A Survey and an Empirical Evaluation of Multi-view Clustering Approaches

LIHUA ZHOU

School of Information Science & Engineering, Yunnan University, Kunming 650091, Yunnan, PR China, lhzhou@ynu.edu.cn

Guowang Du

School of Information Science & Engineering, Yunnan University, Kunming 650091, Yunnan, PR China, dugking@mail.ynu.edu.cn

Kevin Lü

Brunel University, Uxbridge UB8 3PH, UK, kevin.lu@brunel.ac.uk

Lizheng Wang

a. School of Information Science & Engineering, Yunnan University, Kunming 650091, Yunnan, PR China

b. Dianchi College of Yunnan University, Kunming 650228, Yunnan, PR China

lzhwang@ynu.edu.cn

Jingwei Du

School of Information Science & Engineering, Yunnan University, Kunming 650091, Yunnan, PR China, jingweidu@mail.ynu.edu.cn

Multi-view clustering (MVC) holds a significant role in domains like machine learning, data mining, and pattern recognition. Despite the development of numerous new MVC approaches employing various techniques, there remains a gap in comprehensive studies evaluating the characteristics and performance of these approaches. This gap hinders the in-depth understanding and rational utilization of the recently developed MVC techniques. This study formalizes the basic concepts of MVC and analyzes their techniques. It then introduces a novel taxonomy for MVC approaches and presents the working mechanisms and characteristics of representative MVC approaches developed in recent years. Moreover, it summarizes representative datasets and performance metrics commonly employed for evaluating MVC approaches. Furthermore, we have meticulously chosen thirty-five representative MVC approaches to conduct an empirical evaluation across seven real-world benchmark datasets, offering valuable insights into the realm of MVC approaches.

CCS CONCEPTS • General and reference~Document types~Surveys and overviews • Computing methodologies~Machine learning~Learning paradigms~Unsupervised learning~Clustering analysis

Additional Keywords and Phrases: Multi -view clustering, Consensus and complementary principles, Information fusion, Weighting, Clustering routine

ACM 0360-0300/2024/02-ART
https://doi.org/10.1145/3645108

1 INTRODUCTION

With advances in information acquisition technologies, multi-view data, collected from diverse domains or obtained from various feature extractors, has become ubiquitous. Each domain or feature is referred to as a particular view, which reflects a specific property of an object, for example, color and texture describe an image from two different aspects. Different views generally provide compatibility and complementary information to each other [1-4]. The effective utilization of this information can facilitate an increase in the accuracy of learning algorithms [5].

One way for solving the multi-view problem is to treat each view independently, or to concatenate vectors from different views into a new vector and then to apply single-view learning algorithms straightforwardly on the concatenated vector. The strategy that treats each view independently may overlook the complementary and diverse information of different views, thereby losing the advantages of multi-view data. On the other hand, the strategy of concatenation disregards the specific properties of each view and may result in high dimensionality, potentially reducing the interpretability of different views and causing overfitting on a small training sample [2, 6]. Unlike the aforementioned strategies, multi-view learning (MVL) seeks to integrate features and structural information from multiple views, exploiting the richer properties of data to enhance learning performance. Therefore, MVL has emerged as an important research area [4, 5, 7-10].

Multi-view clustering (MVC), playing an imperative role in MVL for reducing data complexity and facilitating interpretation, aims to group samples (objects/instances/points) with similar feature structures or patterns into the same group (cluster) and samples with dissimilar ones into different groups, by combining the available feature information from different views and searching for consistent clusters across different views [11, 12]. As a powerful alternative learning tool for uncovering the underlying structure shared by multiple views in the absence of label information, MVC has attracted increasing attention in recent years, and has been successfully used in widespread applications [13]. However, MVC is not a trivial task. On one hand, each view may have its own feature or distribution, and multiple views may take different forms as well as exhibit heterogeneous and diverse properties. Moreover, in some cases, the data in each separate view may not be compatible with the others [14]. These differences pose a challenge to integrate information from complex distribution and diversified heterogeneous features of multi-view data to obtain the true categories of the data. If a clustering approach is unable to cope appropriately with multiple views, these views may even degrade the performance of MVC. On the other hand, the clustering approach needs to elaborately design a way to maximize clustering quality within each view, while simultaneously making clustering results across different views as consistent (agree with each other) as possible. When multiple sets of features are available for each individual sample, how to express the relationship of multiple views and how to integrate these views to identify essential grouping structure are two important questions needed to be answered by MVC [3]. In addition, the multi-view data collected in the real world are often incomplete (missing samples or features), uncertain or dynamic. This is due to the complexity in data collection and transmission. The challenges introduced by missing samples or features, uncertainty and dynamic changes add complexity to MVC. Therefore, incomplete multi-view clustering (IMVC), uncertain MVC and dynamic MVC have also been developed.

In order to meet these challenges, many research efforts have been devoted and a number of MVC approaches have been developed, such as co-regularized multi-view spectral clustering [15], nonnegative matrix factorization (NMF)-based MVC [16], and bipartite graph-based multi-view spectral clustering [17, 18]. These approaches employ different techniques to tackle the MVC issues from different viewpoints. In addition, several survey studies have been conducted to investigate the theories and techniques of existing MVC approaches [3, 6, 11, 19], where [11] reviewed generative and discriminative approaches, elaborated on the relationships between MVC and several closely related learning approaches (multi-view representation, ensemble clustering, multi-view supervised and semi-supervised learning, and multi-task clustering), and introduced several real-world applications of MVC. [3] classified MVC approaches into five categories, i.e. co-training style approaches, multi-kernel learning, multi-view graph clustering, multi-view subspace clustering, and multi-task multi-view clustering, provided a few examples for each category of approaches, as well as listed some publicly available multi-view datasets. [6] summarized graph-based approaches, space-learning-based approaches, and binary-code-learning-based approaches. [19] reviewed the existing studies on approaches for IMVC, categorizing them into MF-based IMVC, kernel learning-based IMVC, graph learning-based IMVC, and deep learning-based IMVC. The study then selected representative IMVC approaches for an experimental comparative analysis. [3, 6, 11, 19] also examined into various open problems that may necessitate further investigation. These include challenges related to large-scale data (size and dimension), complex data with noises or mixed types, imbalance information, missing view/value recovery, local minima, and deep learning, among others.

However, these survey studies mainly focus on approaches for complete multi-view data (where each feature is collected and each sample appears in each view), or those for incomplete multi-view data with missing samples or features. They do not explore approaches designed to handle both complete and incomplete data at the same time. Moreover, these studies tend to overlook approaches for uncertain multi-view data and dynamic multi-view data. In addition, these survey studies predominantly focus on approaches proposed before 2019. Many novel and important approaches developed after that, such as deep learning-based approaches [20, 21], have not been considered by existing survey studies. This gap makes it challenging to comprehensively track the current progress of the field.

Furthermore, among these survey studies, only [6] and [19] provided experimental quantitative evaluations. In [6], eight approaches (one k-means, three graph-based, three space-learning-based, and one binary-code-learning-based) were tested on seven complete multi-view datasets. [19] executed evaluations for seventeen approaches (two single-view, eight MF-based, one adaptive neighbors-based, two kernel learning-based, and four graph-based) on five incomplete multi-view datasets. Neither [6] nor [19] conducted experimental evaluations for the deep learning-based approaches. As a result, although a number of MVC approaches have been proposed, and several reviews assessing these approaches have been conducted, there is still a lack of comprehensive review and quantitative measurement and comparisons of MVC approaches. Especially noteworthy is the lack of evaluation for the approaches developed after 2019.

Hence, there is a need for a thorough review and quantitative assessment of MVC approaches, particularly those employing deep learning techniques. Notably, these crucial approaches have been overlooked in existing survey studies. Thus, this paper focuses on investigating MVC approaches, including those proposed after 2019, which have

not been reviewed by existing survey studies. Additionally, this paper conducts an experiment-based evaluation study for the most representative MVC approaches. The aim is to provide a quantified assessment of representative MVC approaches. This assessment will provide researchers with a comprehensive understanding of various approaches, while offering practitioners measurable insights to guide their selection of suitable methods for specific applications.

To achieve this objective, we first analyze and formalize basic concepts by introducing common key techniques for MVC. This includes the introduction of multi-view data, the definition of the MVC problem, principles related to MVC, strategies for fusing information and weighting views, the MVC routine, and the model structure, as well as the optimization scheme. Subsequently, we propose a novel taxonomy of MVC approaches and present the characteristics of representative MVC approaches. The MVC approaches are classified into four categories: complete MVC, incomplete MVC, uncertain MVC, and dynamic MVC approaches, based on the data types they handle. Additionally, the approaches for complete MVC and incomplete MVC are further categorized into eight and six sub-categories, respectively, based on their adopted working mechanisms and techniques.

Moreover, we summarize commonly used datasets and performance metrics for evaluating clustering performance in the field of MVC. The representative MVC datasets are divided into five categories: text, image, text-gene, image-text, and video datasets. The performance metrics include internal indices, which evaluate a clustering algorithm by summarizing results in a single quality score, and external indices, which evaluate the clustering result by comparing it with externally supplied true labels.

Finally, we conduct an empirical evaluation on thirty-five representative MVC approaches using seven real-world benchmark datasets. The thirty-five MVC approaches selected for our experimental evaluation studies include twenty-nine complete MVC approaches and six incomplete MVC approaches. These approaches cover the main categories in the taxonomy proposed in this paper, each characterized by distinct structures or constraints.

The seven datasets utilized in our study, each with varying scales, comprise three text datasets, three image datasets, and one image-text dataset. These datasets are widely recognized as typical in the MVC communities. Additionally, these seven datasets, serving as well-known benchmarks, represent different typical application scenarios, each with distinct views, classes, instances, features, and feature dimensions.

The clustering approaches and datasets have been meticulously selected to investigate the clustering performance of various approaches, validate the factors affecting clustering performance, and test the ability of various approaches to handle different data scales, ensuring the validity and relevance of this evaluation study. Furthermore, to support researchers and practitioners conveniently, we have compiled related references, implementations, and datasets on GitHub[1].

Our contributions are summarized as follows.

- We formalize and analyze basic concepts and common key techniques for MVC, which provides the background knowledge for understanding MVC and its related issues.

1 https://github.com/dugzzuli/A-Survey-of-Multi-view-Clustering-Approaches

- We propose a novel taxonomy of MVC approaches based on the detailed analysis of existing approaches, which, offering a comprehensive picture of MVC approaches developed.
- We provide a critical review on the working mechanisms and characteristics of the representative MVC approaches, and summarize representative MVC datasets and performance metrics commonly used in the MVC assessment and evaluation.
- We present the current progress of MVC, addressing the existing gap that the latest approaches proposed after 2019, particularly the deep (contrastive) learning-based MVC approaches, have not been evaluated in existing survey studies.

-We selected thirty-five representative MVC approaches based on the proposed taxonomy to conduct an empirical evaluation on seven real-world benchmark datasets, exploring how they perform across different types of datasets. The experimental results reveal that most MVC approaches struggle with large-scale datasets. No MVC approach consistently maintains high performance across all types of datasets. Factors such as model structures, regularization constraints, and weights corresponding to different views contribute to improving clustering performance. For example, the deep learning-based approach with multilevel representations and adversarial regularization performs well across many datasets. These findings provide valuable insights for future practitioners, offering information on the performance features of existing approaches. This serves as a practical guide for the development of applications, providing empirical evidence for selecting suitable approaches in specific circumstances.

The rest of the paper is organized as follows: In section 2, we analyze and introduce the basic concepts and some common strategies used in MVC approaches. Following that, we provide a novel taxonomy of MVC approaches. In the next four sections, we will present the working mechanisms and characteristics of representative MVC approaches proposed in recent years, each corresponding to the categories outlined in the proposed taxonomy. Afterward, we review the datasets and performance metrics used in the literature on MVC evaluation in Sections 7 and 8, respectively. We then present the empirical results and analyses in Section 9. Finally, we discuss potential directions for future research in Section 10 and conclude this study in Section 11.

2 PRELIMINARIES

In this section, we introduce categories of multi-view data, the definition of MVC, two principles of MVC, common strategies for solving MVC, and propose a new taxonomy of MVC approaches.

