exhibit some *structured data*, which might include trees, groups, clusters, paths, sequences [1]–[3], and graphs [4]. Recent advances in information-retrieval technologies enable collection of such structured data with *heterogeneous* features from *multi-view* data. For example, each web page includes two views of text and images. Image data include multiple features such as color histograms and frequency features of wavelet coefficients. The emergence of such multi-view data has raised a new question: how can such multiple sets of features for individual subjects be integrated into data analysis tasks? This question motivates a new paradigm, called *multi-view learning*, for data analysis with multi-view feature information. Multi-view learning fundamentally makes use of common or consensus information that is presumed to exist across multi-view data to improve data analysis task performance. One successful subcategory of multi-view learning is *multi-view clustering* (MVC), which classifies given subjects into subgroups based on similarities among subjects [5]. Although various approaches have been proposed in this category, graph-based multi-view clustering (GMVC) has recently garnered increasing attention: it has demonstrated state-of-the-art performance for numerous applications [6]–[14].

Fundamentally, GMVC originates from single-view spectral clustering methods. It performs clustering tasks by exploiting consensus features across input *graph matrices* such as adjacency matrices. Among them, some works outperform other methods by seeking a consensus matrix called a *unified matrix* from the input graph matrices with adaptive weights such that the unified matrix directly represents the final clustering result [7], [9], [10], [13], [15].

Most existing methods achieve good performance by exploiting the *consensus part* across multiple views. However, each view might include an *inconsistent part* that does not appear in other views because of noise and outliers. Therefore, this inclusion might lead to severe degradation of downstream clustering performance. Nevertheless, they do not deal explicitly with such inconsistent parts of the input graph matrices. Consequently, they offer only unacceptable performance.

To alleviate this issue, this paper presents a proposal of graph-based multi-view clustering, which particularly incorporates a consistent part and an inconsistent part lying across multiple views. More concretely, we separate the input graph matrices collected from multi-view data into a consistent part and an inconsistent part by orthogonality constraints. This conceptualization shares similar ideas with some recent works [16], [17]. Derivation of the unified matrix is therefore more efficient and more robust than existing GMVC methods, leading to improvement of the clustering performance. We designate this proposed algorithm as CI-GMVC. Numerical evaluations conducted with comparison to state-of-the-art multi-view clustering methods reveal the effectiveness of the proposed CI-GMVC on several real-world datasets. The source code is available at <https://github.com/hiroyuki-kasai>.

II. PRELIMINARY EXPLANATION

This subsection first presents a summary of the notation used in the remainder of this paper. Subsequently, we briefly introduce multi-view clustering and specifically address graph-based multi-view clustering, which are basic techniques of the proposed algorithm: CI-GMVC.

A. Notation

We represent scalars as lower-case letters (a, b, \dots), vectors as bold lower-case letters ($\mathbf{a}, \mathbf{b}, \dots$), and matrices as bold-face

capitals ($\mathbf{A}, \mathbf{B}, \dots$). The i -th element of \mathbf{a} and the element at the (i, j) position of \mathbf{A} are represented respectively as a_i and A_{ij} . Also, the vector of the i -th row vector $A_{i:}$ and the vector of the j -th column vector $A_{:j}$ are denoted respectively as \mathbf{a}_i and \mathbf{a}^j . $\mathbf{1}_d$ is used for the d -dimensional vector of ones, and \mathbf{I}_d is the identity matrix of size $d \times d$. $\mathbb{R}_+^{n \times m}$ represents a nonnegative matrix of size $n \times m$. Additionally, $\mathbf{A} \geq 0$ and $\mathbf{A} > 0$ respectively represent $A_{ij} \geq 0$ and $A_{ij} > 0$ for all (i, j) . Furthermore, $\mathbf{A} \geq \mathbf{B}$ and $\mathbf{A} > \mathbf{B}$ represent $A_{ij} \geq B_{ij}$ and $A_{ij} > B_{ij}$ for all (i, j) . Operators $\text{Tr}(\cdot)$ and $(\cdot)^T$ respectively stand for the matrix trace and transpose. Operator $\max(\mathbf{A}, \mathbf{B})$ outputs A_{ij} when $A_{ij} \geq B_{ij}$ at the (i, j) -th position, and B_{ij} otherwise. Operator $\text{vec}(\mathbf{A})$ performs the vectorization of \mathbf{A} . Operator $\text{diag}(\mathbf{a})$ represents \mathbf{A} of which diagonal and off-diagonal elements are \mathbf{a} and zeros, respectively. Regarding multi-view data, N , V , and C respectively represent the number of sample data, views, and classes. The dimensions of sample data of the c ($\in [C]$)-th class in the v ($\in [V]$)-th view are denoted as d_v .

B. Multi-view clustering: MVC

Multi-view learning fundamentally makes use of common or consensus information across multi-view data [18]–[21]. *Multi-view discriminant learning*, a supervised learning technique, has been studied extensively [22]–[29]. It generally originates from single-view linear discriminant analysis such as Fisher linear discriminant analysis (LDA or FDA) [30]. Regarding unsupervised learning techniques, however, *multi-view clustering* (MVC) clusters given subjects into several groups such that the points in the same group are similar and the points in different groups are dissimilar to one another by combining multi-view data [5]. One naive approach of MVC is to perform a single-view clustering method against *concatenated* features collected from different views. However, this approach might fail when higher emphases are put to certain specific views than to others. Consequently, this category of research has attracted more attention, to include multi-view subspace clustering that learns common coefficient matrices [31], [32], multi-view nonnegative matrix factorization clustering that learns common indicator matrices [33], multi-view k -means [34], multi-kernel based multi-view clustering [35] and CCA based multi-view clustering [36].

C. Graph-based multi-view clustering: GMVC

Different from the approaches described above, graph-based MVC (GMVC) learns common eigenvector matrices or shared matrices, and empirically demonstrates state-of-the-art results in various applications. General steps consist of (i) generating an input graph matrix, called SIG, (ii) generating the graph Laplacian matrix, (iii) computing the embedding matrix, and (iv) performing clustering into groups using an external clustering algorithm. These steps are shared with the normalized cut [37] and the spectral clustering [38]. Furthermore, GMVC is related closely to multi-view spectral clustering [39]–[45]. Recently, some works of GMVC address the unified matrix $\mathbf{U} (\in \mathbb{R}_+^{N \times N})$ with different effects of multiple graph matrices

as [7], [9], [10], [13], [15]. More noteworthy is that, whereas many MVC methods rely on an external clustering algorithm after learning, they seek a unified graph matrix such that it *internally* indicates cluster information.

More specifically, considering an adaptive weight $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_v, \dots, \alpha_V]^T (\in \mathbb{R}^V)$ on all V SIG matrices $\{\mathbf{S}_1, \dots, \mathbf{S}_v, \dots, \mathbf{S}_V\} (\in \mathbb{R}_+^{N \times N})$, the following minimization problem is formulated with respect to \mathbf{U} :

$$\begin{aligned} \min_{\mathbf{U}} \quad & \sum_{v=1}^V \alpha_v \|\mathbf{U} - \mathbf{S}_v\|_F^2 \\ \text{subject to} \quad & U_{ij} \geq 0, \mathbf{1}^T \mathbf{u}_i = 1, \end{aligned} \quad (1)$$

where $\boldsymbol{\alpha}$ represents the weight vector $\boldsymbol{\alpha}$, calculated as $\alpha_v = 1/(2\sqrt{\|\mathbf{U} - \mathbf{S}_v\|_F^2})$, as in [9], [13], [15]. Furthermore, the graph Laplacian matrix $\mathbf{L}_\mathbf{U} (\in \mathbb{R}^{N \times N})$ of \mathbf{U} is introduced such that \mathbf{U} directly produces the clustering result without relying on external clustering methods. For this purpose, acknowledging that \mathbf{U} can be partitioned into C groups directly when $\text{rank}(\mathbf{L}_\mathbf{U})$ is equal to $n - C$ [46], and using Fan's theorem [47], the following formulation is proposed:

$$\begin{aligned} \min_{\mathbf{U}, \mathbf{F}, \boldsymbol{\alpha}} \quad & \sum_{v=1}^V \alpha_v \|\mathbf{U} - \mathbf{S}_v\|_F^2 + 2\lambda \cdot \text{Tr}(\mathbf{F}^T \mathbf{L}_\mathbf{U} \mathbf{F}) \\ \text{subject to} \quad & U_{ij} \geq 0, \mathbf{1}^T \mathbf{u}_i = 1, \mathbf{F}^T \mathbf{F} = \mathbf{I}_C, \end{aligned} \quad (2)$$

where $\lambda (> 0)$ is a regularization parameter. $\mathbf{F} (\in \mathbb{R}^{N \times C})$ is an embedding matrix, which lies on the orthogonal matrix, i.e., the Stiefel manifold $\text{St}(p, d)$; the Riemannian submanifold of orthonormal matrices $\mathcal{M} = \{\mathbf{X} \in \mathbb{R}^{d \times p} : \mathbf{X}\mathbf{X}^T = \mathbf{I}_p\}$.