2.1 Multi-view data

Multi-view data refers to the same sample is described from various perspectives, with each perspective capturing a class of features known as a view [4] [22]. Figure 1 illustrates four intuitive examples of multi-view data, where (a) a person's iris, fingerprint, and face; (b) multi-wavelength whirlpool galaxy; c) car images taken from different viewpoints; and d) "thank you" described in many languages. Although these different features may have distinct physical meanings, take different forms as well as exhibit heterogeneous and diverse properties, they all represent the

same sample from different perspectives [3, 23]. Compared to single-view data that describes samples from a single perspective, multi-view data is semantically richer, more informative and diverse, but also more complex to analyze.

Multi-view data can be categorized as either complete or incomplete, depending on whether all expected perspectives or views are available for analysis. In incomplete multi-view data, only partial features could be available in some views for some data samples (missing features/values), or some data samples could be missing their whole observations in some views (i.e., missing samples/views), or even all samples of a cluster are not observed in one view (missing clusters) [24] [25]. Missing cluster and missing sample can be considered as one special form of missing feature. Figure 2 illustrates an example of incomplete multi-view data. In the figure, samples within the same cluster are represented by the same shape, distinguished by color. The dotted shape indicates missing samples, and 'NA' represents missing feature values. The incomplete multi-view data is ubiquitous in the real world due to sensor failure, equipment malfunction, data corruption, and other factors.

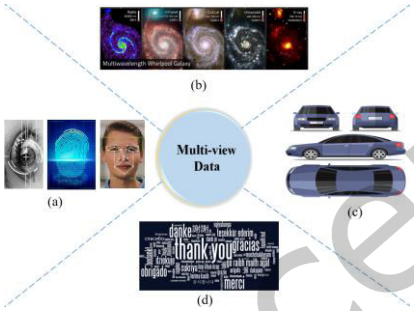

Figure 1: Examples of multi-view data

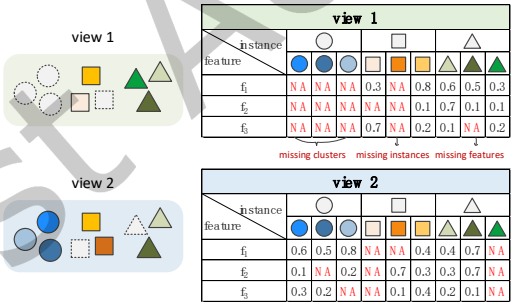

Figure 2: An example of incomplete multi-view data

Multi-view data may also be certain or uncertain. Certain multi-view data implies that the features of each sample and their values are affirmative, while uncertain multi-view data refers to the data that deviates from the desired distribution or pattern due to randomness, privacy issue, and imprecision in measurement [26]. In real-world applications, data collected from various sources like smartphones, satellites, and transportation typically exhibit uncertainty, often quantified through probability distributions.

In addition, in many practical applications, the number of views dynamically changes over time rather than remaining fixed. For example, in brain-computer interface systems, the acquisition of brain signals plays a crucial role in analyzing brain states. Since the brain signal changes with the object's mental state, it becomes necessary to collect the signal at different times. Therefore, the signal data at each time serves as a view, and the number of views changes dynamically over time [27].

2.2 The problem definition

Let $\boldsymbol{X} = \{\boldsymbol{X}^1, \boldsymbol{X}^2, \cdots, \boldsymbol{X}^m\}$ be a dataset with m views and $\boldsymbol{X}^v = \{\boldsymbol{x}_1^v, \cdots, \boldsymbol{x}_n^v\} \in \Re^{d_v \times n}$ be the v-th view data, where d_v is the feature dimensionality of the v-th view, and n is the number of data samples, $\boldsymbol{x}_i^v$ represents the feature vector of the i-th sample in the v-th view.

Let $\boldsymbol{C} = \{C_1, C_2, \cdots, C_K\}$ be a set of K clusters, called as a cluster structure, where C_j is the set of $|C_j|$ samples assigned to the j-th cluster, $\sum_j |C_j| = n$. Let $\boldsymbol{Y} \in \Re^{n \times K}$ be a membership matrix, whose (i, j)-th entry y_{ij} represents the probability that the i-th sample belongs to the j-th cluster. The sum of each row entries of $\boldsymbol{Y}$ should be 1 to make sure each row is a probability, i.e. $\sum_j y_{ij} = 1$. If only one entry of each row is 1 and all others are 0, it is a hard clustering (each sample can only be assigned to one specific cluster with the probability of 1), otherwise it is a soft clustering (each sample is assigned to one cluster with some probability between 0 and 1).

Given $\boldsymbol{X} = \{\boldsymbol{X}^1, \boldsymbol{X}^2, \cdots, \boldsymbol{X}^m\}$, the purpose of MVC is to divide n samples into K clusters by integrating the heterogeneous features of $\boldsymbol{X}$ without any label information, such that data samples within the same cluster are more similar than those in different clusters. That is, finally we will get a membership matrix $\boldsymbol{Y}$ or a cluster structure $\boldsymbol{C} = \{C_1, C_2, \cdots, C_K\}$ to indicate which samples are in the same group while others are in other clusters.

The key questions in MVC include: 1) How to effectively utilize the otherness information amongst different views in a multi-view dataset? 2) how to completely discover the consistency information between the different views? 3) how to reduce computation and space complexities? 4) how to enhance the robustness against noise and outliers? and 5) How to prevent being trapped in suboptimal local minima/maxima? Both the otherness and consistency information in multi-view data are very useful for effective clustering analyses. Multi-view data usually lies in high-dimensional space, where redundant and irrelevant features may result in the curse of dimensionality. Meanwhile, data often contain noise and outliers, which may destroy the underlying clustering structure. Moreover, approaches for solving MVC are usually non-convex, making them prone to becoming stuck into suboptimal local minima, especially when there are outliers and missing data.

2.3 Principles related to MVC

There are two significant principles ensuring the effectiveness of MVC: consensus and complementary principles [4]. The consistent of multi-view data means that there is some common knowledge across different views (e.g., both two

pictures about dogs have contour and facial features), while the complementary of multi-view data refers to some unique knowledge contained in each view that is not available in other views (e.g., one view shows the side of a dog and the other shows the front of the dog, these two views allow for a more complete depiction of the dog). Therefore, the consensus principle aims to maximize the agreement across multiple distinct views for improving the understanding of the commonness of the observed samples, while the complementary principle states that in a multi-view context, each view of the data may contain some particular knowledge that other views do not have, and this particular knowledge can mutually complement to each other.

Both complementary and consensus principles play important roles for improving the performance of MVL algorithms [28]. By exploring the consistency and complementary properties of different views, MVL is rendered more effective, promising, and exhibits better generalization ability than single-view learning [4].

2.4 The information fusion strategy

The strategies for integrating information from multiple views can be divided into three categories: direct-fusion, early-fusion, and late-fusion based on the fusion stage. They are also referred to as data level, feature level, and decision level fusion respectively, i.e. fusion in the data, fusion in the projected features, and fusion in the results. Direct-fusion approaches involve the direct incorporation of multi-view data into the clustering process by optimizing specific loss functions. Early-fusion combines multiple features or graph structure representations of multi-view data into a single representation or a consensus affinity graph across multiple views. Subsequently, any well-known single-view clustering algorithm, such as k-means, can be applied to partition data samples. Most approaches learn a graph structure representation for each view by deploying features of different views, then a consensus affinity graph is built. Some other approaches directly learn a common graph matrix from the original feature space. In contrast, the approaches of the late fusion first perform data clustering on each view individually and subsequently fuse the results for all the views to obtain the final clustering results through consensus [29]. The advantage of late-fusion is that it reduces the interference of other information channels to every separate partition, such that the effect of random noise can be reduced. Figure 3 illustrates the three fusion strategies. There is no theoretical foundation to decide which one is the best.

2.5 The clustering routine

There are two clustering routines, i.e. one-step routine and two-step routine, to execute MVC. The two-step routine first extracts the low-dimensional representation of multi-view data and then uses traditional clustering approaches, such as *k*- means, to process the obtained representation. In other words, the two-step routine often needs a post-processing process, such as applying a simple clustering method to the learned representation or conducting a fusion operation on the clustering results of individual views, to produce the final clustering results. This two-step learning strategy may lead to unsatisfactory clustering performance since the learning of the multi-view representation is not informed by the final clustering goal. The correlation between these two steps is not fully explored, potentially make the learned low-dimensional representation unsuitable for subsequent clustering tasks. In contrary, the one-step routine

integrates representation learning and clustering task into a unified framework, simultaneously learning a graph for each view, a partition for each view, and a consensus partition. Based on an iterative optimization strategy, high-quality consensus clustering results can be obtained directly and employed to guide the graph construction and the updating of basic partitions. This, in turn, contributes to the formation of a new consensus partition. Through joint optimization, co-training involves simultaneous clustering and representation learning, leveraging the inherent interactions between two tasks and realizing the mutual benefit of these two steps [1]. In the one-step routine, the cluster label of each data sample can be directly assigned and without the need any post-processing, reducing the instability of the clustering performance caused by the uncertainty of post-processing operations.

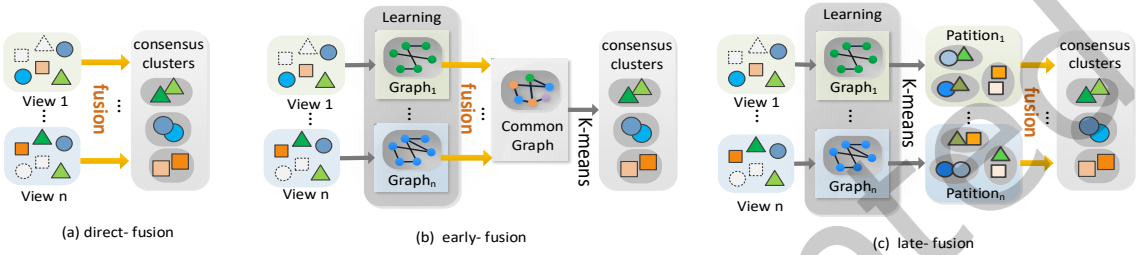

Figure 3: The illustration of three fusion strategies

2.6 The weighting strategy

In multi-view clustering, how to merge multiple views is a key issue. It is essential to assign an appropriate weight to each view based on its importance for the merging process. This ensures a deeper exploration of the complementary information present in heterogeneous data and effectively reduces the adverse effects of noise and outliers. The simplest approach is to assign the same weight to all views (equal-weighted) or not to consider the weight, treating all views equally and assuming their reliability. However, the clustering results may be degraded if different views are not distinguished [30]. In real-world clustering problems, data views inherently vary in strength, and each view possesses specific statistical properties while being susceptible to different forms of noise pollution. Thus, it is unreasonable to treat different views equally during the clustering process. To distinguish the different contributions of different views, auto-weighted (self-weighted) approaches, which involve automatically learning the weight of each view [31], have been proposed. Auto-weighted approaches can be classified into two categories: parameter weighted and parameter-free weighted strategies. Parameter weighted approaches learn the weight of each view by introducing additional hyper parameters, which control the smoothness or sparsity of the weight distribution. However, setting these hyper parameters in the clustering task is often challenging due to the need for an extensive search in a large parameter space. The clustering results can be sensitive to these hyper parameters, and the optimal values may vary across different datasets. The parameter-free weighted strategy automatically assigns weights by a self-conducted weight learning, without the requirement of any hyper parameters while maintaining precision[32].

Except for weighting the view, [33] learned the weights of different features in each view and the weights of each sample in different views by introducing a feature-level and a sample-level attention mechanism. This approach

hierarchically distinguishes the weights of different features in one view and the weights of the same sample in different views. Using a unique weighting strategy, [34] considered the confidence of both views and samples under the assumption that samples may have different confidence levels under the same view. [35] assigned weights to the views of data samples and feature representations in each view, emphasizing discriminatory features and views over others. [36] learned the cluster-wise weights instead of view-wise weights, with the cluster-weighted scheme enhancing the interpretability of the clustering results. [37] learned automatically the view weights based on the concept of mutual information and then imposed simultaneously them on the content-based and context-based multi-view data representations. [38] measured the weights of the previous views and the last view when the number of views increases over time.