III. PROPOSED CI-GMVC: CONSISTENCY-AWARE AND INCONSISTENCY-AWARE GMVC

As explained in Section II-C, GMVC considers consensus features across input graph matrices, i.e., SIG matrices, across multiple views. The hypothesis in this scheme relies on the assumption that all SIG matrices $\{\mathbf{S}_1, \dots, \mathbf{S}_v, \dots, \mathbf{S}_V\}$ share common features across multi-views with appropriate weights. Subsequently, \mathbf{U} can be obtained efficiently from these matrices. However, as explained in Section I, the SIG matrices \mathbf{S}_v are not perfect: they might be corrupted because of noisy input data, severe outliers, and latent fundamental inconsistent structures [16]. Consequently, any analysis of these matrices might result in unacceptable clustering results. Therefore, it is necessary to handle inconsistent parts among multi-view data to avoid lower quality of subsequent clustering.

To alleviate this issue, we separate \mathbf{S}_v into a consistent part and an inconsistent part that exist as mixed inside \mathbf{S}_v , and calculate \mathbf{U} *only* from the consistent parts of \mathbf{S}_v . More concretely, we assume that \mathbf{S}_v consists of a consistent part $\mathbf{A}_v (\in \mathbb{R}^{N \times N})$ and a *nonnegative* inconsistent part $\mathbf{E}_v (\in \mathbb{R}_+^{N \times N})$ as $\mathbf{S}_v = \mathbf{A}_v + \mathbf{E}_v$ which follows [16], [17]. Assuming further that the elements of \mathbf{E}_v are *not shared* in other $\mathbf{E}_w (v \neq w)$, in other words, expecting that they are *element-wise orthogonal*, we assign a penalty as

$$\sum_{v,w=1}^V \gamma \cdot \text{vec}(\mathbf{E}_v)^T \text{vec}(\mathbf{E}_w) = \sum_{v,w=1}^V \gamma \cdot \text{Tr}(\mathbf{E}_v \cdot (\mathbf{E}_w)^T),$$

where $\gamma(> 0)$ is a weighting hyperparameter. Furthermore, a penalty of larger elements of \mathbf{E}_v is regarded as stabilizing the optimization of the objective function as

$$\sum_{v=1}^V \beta \|\text{vec}(\mathbf{E}_v)\|_2^2 = \sum_{v=1}^V \beta \cdot \text{Tr}(\mathbf{E}_v \cdot (\mathbf{E}_v)^T),$$

where $\beta(> 0)$ is a weighting hyperparameter. Integrating the two penalties above yields the following penalty term.

$$\sum_{v,w=1}^V b_{vw} \cdot \text{Tr}((\mathbf{S}_v - \mathbf{A}_v) \cdot (\mathbf{S}_w - \mathbf{A}_w)^T).$$

Therein, b_{vw} represents the (v, w) -elements of $\mathbf{B}(\in \mathbb{R}^{V \times V})$, of which diagonal and off-diagonal elements respectively correspond to β and γ . Finally, we derive the objective formulation mathematically as presented below:

$$\begin{aligned} \min_{\substack{\mathbf{U}, \mathbf{F}, \boldsymbol{\alpha} \\ \mathbf{A}_1, \dots, \mathbf{A}_V}} & \sum_{v=1}^V \alpha_v \|\mathbf{U} - \mathbf{A}_v\|_F^2 + 2\lambda \text{Tr}(\mathbf{F}^T \mathbf{L}_U \mathbf{F}) \\ & + \sum_{v,w=1}^V b_{vw} \cdot \text{Tr}((\mathbf{S}_v - \mathbf{A}_v) \cdot (\mathbf{S}_w - \mathbf{A}_w)^T) \end{aligned} \quad (3)$$

subject to $U_{ij} \geq 0, \mathbf{1}^T \mathbf{u}_i = 1, \mathbf{F}^T \mathbf{F} = \mathbf{I}_C, \mathbf{S}_v \geq \mathbf{A}_v \geq 0$.

As shown there, the differences one can note in relation to existing works are that the unified matrix \mathbf{U} is evaluated with \mathbf{A}_v , i.e., the consistent part of \mathbf{S}_v , and the inconsistent part $\mathbf{S}_v - \mathbf{A}_v$ are evaluated with $\mathbf{S}_w - \mathbf{A}_w$ ($v \neq w$) in terms of their mutual orthogonality.

IV. OPTIMIZATION ALGORITHM

The objective function in (3) is not jointly convex on all variables. Therefore, the alternating minimization algorithm is exploited to obtain the solutions. Note that, as mentioned earlier, $\boldsymbol{\alpha}$ is calculated as $\alpha_v = 1/(2\sqrt{\|\mathbf{U} - \mathbf{A}_v\|_F^2})$, as in [9], [13], [15].

A. Update of \mathbf{F} and \mathbf{U}

The updates of \mathbf{F} and \mathbf{U} are similar to those of [15]. However, for the self-contained explanation, we briefly describe their update rules. The optimization problem in (3) with respect to the unified matrix \mathbf{U} yields the following:

$$\begin{aligned} \min_{\mathbf{U}} & \sum_{v=1}^V \alpha_v \|\mathbf{U} - \mathbf{A}_v\|_F^2 + 2\lambda \cdot \text{Tr}(\mathbf{F}^T \mathbf{L}_U \mathbf{F}) \\ \text{subject to} & U_{ij} \geq 0, \mathbf{1}^T \mathbf{u}_i = 1. \end{aligned}$$

This is equivalent to the following minimization problem in terms of \mathbf{u}_i for $i \in [N]$ as

$$\begin{aligned} \min_{\mathbf{u}_i} & \sum_{v=1}^V \|\mathbf{u}_i - (\mathbf{a}_v)_i\|_2^2 + \frac{\lambda}{2V\alpha_v} \mathbf{p}_i^2 \quad (4) \\ \text{subject to} & U_{ij} \geq 0, \mathbf{1}^T \mathbf{u}_i = 1, \end{aligned}$$

where $\mathbf{p}_i = [(p_i)_1, (p_i)_2, \dots, (p_i)_j, \dots, (p_i)_N]^T (\in \mathbb{R}^N)$ and where $(p_i)_j$ is equal to $\|\mathbf{f}^i - \mathbf{f}^j\|_2^2$; $\mathbf{f}^i (\in \mathbb{R}^{1 \times C})$ is the i -th row vector of \mathbf{F} . This problem is solvable as in Section 5.3

in [15]. Finally, the optimization problem in (3) in terms of \mathbf{F} under fixed $\mathbf{U}, \mathbf{A}, \boldsymbol{\alpha}$ is

$$\min_{\mathbf{F}} \text{Tr}(\mathbf{F}^T \mathbf{L}_U \mathbf{F}), \quad \text{subject to } \mathbf{F}^T \mathbf{F} = \mathbf{I}_C. \quad (5)$$

The solution is obtainable to calculate the C eigenvectors of \mathbf{L}_U , of which eigenvalues are the C smallest ones [38].

B. Update of \mathbf{A}

Keeping $\mathbf{F}, \mathbf{U}, \boldsymbol{\alpha}$ as fixed, the minimization problem about \mathbf{A}_v is defined as

$$\begin{aligned} \min_{\mathbf{A}_1, \dots, \mathbf{A}_V} & \sum_{v=1}^V \alpha_v \|\mathbf{U} - \mathbf{A}_v\|_F^2 \\ & + \sum_{v,w=1}^V b_{vw} \cdot \text{Tr}((\mathbf{S}_v - \mathbf{A}_v) \cdot (\mathbf{S}_w - \mathbf{A}_w)^T) \\ \text{subject to} & \mathbf{S}_v \geq \mathbf{A}_v \geq 0. \end{aligned}$$

The first-order necessary optimality conditions of this problem are that its gradient with respect to \mathbf{A}_v is expected to be zero, which means

$$2\alpha_v(\mathbf{U} - \mathbf{A}_v) + \sum_{w=1}^V b_{vw}(-\mathbf{S}_w + \mathbf{A}_w) = \mathbf{0},$$

for $v = 1, 2, \dots, V$. Consequently, we obtain the following.

$$2\alpha_v \mathbf{A}_v + \sum_{w=1}^V b_{vw} \mathbf{A}_w = 2\alpha_v \mathbf{U} + \sum_{w=1}^V b_{vw} \mathbf{S}_w. \quad (6)$$