2.7 The model structure

The MVC models consist of shallow structure or deep structure. Models with shallow structure learn the low-dimensional representation of multi-view data via a one-level structure, ignoring the hierarchical and non-linear structural information hidden in each view. For example, classical nonnegative matrix factorization (NMF) only factorizes the data matrix $\boldsymbol{X}$ into two nonnegative factor matrices $\boldsymbol{U}$ and $\boldsymbol{V}$, such that $\boldsymbol{X} \approx \boldsymbol{U}\boldsymbol{V}$, which may limit its ability to learn higher level and more complex hierarchical information. Models with deep structure learn a low-dimensional representation of multi-view data via a multi-level structure, enabling them to capture complex hierarchical and structural information. For example, the deep semi-NMF model [39] factorized data matrix $\boldsymbol{X}$ into $l+1$ factors $\boldsymbol{X}^{\pm} \approx \boldsymbol{U}_1^{\pm}\boldsymbol{U}_2^{\pm}\cdots\boldsymbol{U}_l^{\pm}\boldsymbol{V}_l^{+}$, which allows for a hierarchy of l layers of implicit data representations that can be given by the factorizations: $\boldsymbol{V}_{l-1}^{+} \approx \boldsymbol{U}_l^{\pm}\boldsymbol{V}_l^{+}, \cdots, \boldsymbol{V}_2^{+} \approx \boldsymbol{U}_3^{\pm}\cdots\boldsymbol{U}_l^{\pm}\boldsymbol{V}_l^{+}, \boldsymbol{V}_1^{+} \approx \boldsymbol{U}_2^{\pm}\cdots\boldsymbol{U}_l^{\pm}\boldsymbol{V}_l^{+}$.

2.8 The optimization scheme

In general, MVC is an NP-hard optimization problem. The most commonly used solution to this problem is the alternating iterative optimization scheme, which decomposes the problem into several tractable sub-problems. Each variable or group of variables is updated alternately while keeping the others fixed. For example, [40] solved the deep multi-view concept learning (DMCL) model using block coordinate descent [41] that each time optimizes one group of variables while keeping the other groups fixed. The augmented Lagrange multiplier [42] and alternating direction method of multipliers (ADMM) [43] are widely used to solve the convex optimization problem.

2.9 The proposed taxonomy

In this paper, we propose a novel taxonomy according to the data types that MVC approaches deal with. This taxonomy encompasses approaches for complete, incomplete, uncertain, and dynamic multi-view data, denoted as Complete MVC, Incomplete MVC, Uncertain MVC, and Dynamic MVC for brevity. According to the working mechanisms and techniques that various approaches adopted, the approaches for complete multi-view data are further divided into eight sub-categories: (1) NMF-based approaches, (2) multiple kernel learning-based approaches, (3) graph-based approaches, (4) subspace-based approaches, (5) deep learning-based approaches, (6) contrastive learning-

based approaches, (7) co-learning-based approaches, and (8) self-paced learning-based approaches. Similarly, the approaches for incomplete multi-view data are further divided into six sub-categories: (1) NMF-based approaches, (2) multiple kernel learning-based approaches, (3) graph-based approaches, (4) subspace-based approaches, (5) deep learning-based approaches, and (6) contrastive learning-based approaches. The approaches for uncertain multi-view data and those for dynamic multi-view data are not further classified into sub-categories, because there are not many approaches proposed in these two categories. The proposed taxonomy of MVC approaches is illustrated in Figure 4.

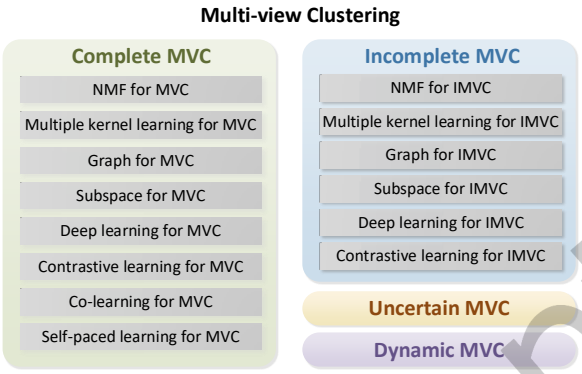

Figure 4: The proposed taxonomy of MVC approaches

The mechanisms and principles of the representative MVC approaches proposed in recent years are summarized in the next four sections. Note that, due to the space limitation, the tables summarizing the characteristics of various approaches are placed in Appendix.

3 COMPLETE MULTI-VIEW CLUSTERING

In this section, we present the working mechanisms of the representative MVC approaches for complete multi-view data and their characteristics. The general procedure of the approaches for Complete MVC is shown in Figure 5, while Table 1 (Appendix) summarizes the characteristics of the representative Complete MVC approaches. These characteristics include motivation, model structure, information fusion and weighting strategy, clustering routine, model peculiarity, and time as well as space complexity.

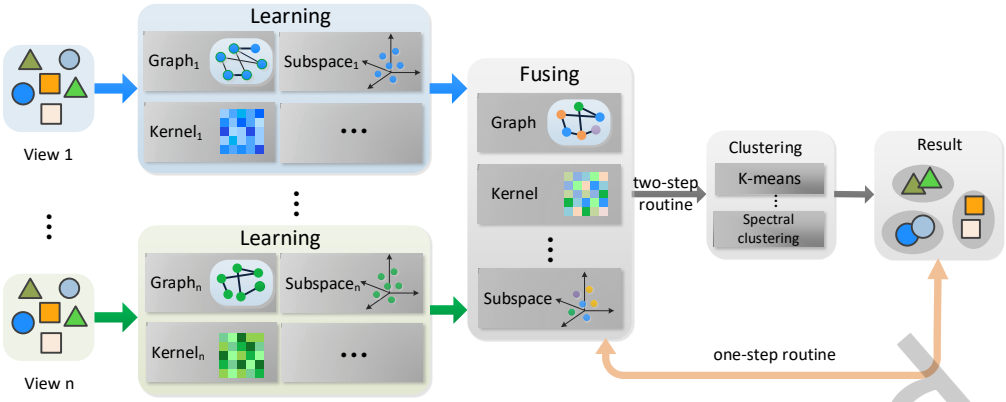

Figure 5: The general procedure of the Complete MVC approaches

3.1 NMF-based approaches for MVC

NMF-based multi-view clustering (MVC) approaches decompose a data matrix with nonnegative elements into two low-rank matrices. The product of these two low-rank matrices is designed to approximate the original data matrix. The nonnegative constraint results in a parts-based representation of samples, which accords with the cognitive process of the human brain from the psychological and physiological evidence. This characteristic makes the clustering results easy to be interpreted.

To obtain the desired dimensional-reduced representation, various constraints are integrated into traditional NMF. For example, [16] introduced the consensus constraint to push clustering solution of each view towards a common consensus. Additionally, [44-46] imposed the multi-manifold regularization to preserve the locally geometrical structure of the data space, [47] utilized the sparseness constraint to extract the robust feature of each view, and [48, 49] designed the orthogonality constraint to capture the intra-view diversity, in turn to obtain the desirable representations for each view. [50] exerted a graph Laplacian regularization on the indicator matrix learned via matrix factorization assisted *k*-means. This approach aims to capture the intrinsic geometric structure of original data. [51] preserved the geometric structures of multi-view data in both the data space and the feature space. [52] adopted auto-weighted collective matrix factorization (CMF) to extract shared information of multi-view data. Additionally, the approach imposed graph dual regularization terms with orthogonality constraints to preserve the geometrical structure of the decomposed factors.

To reduce the high dependency on the quality of the original views and recognize global relationships amongst data samples, [53] jointly factorized multiple networks transformed from multi-view data. The approach also incorporated sparse as well as multi-manifold regularization into NMF to keep the intrinsic geometrical information of the multi-view network manifold space.

To accelerate computational efficiency and decrease memory costs, [21] employed NMF to the embedded anchor graph, and utilized correntropy to increase clustering robustness. [54] divided the optimization problem into three decoupled small-scale problems containing only a small amount of matrix multiplications. [55] exploited a constrained

binary matrix factorization to achieve direct clustering. [56] encoded the multi-view image descriptors into a compact common binary code space and clustered the collaborative binary representations. This process is designed to reduce computation costs and storage requirements through bit-operations. [57] developed an orthogonal mapping binary graph approach to eliminate redundant information and extract local geometric structure information of binary codes.

To capture the complex hierarchical and nonlinear information, [58] introduced deep concept factorization (CF) into MVC. [59] utilized deep matrix decomposition to obtain the partition matrix of each view. [60] constructed a multilayer NMF model with graph regularization to extract abstract representations. The last layer representation from each view was then derived to form a common consensus representation. [61] designed multiple encoder components and decoder components with deep structures to hierarchically factorize the input data. All encoder and decoder components were then integrated at an abstract level to capture heterogeneous information across multi-view data. [62] designed diversity embedding deep matrix factorization to obtain discriminative features and reduce feature redundancy. [63] adopted semi-NMF to learn the hierarchical semantics of multi-view data in a layer-wise fashion. In this process, the nonnegative representation of each view in the final layer was enforced to be the same, maximizing the mutual information from each view. [40] performed hierarchically nonnegative factorization for capturing semantic structures. Additionally, the approach explicitly modeled both consistency and complementary information at the highest abstraction level.

3.2 Multiple kernel learning-based approaches for MVC

The multiple kernel learning (MKL)-based approaches linearly or non-linearly combine predefined kernels corresponding to different views to improve clustering performance. These predefined kernels map samples from the original low-dimensional space to a high-dimensional space, such that the samples are linearly separable in high-dimensional space. Due to the effectiveness of handling non-linear data and avoiding the selection of specific kernel function (in general, it is very time-consuming and expensive to select the most suitable kernel from a pre-specified pool of base kernels, e.g., Linear kernel, Polynomial kernel, and Gaussian kernel), MKL-based approaches for MVC have been widely investigated and achieved promising results [64].

[65, 66] learned similarity relationships in kernel spaces to improve the robustness against noise. [67] jointly learned kernel representation tensor and affinity matrix. [68] employed kernel-induced functions to mapped the multi-view data from linear space into nonlinear space for utilizing the nonlinear structure hidden in data. These approaches automatically appointed a reasonable weight for each view.

3.3 Graph-based approaches for MVC

Graph is an important data structure for representing the relationships hidden in multi-view data. Each node in a graph corresponds to a data sample and each edge represents a relationship between two samples. Graph-based clustering approaches seek to learn a consensus affinity graph across all views and then carry out graph-cut algorithms or other techniques (e.g., spectral clustering) on the consensus affinity graph to produce the clustering result. The consensus affinity graphs can be directly learned in original feature space, or obtained by fusing multiple candidate graphs

learned in original feature space separately. It is obvious that the clustering performance highly depends on the quality of the consensus affinity graph, thus it is a crucial task to learn a high-quality consensus affinity graph. Usually, the difference amongst various multi-view graph clustering approaches vary in how they learn the consensus affinity graph across the multi-view data.

To obtain a high-quality consensus affinity graph, [20, 69-72] projected the multi-view data into a low-dimension shared latent embedding space. They then learned the consensus affinity graph from the shared latent embedding space to reduce the corruption and redundancy of the original views. [73-75] employed manifold learning and sparse representation to construct the graph of each view and fused them in an automatically weighted way. [76-78] introduced tensor nuclear norm to minimize the divergence between graphs of different views. [79] introduced a consistent smoothness constraint of overall views and an orthogonality constraint. [80] converted multi-view fuzzy clustering to adaptive graph learning with sparse low-rank constraints to ensure its strong discriminative ability. [81] weighted the different views in terms of their confidence.

To reduce the computational complexity, [82] relaxed the constraint of the global similarity matrix. [83] devised a refined version of *k*-nearest neighbor graph to keep data points and aggressively reduced the number of edges. [84, 85] developed MM (Majorization-Minimization)-based optimization approaches. [86, 87] jointly optimize the learning of consensus graph and discretization of cluster labels. [88] learned a structured graph to directly extract the clustering indicators, without performing other discretization procedures.[89] constructed a bipartite graph to depict the relationship between samples and anchor points, and imposed a connectivity constraint to guarantee that the connected components indicate clusters directly.

3.4 Subspace-based approaches for MVC

Subspace-based approaches for MVC aim to directly learn a unified low-dimensional representation shared by multiple views from multiple subspaces or a pre-learned latent space. This is based on the assumption that the multi-view observations are generated from an underlying latent representation [90]. Subsequently, *k*-means or spectral clustering is applied to the unified representation to divide data samples lying in a union of multiple low-dimensional subspaces into different clusters. This process ensures that samples in the same cluster come from one subspace. The key challenges for subspace-based multi-view clustering approaches include learning a robust representation, addressing data with nonlinear structures, and enhancing computation efficiency.