Here, because the left-hand terms are represented as $b_{v1}\mathbf{A}_1 + \dots + (b_{vv} + 2\alpha_v)\mathbf{A}_v + \dots + b_{vV}\mathbf{A}_V$, its vectorization form is represented as

$$[b_{v1} \ b_{v2} \ \dots \ (b_{vv} + 2\alpha_v) \ \dots \ b_{vV}] \begin{pmatrix} \text{vec}(\mathbf{A}_1)^T \\ \vdots \\ \text{vec}(\mathbf{A}_V)^T \end{pmatrix}.$$

Adding all V terms of \mathbf{A}_v yields the following.

$$\begin{pmatrix} b_{11} + 2\alpha_1 & \dots & \dots & \dots & b_{1V} \\ \vdots & \ddots & \dots & \dots & \vdots \\ b_{v1} & \dots & b_{vv} + 2\alpha_v & \dots & b_{vV} \\ \vdots & \dots & \dots & \ddots & \vdots \\ b_{V1} & \dots & \dots & \dots & b_{VV} + 2\alpha_V \end{pmatrix} \begin{pmatrix} \text{vec}(\mathbf{A}_1)^T \\ \vdots \\ \text{vec}(\mathbf{A}_V)^T \end{pmatrix}.$$

Denoting $2\text{diag}(\boldsymbol{\alpha}) + \mathbf{B} (\in \mathbb{R}^{V \times V})$ as \mathbf{C} , and the right-hand terms in (6) as $\mathbf{H}_v (\in \mathbb{R}^{V \times V})$, we calculate \mathbf{A}_v ($v \in [V]$) as

$$\begin{pmatrix} \text{vec}(\mathbf{A}_1)^T \\ \vdots \\ \text{vec}(\mathbf{A}_V)^T \end{pmatrix} = \mathbf{C}^+ \cdot \begin{pmatrix} \text{vec}(\mathbf{H}_1)^T \\ \vdots \\ \text{vec}(\mathbf{H}_V)^T \end{pmatrix}, \quad (7)$$

where \mathbf{C}^+ is the inverse or the pseudo-inverse matrix of \mathbf{C} .

Finally, considering the constraint of $\mathbf{S}_v \geq \mathbf{A}_v \geq 0$, the final solution of \mathbf{A}_v is obtainable by outputting \mathbf{A}_v'' as

$$\mathbf{A}_v' = \max(\mathbf{A}_v, \mathbf{0}), \quad \mathbf{A}_v'' = \min(\mathbf{A}_v', \mathbf{S}_v). \quad (8)$$

The overall algorithm of the proposed CI-GMVC is summarized in **Algorithm 1**.

Algorithm 1 CI-GMVC optimization algorithm**Require:** Unified matrix \mathbf{U} .**Ensure:** SIG matrix $\mathbf{S}_1, \dots, \mathbf{S}_V$, cluster number C , λ , β , γ .

- 1: Initialize $\alpha_v = 1/V$ and $\mathbf{A}_v = \mathbf{S}_v$ for $v \in [V]$.
- 2: Initialize \mathbf{U} from α weighted summation of $\mathbf{S}_1, \dots, \mathbf{S}_V$.
- 3: Initialize \mathbf{F} using (5).
- 4: Update α as $\alpha_v = 1/(2\sqrt{\|\mathbf{U} - \mathbf{A}_v\|_F^2})$.
- 5: Update \mathbf{U} with \mathbf{F} , \mathbf{A} , α fixed using (4).
- 6: Update \mathbf{F} with \mathbf{U} , \mathbf{A} , α fixed using (5).
- 7: Update \mathbf{A} with \mathbf{F} , \mathbf{U} , α fixed using (7) and (8).
- 8: Repeat the steps presented above until \mathbf{U} converges or the predefined maximum number of iterations is reached.

V. NUMERICAL EVALUATIONS

This section presents empirical evaluation of the proposed CI-GMVC with some real-world datasets. We compare the proposed algorithm with state-of-the-art methods, which include Multi-view Spectral Clustering (MSC)¹ [41], Co-regularized Spectral Clustering (CoregSC)² [40], Multiple Graph Learning (MGL)³ [7], Multi-view Clustering with Graph Learning (MCGL)⁴ [13], and Graph-based System (GBS)⁵ [15]. As for GBS and our proposed CI-GMVC methods, the SIG matrix is generated by following Algorithm 1 in [15] with the number of neighbors $k = 15$. The hyper-parameters for our proposed CI-GMVC are $\beta = 10^{-12}$ and $\gamma = 10^{-5}$, which are obtained from preliminary experiments.