To learn a robust representation, [91] employed an exclusivity constraint term to enhance the diversity of specific representations amongst different views. Additionally, the approach imposed a clustering structure constraint on the learned subspace self-representation to obtain a clustering-oriented subspace self-representation. [92] built an anti-block-diagonal indicator matrix, incorporating a small amount of supervisory information. This was done regularize the shared affinity matrix corresponding to the latent representation for ensuring its block-diagonal structure. [93] defined a local graph regularization term about the consensus latent subspace representation. This term was designed to preserve the manifold structure of data and ensure consistency across different views. [94] employed the maximum dependence constraint between the similarity matrix and its latent intact (complete and not damaged) points to build

an similarity matrix. [95] designed the multiplicative decomposition scheme and the variable splitting scheme to extract the components from their corresponding view-specific coefficients, where inconsistent elements are filtered out. [1] fused multi-view information in a partition space to reduce the effect of noise. [96] adopted the view-consensus grouping effect and low-rank constraint via the nuclear norm to regularize the view-commonness representation. [97] captured the correlations between shared information across multiple views and employed view-specific information to describe specific property of each independent view. [98] adopted the weighted nuclear norm (instead of nuclear norm) to approximate the rank of the common coefficient matrix. [99] hired the self-attention mechanism to derive dynamic weights of different views for fusing consistent and view-specific information from multiple views.

To capture the nonlinear nature of data, [100, 101] exploited a low-rank kernel mapping and the non-convex Schatten p-norm regularizer as well as the correntropy to learn a joint subspace representation of all views. [102-104] projected data from original space into a kernel/tensor space to encode the low-rank property of the self-representation tensor. [105] merged the self-representative subspaces of different views on a Grassmann manifold. [106] [107] utilized kernel trick and kernel dependence measure to encode complementary information from different views. [108] employed deep matrix factorization to obtain multi-view multi-layer low-rank subspace representations. [109-112] employed the deep convolutional/auto-encoder to learn the latent low-dimensional hierarchical representation. [113] [114] employed auto-encoder networks on multiple views to achieve multi-level latent smoothness.

To improve the computation efficiency, [90] directly recovered the row space of the latent representation without the graph construction procedure. [115] jointly optimize the anchor selection and subspace graph construction in a parameters-free way.

3.5 Deep learning-based approaches for MVC

Deep learning-based MVC approaches utilize neural networks to learn latent representations, and then divide data sample into different groups based on the learned latent representations. Because the learned latent representations can reveal the potential peculiarities of complementarity and consistency information amongst multi-view features, such as the mutual agreement, nonlinear relationships, thus they effectively mitigate the effect of noise and dimensionality curse.

[116] utilized auto-encoders to learn latent representations shared by multiple views, and leveraged adversarial training to further capture the data distribution and disentangle the latent space. [117-119] used auto-encoders to learn individually the embedded representations of multiple views with the consideration of both consensus and complementarity of multiple views. [120] employed various auto-encoders, e.g., stacked autoencoder (SAE), convolutional autoencoder (CAE) and convolutional variational autoencoder (Conv-VAE) to exact multi-view features.

To encode graph structure and node attribute to the node representation, [121] utilized graph convolution network (GCN) with block diagonal property to encode the discriminative information. Additionally, they utilized Euler transform to augment the node attribute as a new view descriptor for non-Euclidean structure data. [122] adopted graph

neural network (GNN) to implement dual fusion-propagation for capturing the multiple information amongst different views. [123] built a neural network composed of multiple blocks to learn sparse regularizers.

To distinguish the importance of different views and semantics, [124] adopted adversarial learning and attention mechanism to align the latent feature distributions of different views and quantified the importance of modalities. [125] adopted the self-attention mechanism to calculate the alignment matrix for capturing the category-level correspondence of the unaligned data. [126] developed a differentiable bi-level optimization network to enhance the interpretability of deep MVC.

3.6 Contrastive learning-based approaches for MVC

Recently, contrastive learning, aiming at maximizing the agreement between positive sample pairs and minimizing that between negative sample pairs, has been widely used in MVC. This is attributed to its excellent ability to capture the consistency of multiple views. It offers a way to align representations from different views at the sample level, forcing the label distributions to be aligned as well [127].

[128] adopted contrastive learning to infer the inter-cluster relationships and intra-cluster boundaries from the local context of each node. [129] used contrastive learning to train GCN for integrating the topological structures and node features of neighbors. [130] used contrastive learning to align sample-level representations across multiple views for capturing the view-invariance information. [131] used contrastive learning to achieve the consistency objectives for the high-level features and the semantic labels. [132] respectively exploited a cross-view contrastive learning and a mutual contrastive teacher-student learning. These approaches aimed to obtain a redundancy-free consistent representation at the instance level and capture the intra-view discriminative information at semantic level. [133] designed instance-level and cluster-level contrastive learning to exploit the representations of the augmented weak-weak view pair and the strong-weak view pairs.

To improve the performance, [134] developed a contrastive fine-modeling via maximizing the similarity of pair-view to guarantee the consistency of multiple views. [135] designed a graph contrastive loss to regularize the learning of the consensus graph. [136] introduced a contrastive reconstruction loss to realize sample-level approximations between the reconstructed graphs and the raw graphs. However, [137] proved that contrastive alignment can be detrimental to the clustering performance, especially when the number of views increases.

3.7 Co-learning-based approaches for MVC

Co-learning approaches aim to improve the clustering performance by exchanging information amongst different objects. Co-learning approaches include co-regularization MVC, co-training MVC, multi-task MVC, and co-clustering MVC. The general procedures of co-training, multi-task, and co-clustering MVC are shown in Figure 6.

The typical representatives of co-regularization MVC and co-training MVC are CoRegSC (co-regularization spectral clustering) [15] and CoTrainSC (co-training spectral clustering) [138]. CoRegSC formally measured the agreement on distinct views and solved the corresponding objective problem to make the cluster structures in different views agree with each other. CoTrainSC employed separate but correlated learners on each view to acquire the cluster

structure. The acquired cluster structure then guided the learners in other views, facilitating the exchange of information among different views. This allowed the disagreement amongst views to be propagated back to the learners, helping them learn a more accurate cluster structure by minimizing the disagreement in the next iteration. Through an iterative alternate training procedure, the cluster structures of multiple views tended towards consensus. [139] pointed out that the success of co-training algorithms mainly relies on three assumptions: (a) sufficiency - each view is sufficient for classification on its own, (b) compatibility - the target functions of both views predict the same labels for co-occurring features with a high probability, and (c) conditional independence - views are conditionally independent given the label. [140] implemented collaborative learning between visible and hidden views to combine the individual information and the shared information in different views.

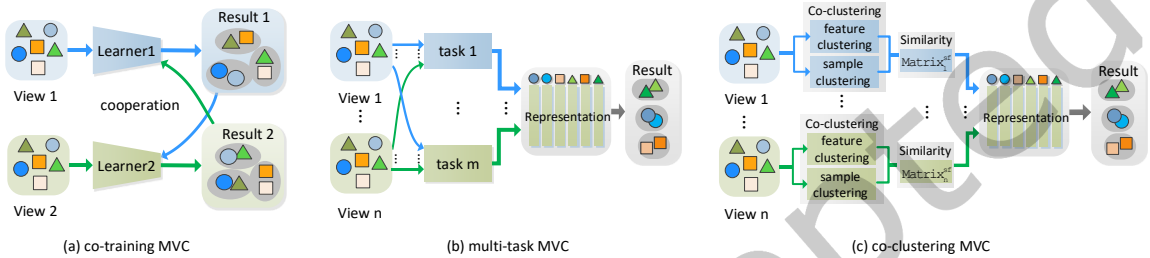

Figure 6: General procedures of co-training, multi-task, and co-clustering MVC

Multi-task clustering improves individual clustering performance by learning the relationship amongst related tasks, while MVC makes use of consistency amongst different views to achieve better performance. Multi-task MVC (MTMVC) means the tasks are closely related and each task can be analyzed from multiple views. The goal of MTMVC is to generate a consensus partitioning for every task by exploiting the shared relationships between the views within a task and between different tasks [141]. [141] formulated the MTMVC problem as a multi-objective optimization problem. [142, 143] developed a semi-nonnegative matrix tri-factorization based MTMVC clustering algorithm to deal with the data with negative feature values.

Co-clustering refers to conducting two-sided clustering along the samples and features simultaneously under the assumption that samples exhibit a pattern only under a subset of features [14]. Co-clustering MVC combines MVC with co-clustering for taking advantage of the diversity of features provided by multiple sources and the dual relationship between the feature space and sample space.

To employ the agreement and disagreement amongst views, [144] shared a common clustering results along the sample dimension and kept the clustering results of each view specific along the feature dimension. The mechanism of maximum entropy was leveraged to control the importance of different views. [145] built a view-level bipartite graph to draw the co-occurring structure of data for exploiting the duality between samples and features of multi-view data.

To improve the robustness against noisy features, [146] employed the dynamic multi-view co-clustering algorithm with mutual information. This approach was employed to learn the view weights, which was then applied to the discriminative feature representations of multiple views, instead of the original representations. [147] employed multiple co-clustering algorithms to calculate the sample-sample, feature-feature, and sample-feature similarity information of each view data. [14] used co-clustering as the basic clustering block of MVC, and utilized the features and view weighting schemes to iteratively specify the perceived features and view importance.

In addition, [148] conducted co-clustering based on the matrix factorization under the constraints of indicator matrix to improve the computational efficiency. [149] developed a differentiable deep network to learn interpretable and consistent collaborative representation from multi-source features, and maintain sparsity between multi-view feature space and single-view sample space.

3.8 Self-paced learning-based approaches for MVC

Self-paced learning, an effective technique to avoid bad local minima and improve the generalization result, simulates human learning process. It starts by training a model on 'easy' examples which have smaller loss values and then gradually takes 'complex' examples into consideration. The general self-paced learning model is composed of a weighted loss term on all examples (with higher losses) and a regularizer term imposed on example weights. By gradually increasing the penalty on the regularizer during model optimization, more examples are automatically included in training from 'easy' to 'complex' via a pure self-paced approach.

[106] employed low-rank tensor constraint to assign different weights on different singular values of the representations, and used the self-paced learning to treat multiple instances differently. [150] exploited a self-paced and auto-weighted strategy to reduce the risk of trapping into bad local optima. [151] applied a soft-weighting scheme of self-paced learning for instances to mitigate the negative impact of noises and outliers. Additionally, they designed a self-paced feature selection manner and a weighting term for views to alleviate the feature and view quality issues.

3.9 Discussion

Although numerous MVC algorithms have been proposed, there is no criterion to decide which MVC algorithm is the best due to the unique merits of each approach. In summary, NMF-based approaches provide straightforward interpretability for clustering results, i.e., each observation can be explained as an additive linear combination of nonnegative basis vectors. However, a challenge lies in limiting the search of factorizations to those that can give meaningful and comparable clustering solutions across multiple views simultaneously. Graph-based approaches can learn the correlation and complex structures hidden in data, but the performance is highly dependent on some prior factors. Subspace learning is effective in reducing the "curse of dimensionality", but they have initialization dependence. Deep learning-based approaches can explore the relationships amongst different samples and avoid the possible corruption as well as the curse of dimensionality, but many deep learning-based MVC approaches involve more parameters and lack theoretic interpretability. Co-clustering can use samples to induce feature clustering and use features to induce sample clustering based on the duality, but co-clustering-based MVC approaches usually suffer from

high computational and storage complexities. Self-paced learning approaches avoid bad local minima, but it is difficult to distinguish 'easy' and 'complex' examples.

Several MVC approaches integrated different techniques to enhance the reliability of clustering results via inheriting the merits of different approaches. For example, [152] exploited several candidate multi-view clustering to maximize the worst-case performance gain against the best single view clustering. This ensures that the clustering performance, when utilizing multiple views, is never statistically significantly worse than that achieved by using a single view alone. [153] designed a self-tuning MVC approach that introduced a sum-of-norm loss function, regularization, and statistics techniques to reduce the initialization sensitivity and automatically determine the cluster number.

4 INCOMPLETE MULTI-VIEW CLUSTERING

The problem of clustering incomplete multi-view data is known as incomplete multi-view clustering (IMVC) (or partial multi-view clustering) [154]. The goal of IMVC is to discover the common cluster patterns hidden in incomplete multi-view data based on the observed data samples and features in different views. In incomplete data, due to the degradation of the acquired data quality, the natural alignment property of same samples across multiple views and sample completeness may not be preserved [155]. This may leads to a serious of information loss, the information imbalance aggravation amongst different views [19] and the integration difficulties of multiple views [156]. Consequently, it results in the insufficient excavation of the complementary and consistent information [157]. IMVC faces more challenges compared to MVC. It needs issues such as mitigating the impact of missing information and effectively extracting and utilizing the underlying semantic information from missing views/features to cluster the multi-view data. The existing approaches for MVC cannot be directly applied to incomplete multi-view data.

An intuitive strategy to solve IMVC is data adaption, i.e., removing incomplete samples or filling incomplete views (imputation) to obtain complete multi-view data. Subsequently, existing MVC methods can be employed directly. The removing incomplete samples is simple and straightforward, but it loses original samples and may cause heavy information loss when most of the samples have missing values. Imputation methods include zero imputation, mean imputation, *k*-nearest-neighbor imputation, random imputation, regression imputation, expectation maximum (EM) imputation, multiple imputation (imputes the missing value multiple times, MI) and so on [158]. Because inaccurate interpolation or filling of missing data may induce noisy features, which may destroy the distribution of original data and damage the clustering performance [159], imputation-free approaches have been developed, for example, [160] dealt with the missing samples or features by introducing an sample-view indicator matrix to indicate whether an sample exists in a view or not.

In recent years, a number of approaches for IMVC have been developed. In this section, we present the working mechanisms of the representative IMVC approaches. The general procedure of the approaches for IMVC is shown in Figure 7, and the characteristics of the representative IMVC approaches are summarized in Table 2 (Appendix).

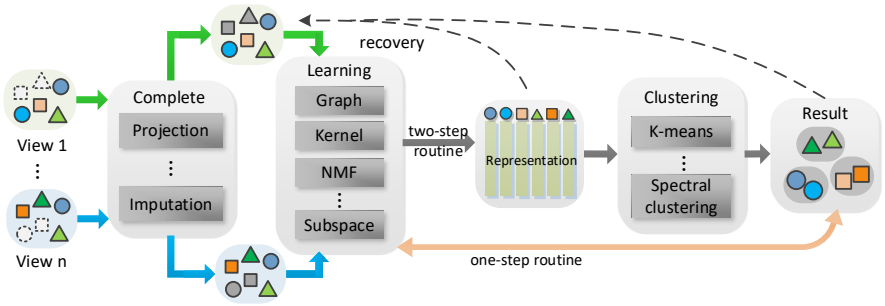

Figure 7: The general procedure of the IMVC approaches

4.1 NMF-based approaches for IMVC

To acquire a robust low-dimensional consensus latent representation for all views, [161-164] adopted graph-regularized matrix factorization model to reveal the geometrical structure of data. Specifically, [161] designed a semantic consistency constraint to reduce the influence of the unbalanced incomplete views, [162, 163] used orthogonal constraint to alleviate the problem of the inconsistent clustering structure, and [164] utilized the guidance of the virtual label to enhance the distinctiveness of learned consensus latent representation. [165] combined NMF with low-rank tensor to capture the higher-order and complementary information. [25] employed NMF to separately handle individual clusters/instances, and introduced locally geometrical information to reduce the negative impact caused by multi-view interaction. [166, 167] incorporated index matrices of missing samples into matrix factorization and introduced graph Laplacian regularization to promote the compactness of the low-dimensional representation. [168] utilized the results of NMF with the Laplacian regularization to reconstruct missing views, and the reconstructed views were also used to seek the latent representation. [158] adopted multiple imputations to deal with missing values under the consideration of the missing value uncertainty, and designed a view weighting strategy to ensemble the clustering results from multiple views.

To reduce the computational and memory costs, [169] employed the regularized matrix factorization and weighted matrix factorization, [170] combined NMF with the graph Laplacian of the cosine similarity matrix directly computed from the original data space.

4.2 Multiple kernel learning-based approaches for IMVC

[171] encouraged incomplete kernel matrices to mutually complete each other and integrated imputation as well as clustering into a unified learning procedure. [172] imputed each incomplete base matrix, and introduced prior knowledge (the consensus clustering matrix is required to lie in the neighborhood of a pre-specified one) to regularize the consensus clustering matrix. [173] dynamically generated a consensus proxy under the guidance of a shared cluster matrix to improve the effectiveness of imputation and clustering, and preserved sufficient kernel details via adjusting the size of base partitions for further improving the quality of the consensus proxy. [174] employed similarity graph to

guide the kernel completion for capturing the global structure and local nonlinear relationship of samples in kernel space.

4.3 Graph-based approaches for IMVC

To alleviate the effect of missing data, [175] designed a locality-preserved reconstruction term to infer the missing views. [176] exploited the intrinsic structure of data to complete the missing views. [162, 163] employed the cross-view feature transformation to complete the unobserved samples and incorporated the completion process into multi-view graph learning. [177] designed the feature space based missing-view inferring to recover the missing views. [178] designed a complete structure inferring strategy to learn the complete structures of all views for disclosing the real distribution of the absent samples under the reconstruction constraint. In addition, [179] transferred feature missing to similarity missing and then used average similarity values in other views to complete the missing similarity entries based on the spectral perturbation theory. [180] employed existing data to reconstruct incomplete data and incorporated the reconstruction objective with self-representation based spectral clustering. On the contrary, [181] combined a distance regularization term as well as low-rank representation-based non-negativity constraints to directly learn graphs from raw data without filling. [182] used the available data of each view to learn a corresponding view-specific partial graph, and designed a cross-view graph fusion term to learn a consensus complete graph for different views.

To distinguish the contributions from different views for alleviating influence of improper views to the quality of the fused consensus graph, [183] considered the view's importance to learn the affinity matrix of the view. [184] learned the sample-level auto weight to fuse graphs. [185] introduced adaptive weights to balance the importance of different views. [186] fused incomplete base partition matrices in an auto-weighted manner to generate a unified partition matrix. [187] adopted an adaptive weighting strategy to alleviate the negative impact of unbalanced incomplete views. In additional, they devised a local and global co-regularization to reveal the mutual effect between local clustering from different incomplete views and global clustering across all views.

To handle data with complex distributions and the non-linear information, [188] exploited spectral analysis to supervise the common representation extracted from all the views. [189] investigated the local information within each view as well as the semantic consistent information shared by all views. They introduced a co-regularization constraint to minimize the disagreement between the common representation and the individual representations with respect to different views. [190] built a similarity matrix to measure the relationships of present instances, and employed a spectral-based method to learn a common probability label matrix and low-dimensional representations of present samples. [191] employed the graph learning and spectral clustering techniques to learn the common representation. [192] incorporated the feature space based missing-view inferring and manifold space based similarity graph learning, and imposed a low-rank tensor constraint to capture the high-order correlations of multiple views. [193] employed projection learning to reduce the effect of the information imbalance between different views caused by the diversity dimensions, and imposed a graph regularization penalty term to capture the geometric structure of data.

In addition, [194] [195] designed improved anchor selection strategies to choose representative anchor points. These anchor points were crucial for learning the consensus instance-to-anchor similarity matrices for all views and constructing view-wise complete anchor graphs as well as the fused complete anchor graphs. [196] [197] utilized anchors to compute the similarities between all data points, and executed anchor-based spectral clustering to obtain the clustering result. [198] proposed a structural anchor-based similarity learning model to obtain the intra-view similarity matrix. They employed the paired anchor samples to compute the inter-view similarities, and then designed a complete anchor-inferred graph learning scheme to improve the efficiency and performance of the spectral clustering. [199] independently built the similarity matrix of each view via the adaptive neighbor assignment strategy to eliminate the necessary of adjusting parameters. Additionally, they employed a non-iterative approach to reduce the computational complexity.

4.4 Subspace-based approaches for IMVC

To alleviate the effect of missing data, [200] employed the affinity matrices learning and tensor factorization regularization to recover the missing views and the subspace structure. They introduced hypergraph-induced hyper-Laplacian regularization to preserve the high-order geometrical structure of data. [201] employed the available instances within views to infer the missing information, and adopted a weighted alignment of projection matrices corresponding to different views to learn a discriminative shared embedding. [202] employed a reconstruction term to recover missing samples from non-missing ones, and decomposed the self-expressiveness coefficient learned from the recovered complete multi-view into a consistent part together with a specific part. This approach aims to reveal the similarity information of view-paired and obtain the unique information of each view. [203] recovered the missing data based on shared latent representation and learned multilevel graphs of recovered views by self-representation. Additionally, it introduced a tensor nuclear norm regularizer to pursue the global low-rank property and explore both intraview and interview correlations.

To obtain more accurate subspace representation, [24] described multi-view data in a local manner to obtain clear block-diagonal structure for data distribution and more accurate subspace representation. [204] devised the soft block-diagonal-induced regulariser to fuse and construct a shared representation for all views. It also inserted multiple indicator matrices into the multi-view self-representation model to achieve clustering results. [205] utilized a tensor nuclear norm regularizer to diffuse the information of multi-view block-diagonal structure across different views. [206] utilized low-rank matrix factorization to obtain a consensus representation matrix. It then then combined with the objective function of nonnegative embedding and spectral embedding subspace clustering for joint optimization. [207] learned missing instances and self-representations in the latent space for capturing the features of missing instances and preserving the original latent spatial structure of the data. It imposed a completeness constraint to guarantee that learning direction of missing instances was close to the original data as possible. [208] learned a similarity graph for each view rather than a consensus graph, and dealt with the incompleteness of the views by fusing information from different views in partition space. [209] jointly performed data imputation and self-representation learning.

To improve robustness against the incompleteness and noises, [210] designed a view evolution scheme to deal with unbalanced incomplete view data (i.e. different views often have distinct incompleteness), and devised the low-rank representation to recover the data. [211] developed a weighted subspace learning mechanism with low rank and sparse constraints to capture relationship between the data samples.

4.5 Deep learning-based approaches for IMVC

[212] optimized multiple groups of decoder deep networks to obtain the completion of data view and multiple shared representations, so as to generate multiple clustering results with high diversity and quality. [213] utilized multi-view auto-encoder to infer the missing features of incomplete samples, and conducted adaptive graph learning as well as graph convolution to extract data structure. [156] developed the auto-encoder with the manifold alignment constraint and consistency alignment constraint, aiming to preserve the compact inherent local structure within the view and the consistency semantics between incomplete views. [214] combined discriminator networks with auto-encoders to learn common low-dimensional representations as well as the shared cluster structure across multiple views, and employed adversarial training to generate possible values of missing features. [215] designed view-specific encoders to extract the high-level information of multiple views, and introduced a self-paced strategy and a weighted fusion layer to select the most confident samples and obtain the consensus representation shared by all views. [159] [216] employed auto-encoders to learn features for each view and utilized an adaptive feature projection to avoid the imputation and fusion for missing data.

[217] utilized convolutional neural networks to extract deep features and employed attention mechanism to fuse these deep features in a weighted way. [218] used several view-specific graph convolutional encoder networks to reveal the high-level features and high-order geometric structure information of data. [219] integrated the element-wise reconstruction and the generative adversarial network (GAN) to infer the missing data, and employed multi-layer non-linear transformations to learn high-level common representation. [220] transferred known similar inter-instance relationships to the missing view and adopted graph networks constructed on the transferred relationship graph to infer missing data. It devised view-specific encoders and an attention-based fusion layer to extract the recovered multi-view data and obtain the common representation. [221] proposed a bi-level optimization framework to dynamically impute missing views from the learned semantic neighbors and automatically select imputed samples for training.

4.6 Contrastive learning-based approaches for IMVC

[222][223] employed contrastive learning with the maximal mutual information across different views to obtain the consistent representation. They used additional prediction networks with minimal conditional entropy of different views to recover the missing data. [224] designed augmentation-free graph contrastive learning and cross-view graph consistency learning to maximize the mutual information of different views within a cluster based on the graphs transferred from the inter-sample relation graphs.

[155] proposed a unified contrastive learning paradigm to simultaneously solve partially view-unaligned problem and partially sample-missing problem, where the available pairs were used as positives and some cross-view samples

were randomly selected as negatives, and the influence of the false negatives caused by random sampling was alleviated via the noise-robust contrastive loss. [225] developed multi-view unified and specific encoding network to fuse different views into a unified representation, designed a diversified graph contrastive regularization to enhance the discriminating power of the learned representation and reduce the information loss caused by the view missing, and utilized the robust contrastive learning loss to reduce the effect of noise and unreliable views.