Datasets summarized in TABLE I are the following. The BBC (BBC) dataset⁶ includes news articles from the BBC news website. The number of articles is 685. Each has one of five topical labels. The Newsgroup (NG) dataset⁷ is collected from the 20 News-group datasets, which has 500 newsgroup documents. This has five topical labels. The WebKB dataset⁸ has four classes; it includes 203 web-pages. Each web-page consists of the anchor text of the hyperlink, its title, and the page content. The One-hundred plant species leaves (100 leaf) dataset⁹ includes three views of which one has 1600 samples. Each belongs to one of the one hundred plant species.

TABLE I
FEATURES OF DATASETS USED FOR THIS EXPERIMENT.

dataset	N	V	C	dimensions of each view			
				d_1	d_2	d_3	d_4
BBC	685	4	5	4659	4633	4665	4684
NGs	500	3	5	2000	2000	2000	–
WebKB	203	3	4	1703	230	230	–
100leaves	1600	3	100	64	64	64	–

¹Source code available at <https://github.com/frash1989/ELM-MVClustering/tree/master/RMSC-ELM>.

²Source code available at http://legacydirs.umiacs.umd.edu/~abhishek/code_corespectral.zip.

³Source code available at <http://www.escience.cn/people/fpnie>.

⁴Source code available at <https://github.com/kunzhan/MVGL>.

⁵Source code available at <https://github.com/cswanghao/gbs>.

⁶<http://mlg.ucd.ie/datasets/segment.html>.

⁷<http://lig-membres.imag.fr/grimal/data.html>.

⁸<https://linqs.soe.ucsc.edu/data>.

⁹<https://archive.ics.uci.edu/ml/datasets/One-hundred+plant+species+leaves+data+set>.

A. Convergence behavior

The objective function in (3) is not convex on all variables. Therefore, this subsection confirms the convergence behaviors of our proposed CI-GMVC compared with GBS, which outperforms others. For a fair comparison, we evaluate the objective function without regularizers (3), i.e., $\sum_{v=1}^V \alpha_v \|\mathbf{U} - \mathbf{S}_v\|_F^2$ in (1), and use the same stopping condition as that used for GBS. As Fig. 1 shows, the convergences of CI-GMVC on the WebKB and 100 leaf datasets are faster than those of GBS.

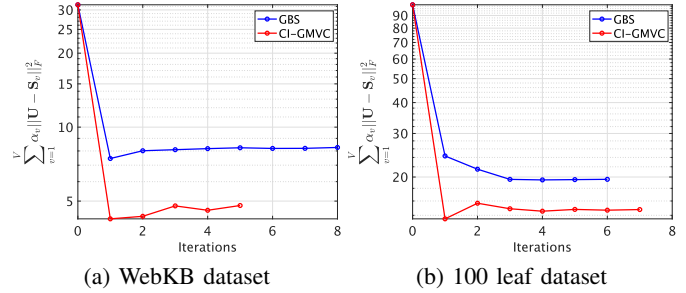


Fig. 1. Convergence behaviors of CI-GMVC compared to those of GBS.

B. Clustering performance

This subsection presents comparisons of clustering performance. The results of ACC and NMI are shown respectively in TABLE II and TABLE III, where the average scores of the accuracy (ACC) and the normalized mutual information (NMI) are shown. The best performances are presented in bold. Results aside from those of GBS and CI-GMVC are from those in [15]. The results demonstrated that the proposed CI-GMVC is comparable to or outperforms other state-of-the-art methods.

TABLE II
AVERAGE CLUSTERING PERFORMANCE (ACC).

method	BBC	NGs	WebKB	100leaves
CoregSC [40]	47.01	27.68	59.70	77.06
MSC [41]	62.32	31.12	47.34	73.79
MGL [7]	53.96	82.18	73.84	69.04
MCGL [13]	35.33	24.60	54.19	81.06
GBS [15]	69.34	98.20	74.38	82.44
CI-GMVC (proposed)	70.36	98.40	77.34	82.44

TABLE III
AVERAGE CLUSTERING PERFORMANCE (NMI).

method	BBC	NGs	WebKB	100leaves
CoregSC [40]	28.63	8.80	31.39	91.65
MSC [41]	55.31	9.72	22.37	90.14
MGL [7]	36.97	83.04	43.62	87.53
MCGL [13]	7.41	10.72	8.60	91.30
GBS [15]	56.27	93.92	37.83	93.43
CI-GMVC (proposed)	58.59	94.61	47.01	93.52

VI. CONCLUSIONS

The proposed graph-based multi-view clustering method CI-GMVC particularly incorporates the consistency and the inconsistency structure lying across multiple views. Numerical evaluations using several real-world datasets demonstrated the effectiveness of the proposed CI-GMVC.