4.7 Discussion

In IMVC, most approaches usually tend to impute/recover/infer values for the missing data of incomplete multi-view data sets in original data space, and then explore cluster information. The imputation quality depends on the estimation of the distribution of available data. Because incomplete multi-view data usually have biased data distribution, the imputation of missing data might induce noise even incorrect information, especially for data with a large missing rate. [168][158] et al. complete missing information to capture the features of missing instances. This way retains the original latent spatial structure of the data, but the common features of multiple views learned from the complete data may have large variances in both inter-class and intra-class. Because the clustering result is determined by a whole similarity matrix, the imputation has an impact on the clustering of all samples, no matter whether a sample is complete or not. When an imputation is not of high quality, it could adversely affect the clustering performance of all samples, especially for those with complete views. In addition, the filled values usually lack physical meaning and the imputation of missing data is time consuming.

The NMF-based approaches learn a consensus representation shared by all views for clustering. They explore certain information amongst the observed views and reduce the negative influence of the missing views via the partial view aligned or weighted regularization strategy. The graph based approaches transform the feature missing problem into the graph space and explore the similarity information amongst the observed instances in all views to learn the orthogonal consensus representation. These representations are more robust to noise and are suitable to the non-linear separable data.

It is difficult for IMVC approaches adopting shallow models to deal with the dependence, as well as discrepancy amongst different views and nonlinear relations amongst samples. This limitation restricts their capability to learn discriminative feature representations. Meanwhile, some of them suffer from high computation costs (e.g., in inverse operations of matrices). Compared with the approaches adopting shallow models, deep learning-based approaches have the potential to extract more salient features from the data and thus obtain very promising results.

Many IMVC approaches handle the incomplete multi-view data in a two-stage process, i.e., first exploring the consistency amongst multiple views from the complete part of data, and then extending the learned consistency to the incomplete part of data. However, feature learning performed on partial data might cause the distribution discrepancy between the features of complete data and that of incomplete data, resulting in the degradation of model generalization capability.

The unification of recovery of the missing views, the representation learning and the clustering task provides a novel insight to IMVC [226]. However, these approaches often face challenges associated with high computational and

storage complexities, limiting their applicability to large-scale datasets. Several novel approaches have been proposed, for example, [227] integrated missing view imputation into the fuzzy clustering process to realize cooperative learning between the visible and hidden views. They established hidden links between missing views, hidden views and complete views to improve the quality of the imputed missing views as well as the learned hidden views. Additionally, an adaptive view weighting mechanism was introduced to improve the robustness of the model. [228] robustly learned the common compact binary codes for incomplete multi-view features and optimized the cluster structures in an online fashion. [229] introduced genetics to clustering algorithms to learn simultaneously the consistent information and the unique information based on the subspace decomposition.

5 UNCERTAIN MULTI-VIEW CLUSTERING

Uncertain data clustering must first model uncertain data by using either of fuzzy model, evidence-oriented model, or probabilistic model. Moreover, similarity measure plays an imperative role [230]. When two distributions of two uncertain data are heavily overlapped in locations, the geometric distance-based similarity function cannot correctly capture the change between uncertain data with their distributions. On the other hand, similarity measure based on probability distribution, such as the divergence-based similarity function, cannot discriminate the change between data points when they are not closed to each other or completely separated.

[230] combined a self-adaptive mixture similarity function composed of geometric distance and S-divergence with *k*-medoids MVC to reduce the adverse effect of outliers and noises. [26] integrated induced kernel distance and Jeffrey-divergence in terms of the degree of overlap concerning each view in a dataset to construct a self-adaptive mixture similarity measure (SAM). Subsequently, they developed a multi-view spectral clustering algorithm with SAM as well as pairwise co-regularization to group uncertain data. [231] developed an approximate Bayes approach to directly estimate the cluster assignment and co-assignment probabilities in the range with several different clustering patterns. [232] proposed a Bayesian probabilistic model via variation inference to automatically learn the multiple expert views, multiple clustering structures and expert confidences. This approach enables the discovery of various ways to cluster data, considering potentially diverse inputs from multiple uncertain experts. Table 3 (Appendix) summarizes the characteristics of representative uncertain MVC approaches.

6 DYNAMIC MULTI-VIEW CLUSTERING

Most existing MVC approaches fuse all views at one time, but in some applications, the number of views changes with time. For the newly collected views, re-fusing all views at each time is too expensive to store all historical views [233].

To deal with dynamic change of views, [233, 234] developed incremental multi-view spectral clustering (IMSC) to integrate different views one by one in an incremental way, where [234] first learned an initial model from a small number of views, next updated the model when a new view is available, and then used the updated model to learn a consensus result. On the other hand, [233] opted to store a single consensus similarity matrix, representing the structural information of all historical views. Once the newly collected view is available, the consensus similarity

matrix is reconstructed by learning from its previous version and the current new view. [233] also incorporated sparse and connected graph learning to reduce the noises and preserve the correct connections within clusters.

In addition, [235] designed an alternative iterative algorithm based on the introduction of incremental learning mechanism to incrementally update the feature selection matrix. This feature selection was then incorporated into an extended weighted NMF to learn a consensus clustering indicator matrix.

There is a major challenge for incremental clustering called the stability-plasticity dilemma [236]. On the one hand, the clustering results should be plastic to the new input data from non-stationary distributions. On the other hand, the clustering results should retain the performance of previous input data. Table 4 (Appendix) summarizes the characteristics of representative dynamic MVC approaches.

7 DATASET

Multi-view datasets widely used by most MVC approached for evaluating the clustering performance consist of five categories: text, image, text-gene, image-text, and video.

7.1 Text Dataset

The text datasets consist of news dataset (3Sourses, BBC, BBCSport, Newsgroup), multilingual documents dataset (Reuters, Reuters-21578), citations dataset (Citeseer), WebKB webpage dataset (Cornell, Texas, Washington and Wisconsin), articles (Wikipedia), and diseases dataset (Derm). The statistics of text databases are reported in Table 5 (Appendix).

7.2 Image Dataset

The image datasets consist of facial image datasets (Yale, Yale-B, Extended-Yale, VIS/NIR, ORL, Notting-Hill, YouTube Faces), handwritten digits datasets (UCI, Digits, HW2source, Handwritten, MNIST-USPS, MNIST-10000, Noisy MNIST-Rotated MNIST), object image dataset (NUS/WIDE, MSRC, MSRCv1, COIL-20, Caltech101), Microsoft Research Asia Internet Multimedia Dataset 2.0 (MSRA-MM2.0), natural scene dataset (Scene, Scene-15, Out-Scene, Indoor), plant species dataset (100leaves), animal with attributes (AWA), multi-temporal remote sensing dataset (Forest), Fashion (such as T-shirt, Dress and Coat) dataset (Fashion-10K), sports event dataset (Event), image dataset (ALOI, ImageNet, Corel, Cifar-10, SUN1k, Sun397). The statistics of image databases are reported in Table 6 (Appendix).

7.3 Text-gene Dataset

The prokaryotic species dataset (Prok) is a text-gene dataset, which consists of 551 prokaryotic samples belonging to 4 classes. The species are represented by 1 textual view and 2 genomic views. The textual descriptions are summarized into a document-term matrix that records the TF-IDF re-weighted word frequencies. The genomic views are the proteome composition and the gene repertoire. The statistics of text-gene databases are reported in Table 7 (Appendix).

7.4 Image-text Dataset

The image-text datasets consist of Wikipedia's featured articles dataset (Wikipedia), drosophila embryos dataset (BDGP), NBA-NASCAR Sport dataset (NNSpt), indoor scenes (SentencesNYU v2 (RGB-D)), Pascal dataset (VOC), object dataset (NUS-WIDE-C5), and photographic images (MIR Flickr 1M). The statistics of image-text databases are reported in Table 8 (Appendix).

7.5 Video Dataset

The video datasets consist of actions of passenger dataset (DTHC), pedestrian video shot dataset (Lab), motion of body sequences (CMU Mobo) dataset, face video sequences dataset (YouTubeFace_sel, Honda/UCSD), and Columbia Consumer Video dataset (CCV). The statistics of video databases are reported in Table 9 (Appendix).

8 Performance metrics

The procedure for evaluating clustering results is known as cluster validity. Well-known performance metrics for evaluating performance of MVC include internal validation indexes and external validation indexes [230].

8.1 Internal indices

Internal indices are quality scores computed by employing only the information inherent to the dataset, such as Compactness (the average distance between every pair of data points), Davies–Bouldin index (the ratio of the sum of within-cluster scatters to between-cluster separations), Dunn validity index (inter cluster distances over intra cluster distances), or Separation (the mean Euclidean distance between cluster centroids) [237]. Ideally, the data samples of each cluster to be as close as possible. Thus, a smaller value of the Compactness indicates more compact and better clusters. The Separation quantifies the magnitude of separation between the clusters. A lower value of the Separation represents the closeness of the clusters. Further, the Davies–Bouldin index identifies the overlapping of clusters by calculating the fraction of the sum of with-in-cluster distribution to between-cluster separations. A value near to 0 of the Davies–Bouldin index denotes that the resultant clusters are well-separated and compact. At last, the Dunn validity index measures the ratio of inter-cluster distance to intra cluster distance. A higher value of the Dunn validity index illustrates the well-separated and compact clusters [230].

8.2 External indexes

The external indexes are quality scores computed by comparing clustering labels with the external supplied true labels, such as clustering accuracy (ACC), normalized mutual information (NMI), Purity, Rand Index (RI)/Adjusted Rand Index (ARI), Precision, Recall, and F-score (F1) [51, 238]. For all external evaluation metrics, a higher value close to 1 of each external index indicates preferable clustering results whereas a lower value close to 0 denotes undesirable clustering results.

The formulas of evaluation indicators are shown in Table 10 (Appendix).

9 EMPIRICAL EVALUATION

In this section, we empirically evaluate the performance of thirty-five approaches, including twenty-nine Complete MVC approaches and six Incomplete MVC approaches. The twenty-nine Complete MVC approaches consist of two single view clustering approaches (SCBest and SCCat. SCBest involves conducting spectral clustering on each single view independently and the result of the view with the best clustering performance is reported. SCCat involves concatenating vectors from different views into a new vector and then apply spectral clustering algorithm straightforwardly on the concatenated vector. Additionally, there are six NMF-based approaches (MultiNMF [16], MultiGNMF [239], MVCC [44], MVCF[240], MvCDMF [63], and MCDCF [58]), five graph-based approaches (SwMC [30], AMGL[32], MCGC [241], RMSC [242], and AWP [243]), nine subspace-based approaches (DiMSC [244], ECMSC [245], FMR [107], SM2VC [95], CSMSC [246], DSS-MSC [97], DMSCN [247], MvDSCN [109], and MvSC-MRAR [114]), three deep learning-based approaches (SDMVC [248], CoMVC [127], and MVC-MAE [117]), one contrastive learning-based approach (NMvC-GCN [129]), and three co-learning-based approaches (CoregSC [15], TW-Co-k-means [249], DWMVC [37]). The six Incomplete MVC approaches consist of one graph-based approach (PIMVC [193]), one subspace-based approach (LATER [203]), three deep learning-based approaches (DIMVC [159], APADC [216], DSIMVC[221]), and one contrastive learning-based approach (COMPLETER [222]). These approaches are representative MVC approaches, covering the main categories in the taxonomy proposed in this paper and having different structures or constraints.

The experiments are conducted on seven datasets, including three text datasets (BBCSport, Reuters, Reuters-21578), and three image datasets (NUSWIDE, Caltech101-7, NUSWIDE30K), and one image-text dataset (RGB-D). NUSWIDE30K contains 30,000 instances, while the other datasets contain hundreds or thousands of samples. These datasets are widely recognized as well-known benchmarks in MVC communities, representing different typical application scenarios, each with distinct views, classes, instances, features, and feature dimensions. We choose them to explore how various approaches behave in different types of datasets. ACC and NMI are used as performance metrics. In the experiments, all algorithms are run 5 times on each dataset and the average performance are reported. All evaluations are carried out on a standard Ubuntu-18.04 OS, and eleven approaches (DMSCN, MvDSCN, MvSC-MRAR, SDMVC, CoMVC, MVC-MAE, NMvC-GCN, COMPLETER, DSIMVC, DIMVC, APADC) are implemented by using Pytorch 1.0 with an NVIDIA 2080Ti GPU, while other approaches are implemented by using Matlab with an Intel Core i7-7820X CPU. For each compared approach, a grid search is performed in the parameter set suggested in the corresponding paper and the best results are reported.