ACKNOWLEDGMENTS

H. Kasai was partially supported by JSPS KAKENHI Grant Numbers JP16K00031 and JP17H01732.

REFERENCES

- [1] H. Narimatsu and H. Kasai, "Duration and interval hidden Markov model for sequential data analysis," in *Proc. of International Joint Conference on Neural Networks (IJCNN)*, 2015.
- [2] H. Narimatsu and H. Kasai, "State duration and interval modeling in hidden semi-markov model for sequential data analysis," *Annals of Mathematics and Artificial Intelligence*, vol. 81, no. 3-4, pp. 377–403, 2017.
- [3] H. Narimatsu and H. Kasai, "Overlapped state hidden semi-markov model for grouped multiple sequences," in *45th International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2020.
- [4] R. Hashimoto and H. Kasai, "Sequential semi-orthogonal multi-level nmf with negative residual reduction for network embedding," in *45th International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2020.
- [5] G. Chao, S. Sun, and J. Bi, "A survey on multi-view clustering," *arXiv preprint arXiv:1712.06246*, 2017.
- [6] M. Saha, "A graph based approach to multiview clustering," in *Pattern Recognition and Machine Intelligence (PRMI)*, P. Maji, A. Ghosh, M. N. Murty, K. Ghosh, and S. K. Pal, Eds., 2013, vol. 8251 of *Lecture Notes in Computer Science*, pp. 128–133.
- [7] F. Nie and X. Li, J. ad Li, "Parameter-free auto-weighted multiple graph learning: A framework for multiview clustering and semi-supervised classification," in *Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI)*, 2016.
- [8] G. How, F. Nie, H. Tao, and D. Yi, "Multi-view unsupervised feature selection with adaptive similarity and view weight," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 9, pp. 1998–2011, 2017.
- [9] F. Nie, J. Li, and X. Li, "Self-weighted multiview clustering with multiple graphs," in *Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI)*, 2017.
- [10] H. Tao, C. Hou, J. Zhu, and D. Yi, "Multi-view clustering with adaptively learned graph," in *Ninth Asian Conference on Machine Learning (ACML)*, 2017.
- [11] W. Zhuge, F. Nie, C. Hou, and D. Yi, "Unsupervised single and multiple views feature extraction with structured graph," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 10, 2017.
- [12] F. Nie, G. Cai, and X. Li, "Multi-view clustering and semi-supervised classification with adaptive neighbours," in *Thirty-First AAAI Conference on Artificial Intelligence (AAAI)*, 2017.
- [13] F. Nie, G. Cai, J. Li, and X. Li, "Auto-weighted multi-view learning for image clustering and semi-supervised classification," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1501–1511, 2018.
- [14] K. Zhan, C. Zhang, J. Guan, and J. Wang, "Graph learning for multiview clustering," *IEEE Trans. Cybern.*, vol. 48, no. 10, pp. 2887–2895, 2018.
- [15] H. Wang, Y. Yang, B. Liu, and H. Fujita, "A study of graph-based system for multi-view clustering," *Knowledge-Based Systems*, vol. 24, no. 163, pp. 1009–1019, 2019.
- [16] A. Bojchevski, Y. Matkovic, and S. Gunnemann, "Robust spectral clustering for noisy data: Modeling sparse corruptions improves latent embeddings," in *KDD*, 2017.
- [17] Y. Liang, D. Huang, and C.-D. Wang, "Consistency Meets Inconsistency: A Unified Graph Learning Framework for Multi-view Clustering," in *IEEE International Conference on Data Mining (ICDM)*, 2019.
- [18] C. Xu, D. Tao, and C. Xu, "A survey on multi-view learning," *arXiv preprint arXiv:1304.5634*, 2013.
- [19] S. Sun, "A survey of multi-view machine learning," *Neural Computing and Applications*, vol. 23, no. 7-8, pp. 2031–2038, 2013.
- [20] J. Zhao, X. Xie, Xu. X., and S. Sun, "Multi-view learning overview: Recent progress and new challenges," *Information Fusion*, vol. 38, pp. 43–54, 2017.
- [21] Y. Li, M. Yang, and Z. Zhang, "A survey of multi-view representation learning," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 10, pp. 1863–1883, 2019.
- [22] R. Rosipal and N. Kramer, "Overview and recent advances in partial least squares," in *Proceedings of the 2005 international conference on Subspace, Latent Structure and Feature Selection (SLSFS'05)*, 2005.