The ACCs and NMIs of twenty-nine Complete MVC approaches and six Incomplete MVC approaches are shown in Table 11 (Complete MVC) and Table 12 (Incomplete MVC) (Appendix), where bold numbers represent the best results in the column. From Table 11 and Table 12 (Appendix), we have the following observations and analyses:

(1) The clustering performance of SCBest and SCCat are not necessarily worse than those of MVC algorithms, especially, SCCat achieves the second-highest ACC and NMI values on BBCSport dataset. Such phenomena are contrary to the general expectation that multi-view algorithms can obtain better performance than that of merely using

a single view. This indicates that it is important to design a strategy for integrating multi-view information, otherwise a simple concatenation may perform better.

(2) Amongst all MVC approaches, MvSC-MRAR achieves the highest ACC and NMI values on Reuters, Reuters-21578, NUSWIDE, Caltech101-7, and RGB-D, meanwhile, it also achieves the third-best clustering performance, after that of NMvC-GCN and SCCat, on the dataset BBCSport. Amongst the nine subspace-based MVC approaches, approaches with deep structure generally outperform those with shallow structure. However, MvDSCN and DMSCN do not perform as well as MvSC-MRAR, because they only use the information of the last layer in encoder components, without integration of information from different levels. It indicates that deep learning and the integration of the nonlinear structure, multilevel representation information in multi-view data, and the data distribution of latent representation are helpful for improving the performance of MVC. In addition, NMvC-GCN achieves the highest ACC and NMI values on BBCSport, and its ACC and NMI values are also close to those of MvSC-MRAR on other datasets. It indicates that contrastive learning is effective in MVC.

(3) No one MVC approach can maintain consistent good performance on various datasets. For example, MultiGNMF with local graph regularization achieves the highest ACC and NMI values on Caltech101-7 and RGB-D, while MultiNMF with the consensus constraint and MVCC with the multi-manifold regularization achieves the highest ACC and NMI values on BBCSport and Reuters respectively. These phenomena show that factors such as model structures, regularization constraints and weights corresponding to different views all contribute to improving the clustering performance. Moreover, the information captured for clustering under different conditions are different.

(4) Except for the single-view approaches (SCBest, SCCat), the deep learning-based approaches (SDMVC, CoMVC, MVC-MAE), AWP, and TW-Co-k-means, other approaches did not complete the clustering task (i.e. have not generated clustering results) on the dataset of NUSWIDE30K, due to the run out of memory or heavy time-consuming. However, the ACC and NMI values obtained by SCBest, SCCat, SDMVC, CoMVC, MVC-MAE, AWP, and TW-Co-k-means on NUSWIDE30K are relatively low. This suggests that these approaches do not fully capture the inherent geometrical structure of NUSWIDE30K. The challenge of clustering large-scale multi-view data persists.

(5) In Table 12 (Incomplete MVC), LATER achieves the highest ACC and NMI values on the BBCSport, Reuters-21578, and RGB-D under various missing-rates (ratio of missing samples to all samples), but an exception occurs on the Reuters, DSIMVC achieves the highest ACC and NMI values on NUSWIDE and NUSWIDE30K under various missing-rates. On Caltech101-7, COMPLETER outperforms other approaches when the missing-rate is 0.1, while APADC does well when missing-rate is greater than 0.5, the performance of DIMVC and PIMVC is not as good as other approaches. Furthermore, we also see that ACC and NMI values of many algorithms decrease with the increase of the missing-rate.

Figures 8 (a) and (b) show the runtimes of different approaches on the BBCSport dataset. From the Figure 8 (a) (Complete MVC), it can be seen that SCBest, SCCat, CSMSC, DSS-MSC, TW-Co-k-means, and CoregSC run well, SDMVC, CoMVC, MVC-MAE, NMvC-GCN, SwMC, AMGL, MCGC, RMSC, DiMSC, ECMSC, and SM2VC also run reasonably well. However, MultiNMF, MultiGNMF, MVCC, MVCF, MvCDMF, MCDCF, AWP, DMSCN, MvDSCN, and MvSC-MRAR run slower, especially FMR and DWMVC run unreasonably. The difference of running

time between different approaches stems from many factors, such as the design of the algorithms, the complexity of adopted techniques, the clustering routine, the model structure, the number of samples, the dimension of features, and among others. For example, MvCDMF requires multiple layers of deep NMF decomposition, MvDSCN requires multiplication with a matrix of the square of the similarity matrix during each iteration, which leads to longer execution time. In the Figure 8 (b) (Incomplete MVC), DIMVC is the slowest, and PIMVC is the fastest. Furthermore, the same approach takes longer on datasets with higher missing rates.

10 FUTURE WORK

Although existing MVC approaches have obtained promising performance via different strategies, such as the weighting views or samples, imposing manifold or low-rank constraints, there are still some unresolved problems in the field of MVC, which are worthy of the attention of researchers. In the following, we summarize several directions of future research in order to encourage more research in MVC.

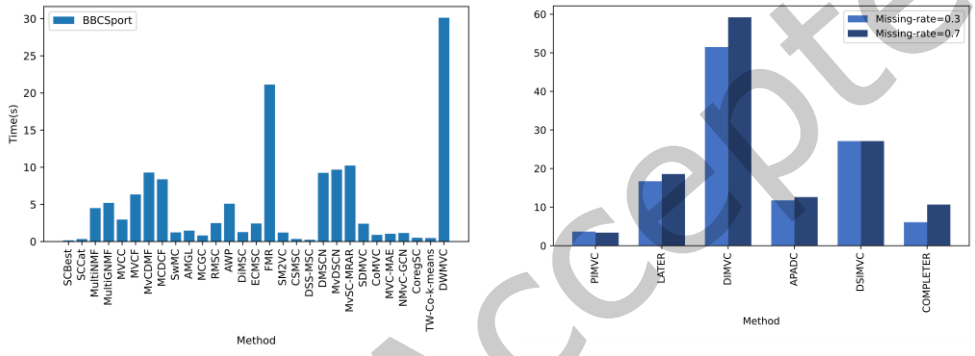

Figure 8: Comparison of runtime of different algorithms on the BBCSport dataset (complete/incomplete)

MVC for Large-scale and high-dimensional data. The datasets used for current experimental evaluation are often million-scale with low dimensionality, and only a few of them are billion-scale or have thousands of feature dimensions. However, in many real applications, the datasets may involve billions of samples or have high dimensionality, e.g., several million posts are shared per minute in Facebook, and each person has millions of genetic variants as genetic features in bioinformatics [11]. As a result, existing MVC solutions may fail to process such real large-scale and/or high-dimensional data within reasonable time cost. Furthermore, the performance of algorithms is also affected in a high-dimensional situation, because high-dimensional data often has a large amount of redundant information. The redundant information not only fails to supplement valid information, but also jeopardizes good data feature representation [6]. It is important to develop approaches to efficiently perform clustering on these large-scale or high-dimensional multi-view data. One possible research direction is developing clustering algorithms based on distributed computation platforms, or keeping the data on disk and designing I/O-efficient clustering algorithms.

MVC for dynamic and uncertain data. Existing MVC mainly focuses on clustering the static data and the settings of dynamic data clustering are overlooked. In real life scenarios, data are not always static, e.g., video data. On

the one hand, some samples appear while other samples disappear. On the other hand, samples may be described by some time-varying information. The techniques for dynamic clustering need to be scalable and better to be incremental so as to deal with the dynamic changes efficiently. How to design effective MVC approaches in dynamic domains remains an open question. Moreover, how to integrate various features for uncertain data clustering is also a subject deserved further to study.

MVC for data with the unknown number of clusters. Most existing MVC works assume that the number of clusters is known. This assumption, however, is too strong in real applications, especially in multi-view scenarios. How to design MVC approaches that can automatically determine the number of clusters is a problem worthy of careful investigation.

MVC with more interpretation. Most deep learning approaches can explore the relationships amongst different samples and avoid the possible corruption as well as the curse of dimensionality, but many deep MVC approaches involve more parameters and lack theoretic interpretability. Thus, we may need to construct deep networks based on optimization approaches and make the networks more interpretable.

11 CONCLUSION

As a powerful learning tool for exploring the structure of multi-view data, MVC is widely used in many disciplines and plays a crucial role in various applications. However, the complex distribution and diversified heterogeneous features of multi-view data impose challenges for MVC. Despite several survey studies to assess MVC approaches have been performed in the past, these survey studies mainly focus on limited approaches for complete multi-view data or those for incomplete multi-view data. They often overlook approaches for complete, incomplete, uncertain, and dynamic multi-view data at the same time. To fill this gap and provide an updated survey for MVC approaches including those proposed after 2019, this study first analyzes and introduces the basic concepts and common key techniques for MVC, proposes a novel taxonomy of MVC approaches. The study then presents the working mechanisms as well as characteristics of the representative MVC approaches proposed in recent years. Following this, it summarizes representative MVC datasets and performance metrics commonly used in the MVC field. Finally, the study selects thirty-five representative MVC approaches from the proposed taxonomy to conduct an empirical evaluation on seven real world benchmark datasets.

The study provides the theoretical knowledge of MVC, a novel taxonomy for MVC approaches, and insightful information on performance features of existing approaches. Its primary goal is to assist researchers interested in MVC and its related areas by offering an in-depth understanding on the current progresses of MVC. Additionally, it aims to serve as an insightful guideline to practitioners for application development.

ACKNOWLEDGMENTS

This research is supported by the National Natural Science Foundation of China (62062066, 61762090, 61966036 and 62276227), Yunnan Fundamental Research Projects (202201AS070015); Yunnan Key Laboratory of Intelligent Systems and Computing (202205AG070003), the Blockchain and Data Security Governance Engineering Research

Center of Yunnan Provincial Department of Education, University Key Laboratory of Internet of Things Technology and Application of Yunnan Province.

REFERENCES

[1] Zhao Kang, Xinjia Zhao, Chong Peng, Hongyuan Zhu, Joey Tianyi Zhou, Xi Peng, Wenyu Chen and Zenglin Xu. 2020. Partition level multiview subspace clustering. Neural Networks 122, 279-288.

[2] Junwei Han, Jinglin Xu, Feiping Nie and Xuelong Li. 2022. Multi-view k-means clustering with adaptive sparse memberships and weight allocation. IEEE Transactions on Knowledge and Data Engineering, 34, 2, 816-827 (2022).

[3] Yan Yang and Hao Wang. 2018. Multi-view clustering: A survey. Big Data Mining and Analytics 1, 2, 83-107.

[4] Chang Xu, Dacheng Tao and Chao Xu. 2013. A survey on multi-view learning. arXiv preprint arXiv:1304.5634.

[5] Zhaoliang Chen, Lele Fu, Jie Yao, Wenzhong Guo, Claudia Plant and Shiping Wang. 2023. Learnable graph convolutional network and feature fusion for multi-view learning. Information Fusion 95, 109-119.

[6] Lele Fu, Pengfei Lin, Athanasios V Vasilakos and Shiping Wang. 2020. An overview of recent multi-view clustering. Neurocomputing 402, 148-161.

[7] Shiliang Sun. 2013. A survey of multi-view machine learning. Neural Computing and Applications 23, 7, 2031-2038.

[8] Jing Zhao, Xijiong Xie, Xin Xu and Shiliang Sun. 2017. Multi-view learning overview: Recent progress and new challenges. Information Fusion 38, 43-54.

[9] Yingming Li, Ming Yang and Zhongfei Zhang. 2018. A survey of multi-view representation learning. IEEE Transactions on Knowledge and Data Engineering 31, 10, 1863-1883.

[10] Shu Li, Wen-Tao Li and Wei Wang. 2020. Co-gcn for multi-view semi-supervised learning. In Proceedings of The AAAI Conference on Artificial Intelligence, 4691-4698.

[11] Guoqing Chao, Shiliang Sun and Jinbo Bi. 2017. A survey on multi-view clustering. arXiv preprint arXiv:1712.06246.

[12] Sanjay Kumar Anand and Suresh Kumar. 2022. Experimental Comparisons of Clustering Approaches for Data Representation. ACM Computing Surveys (CSUR) 55, 3, 1-33.

[13] Han Xiao, Yidong Chen and Xiaodong Shi. 2019. Knowledge graph embedding based on multi-view clustering framework. IEEE Transactions on Knowledge and Data Engineering 33, 2, 585-596.