- [23] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3-4, pp. 321–377, 1936.
- [24] B. Thompson, "Canonical correlation analysis," *Encyclopedia of Statistics in Behavioral Science*, vol. 1, no. 1, pp. 192–196, 2005.
- [25] J. B. Tenenbaum and W. T. Freeman, "Separating style and content with bilinear models," *Neural Computation*, vol. 12, no. 6, pp. 1247–1283, 2000.
- [26] A. Sharma, A. Kumar, H. Daume, and D.W. Jacobs, "Generalized multiview analysis: A discriminative latent space," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.
- [27] G. Cao, A. Iosifidis, K. Chen, and M. Gabbouj, "Generalized multi-view embedding for visual recognition and cross-modal retrieval," *IEEE Trans. Cybern.*, vol. 48, no. 9, pp. 2542–2555, 2016.
- [28] J. Xu, S. Yu, X. You, M. Leng, X.-Y. Jing, and C.L.P. Chen, "Multiview hybrid embedding: A divide-and-conquer approach," in *IEEE Trans. Cybern.*, 2019.
- [29] H. Kasai, "Multi-view Wasserstein discriminant analysis with entropic regularized Wasserstein distance," in *45th International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2020.
- [30] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer Series in Statistics, 2009.
- [31] Y. Wang, W. Zhang, L. Wu, X. Lin, M. Fang, and Y. Pan, "Iterative views agreement: An iterative low-rank based structured optimization method to multi-view spectral clustering," in *Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI)*, 2016.
- [32] Q. Yin, S. Wu, and L. Wang, "incomplete multi-view clustering via subspace learning," in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management (CIKM)*, 2015.
- [33] Z. Akata, C. Bauckhage, and C. Thureau, "Non-negative matrix factorization in multimodality data for segmentation and label prediction," in *16th Computer Vision Winter Workshop*, 2011.
- [34] X. Cai, F. Nie, and H. Huang, "Multi-view k-means clustering on big data," in *Twenty-Third International Joint Conference on Artificial Intelligence (IJCAI)*, 2013.
- [35] D. Guo, J. Zhang, X. Liu, Y. Cui, and C. Zhao, "Multiple kernel learning based multi-view spectral clustering," in *22nd International Conference on Pattern Recognition (ICPR)*, 2014.
- [36] M. B. Blaschko and C. H. Lampert, "Correlational spectral clustering," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.
- [37] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, 2000.
- [38] B. Mohar, "The laplacian spectrum of graphs," in *Graph Theory, Combinatorics, and Applications*, Y. Alavi, G. Chartrand, and O. Oellermann, Eds., pp. 871–898. 1991.
- [39] T. Xia, D. Tao, T. Mei, and Y. Zhang, "Multiview spectral embedding," *IEEE Trans. Syst. Man Cybern. B*, vol. 40, no. 6, pp. 1438–1446, 2010.
- [40] P. Kumar, A. ad Rai and H. Daume, "Co-regularized multi-view spectral clustering," in *Advances in Neural Information Processing Systems (NIPS)*, 2011.
- [41] R. Xia, Y. Pan, L. Du, and J. Yin, "Robust multi-view spectral clustering via low-rank and sparse decomposition," in *Twenty-Eighth AAAI Conference on Artificial Intelligence (AAAI)*, 2014.
- [42] C. Lu, C. Yan, and Z. Lin, "Convex sparse spectral clustering: Single-view to multi-view," *IEEE Trans. Image Process.*, vol. 25, no. 6, pp. 2833–2843, 2016.
- [43] L. Feng, L. Cai, Y. Liu, and S. Liu, "Multi-view spectral clustering via robust local subspace learning," *Soft Comput.*, vol. 21, no. 8, pp. 1937–1948, 2017.
- [44] Wang. Y., L. Wu, X. Lin, and J. Gao, "Multiview spectral clustering via structured low-rank matrix factorization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 10, pp. 4833–4843, 2018.
- [45] L. Zong, X. Zhang, X. Liu, and H. Yu, "Weighted multi-view spectral clustering based on spectral perturbation," in *Thirty-Second AAAI Conference on Artificial Intelligence (AAAI)*, 2018.
- [46] I. S. Dhillon, "Co-clustering documents and words using bipartite spectral graph partitioning," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2001.
- [47] K. Fan, "On a theorem of weyl concerning eigenvalues of linear transformations I," *Proc. Natl. Acad. Sci. USA*, vol. 35, no. 11, pp. 652–655, 1949.