[14] Syed Fawad Hussain, Khadija Khan and Rashad Jillani. 2022. Weighted multi-view co-clustering (WMVCC) for sparse data. Applied Intelligence 52, 1, 398-416.

[15] Abhishek Kumar, Piyush Rai and Hal Daume. 2011. Co-regularized multi-view spectral clustering. Advances in Neural Information Processing Systems 24.

[16] 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, 252-260.

[17] Yeqing Li, Feiping Nie, Heng Huang and Junzhou Huang. 2015. Large-scale multi-view spectral clustering via bipartite graph. In Proceedings of

29th AAAI Conference on Artificial Intelligence, 2750-2756.

[18] Lusi Li and Haibo He. 2022. Bipartite Graph based Multi-view Clustering. IEEE Transactions on Knowledge and Data Engineering 34, 7, 3111-3125.

[19] Jie Wen, Zheng Zhang, Lunke Fei, Bob Zhang, Yong Xu, Zhao Zhang and Jinxing Li. 2022. A survey on incomplete multiview clustering. IEEE Transactions on Systems, Man, and Cybernetics: Systems 53, 2, 1136-1149.

[20] Yanying Mei, Zhenwen Ren, Bin Wu, Yanhua Shao and Tao Yang. 2022. Robust graph-based multi-view clustering in latent embedding space. International Journal of Machine Learning and Cybernetics 13, 2, 497-508.

[21] Ben Yang, Xuetao Zhang, Badong Chen, Feiping Nie, Zhiping Lin and Zhixiong Nan. 2022. Efficient correntropy-based multi-view clustering with anchor graph embedding. Neural Networks 146, 290-302.

[22] Shiliang Sun, Feng Jin and Wenting Tu. 2011. View construction for multi-view semi-supervised learning. In Advances in Neural Networks–ISNN 2011: 8th International Symposium on Neural Networks, 595-601.

[23] Cam-Tu Nguyen, De-Chuan Zhan and Zhi-Hua Zhou. 2013. Multi-modal image annotation with multi-instance multi-label LDA. In Proceedings of The 23rd International Joint Conference on Artificial Intelligence, 1558-1564.

[24] Jianguo Zhao, Gengyu Lyu and Songhe Feng. 2022. Linear neighborhood reconstruction constrained latent subspace discovery for incomplete multi-view clustering. Applied Intelligence 52, 1, 982-993.

[25] Linlin Zong, Faqiang Miao, Xianchao Zhang, Xinyue Liu and Hong Yu. 2021. Incomplete multi-view clustering with partially mapped instances and clusters. Knowledge-Based Systems 212, 106615.

[26] Krishna Kumar Sharma and Ayan Seal. 2021. Multi-view spectral clustering for uncertain objects. Information Sciences 547, 723-745.

[27] Minmin Miao, Wenbin Zhang, Wenjun Hu and Ruiqin Wang. 2020. An adaptive multi-domain feature joint optimization framework based on composite kernels and ant colony optimization for motor imagery EEG classification. Biomedical Signal Processing and Control 61, 101994.

[28] Wei Wang and Zhi-Hua Zhou. 2007. Analyzing co-training style algorithms. In European Conference on Machine Learning, 454-465.

[29] Siwei Wang, Xinwang Liu, En Zhu, Chang Tang, Jiyuan Liu, Jingtao Hu, Jingyuan Xia and Jianping Yin. 2019. Multi-view Clustering via Late Fusion Alignment Maximization. In Proceedings of The 28th International Joint Conference on Artificial Intelligence, 3778-3784.

[30] Feiping Nie, Jing Li and Xuelong Li. 2017. Self-weighted Multiview Clustering with Multiple Graphs. In Proceedings of The 26th International Joint Conference on Artificial Intelligence, 2564-2570.

[31] Tian Xia, Dacheng Tao, Tao Mei and Yongdong Zhang. 2010. Multiview spectral embedding. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 40, 6, 1438-1446.

[32] Feiping Nie, Jing Li and Xuelong Li. 2016. Parameter-free auto-weighted multiple graph learning: a framework for multiview clustering and semi-supervised classification. In Proceedings of The 27th International Joint Conference on Artificial Intelligence, 1881-1887.

[33] Du Guowang, Zhou Lihua, Wang Lizhen and Du Jingwei. 2022. Multi-View Clustering Based on Two-Level Weights. Journal of Computer Research and Developments 59, 4, 15.

[34] Xiaoqian Zhang, Jing Wang, Xuqian Xue, Huaijiang Sun and Jiangmei Zhang. 2022. Confidence level auto-weighting robust multi-view subspace clustering. Neurocomputing 475, 38-52.

[35] Yu-Meng Xu, Chang-Dong Wang and Jian-Huang Lai. 2016. Weighted multi-view clustering with feature selection. Pattern Recognition 53, 25-35.

[36] Jing Liu, Fuyuan Cao, Xiao-Zhi Gao, Liqin Yu and Jiye Liang. 2020. A cluster-weighted kernel K-means method for multi-view clustering. In Proceedings of The AAAI Conference on Artificial Intelligence, 4860-4867.

[37] Hu Shi-Zhe, Lou Zheng-Zheng, Wang Ruo-Bin, Yan Xiao-Qiang and Ye Yang-Dong. 2020. Dual-Weighted Multi-View Clustering Journal of Computer Science 43, 9, 1708-1720.

[38] Zhao Kang, Guoxin Shi, Shudong Huang, Wenyu Chen, Xiaorong Pu, Joey Tianyi Zhou and Zenglin Xu. 2020. Multi-graph fusion for multi-view spectral clustering. Knowledge-Based Systems 189, 105102.

[39] George Trigeorgis, Konstantinos Bousmalis, Stefanos Zafeiriou and Björn W Schuller. 2016. A deep matrix factorization method for learning attribute representations. IEEE Transactions on Pattern Analysis & Machine Intelligence 39, 3, 417-429.

[40] Cai Xu, Ziyu Guan, Wei Zhao, Yunfei Niu, Quan Wang and Zhiheng Wang. 2018. Deep Multi-View Concept Learning. In Proceedings of The 27th International Joint Conference on Artificial Intelligence, 2898-2904.

[41] Chih-Jen Lin. 2007. Projected gradient methods for nonnegative matrix factorization. Neural Computation 19, 10, 2756-2779.

[42] Zhouchen Lin, Risheng Liu and Zhixun Su. 2011. Linearized alternating direction method with adaptive penalty for low-rank representation. Advances in Neural Information Processing Systems 2011, 612-620.

[43] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato and Jonathan Eckstein. 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends® in Machine learning 3, 1, 1-122.

[44] Hao Wang, Yan Yang and Tianrui Li. 2016. Multi-view clustering via concept factorization with local manifold regularization. In 2016 IEEE 16th International Conference on Data Mining (ICDM), 1245-1250.

[45] Linlin Zong, Xianchao Zhang, Long Zhao, Hong Yu and Qianli Zhao. 2017. Multi-view clustering via multi-manifold regularized non-negative matrix factorization. Neural Networks 88, 74-89.

[46] Naiyao Liang, Zuyuan Yang, Zhenni Li, Shengli Xie and Chun-Yi Su. 2020. Semi-supervised multi-view clustering with graph-regularized partially shared non-negative matrix factorization. Knowledge-Based Systems 190, 105185.

[47] Hao Cai, Bo Liu, Yanshan Xiao and Luyue Lin. 2019. Semi-supervised multi-view clustering based on constrained nonnegative matrix factorization. Knowledge-Based Systems 182, 104798.

[48] Hao Cai, Bo Liu, Yanshan Xiao and Luyue Lin. 2020. Semi-supervised multi-view clustering based on orthonormality-constrained nonnegative matrix factorization. Information Sciences 536, 171-184.

[49] 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, 105582.

[50] Xiao Zheng, Chang Tang, Xinwang Liu and En Zhu. 2023. Multi-view clustering via matrix factorization assisted k-means. Neurocomputing 534, 45-54.

[51] Peng Luo, Jinye Peng, Ziyu Guan and Jianping Fan. 2018. Dual regularized multi-view non-negative matrix factorization for clustering. Neurocomputing 294, 1-11.

[52] Mingyang Liu, Zuyuan Yang, Lingjiang Li, Zhenni Li and Shengli Xie. 2023. Auto-weighted collective matrix factorization with graph dual regularization for multi-view clustering. Knowledge-Based Systems 260, 110145.

[53] Lihua Zhou, Guowang Du, Kevin Lü and Lizhen Wang. 2021. A network-based sparse and multi-manifold regularized multiple non-negative matrix factorization for multi-view clustering. Expert Systems with Applications 174, 114783.

[54] Ben Yang, Xuetao Zhang, Feiping Nie, Fei Wang, Weizhong Yu and Rong Wang. 2020. Fast Multi-View Clustering via Nonnegative and Orthogonal Factorization. IEEE Transactions on Image Processing 30, 2575-2586.

[55] Khamis Houfar, Djamel Samai, Fadi Dornaika, Azeddine Benlamoudi, Khaled Bensid and Abdelmalik Taleb-Ahmed. 2023. Automatically weighted binary multi-view clustering via deep initialization (AW-BMVC). Pattern Recognition 137, 109281.

[56] Zheng Zhang, Li Liu, Fumin Shen, Heng Tao Shen and Ling Shao. 2019. Binary Multi-View Clustering. IEEE Transactions on Pattern Analysis & Machine Intelligence 41, 7, 1774-1782.

[57] Jianxi Zhao, Fangyuan Kang, Qingrong Zou and Xiaonan Wang. 2023. Multi-view clustering with orthogonal mapping and binary graph. Expert Systems with Applications 213, 118911.

[58] Shuai Chang, Jie Hu, Tianrui Li, Hao Wang and Bo Peng. 2021. Multi-view clustering via deep concept factorization. Knowledge-Based Systems 217, 106807.

[59] Chen Zhang, Siwei Wang, Jiyuan Liu, Sihang Zhou, Pei Zhang, Xinwang Liu, En Zhu and Changwang Zhang. 2021. Multi-view clustering via deep matrix factorization and partition alignment. In Proceedings of The 29th ACM International Conference on Multimedia, 4156-4164.

[60] 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, 108-116.

[61] Guowang Du, Lihua Zhou, Kevin Lü and Haiyan Ding. 2021. Deep multiple non-negative matrix factorization for multi-view clustering. Intelligent Data Analysis 25, 2, 339-357.

[62] Zexi Chen, Pengfei Lin, Zhaoliang Chen, Dongyi Ye and Shiping Wang. 2022. Diversity embedding deep matrix factorization for multi-view clustering. Information Sciences 610, 114-125.

[63] Handong Zhao, Zhengming Ding and Yun Fu. Multi-View Clustering via Deep Matrix Factorization. In 2017 31th AAAI Conference on Artificial Intelligence (AAAI), 2921-2927.

[64] Zhenwen Ren, Haoyun Lei, Quansen Sun and Chao Yang. 2020. Simultaneous Learning Coefficient Matrix and Affinity Graph for Multiple Kernel Clustering. Information Sciences, 547, 289-306.

[65] Zhenwen Ren, Haoran Li, Chao Yang and Quansen Sun. 2020. Multiple kernel subspace clustering with Local Structural Graph and Low-Rank Consensus Kernel Learning. Knowledge-Based Systems 188, 105040.

[66] Shudong Huang, Zhao Kang, Ivor W Tsang and Zenglin Xu. 2019. Auto-weighted multi-view clustering via kernelized graph learning. Pattern Recognition 88, 174-184.

[67] Yongyong Chen, Xiaolin Xiao and Yicong Zhou. 2020. Jointly learning kernel representation tensor and affinity matrix for multi-view clustering. IEEE Transactions on Multimedia 22, 8, 1985-1997.

[68] Guang-Yu Zhang, Xiao-Wei Chen, Yu-Ren Zhou, Chang-Dong Wang, Dong Huang and Xiao-Yu He. 2022. Kernelized multi-view subspace

clustering via auto-weighted graph learning. Applied Intelligence 52, 1, 716-731.

[69] 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 5092-5101.

[70] Quanxue Gao, Zhizhen Wan, Ying Liang, Qianqian Wang, Yang Liu and Ling Shao. 2020. Multi-view projected clustering with graph learning. Neural Networks 126, 335-346.

[71] Zhenwen Ren, Xingfeng Li, Mithun Mukherjee, Yuqing Huang, Quansen Sun and Zhen Huang. 2021. Robust multi-view graph clustering in latent energy-preserving embedding space. Information Sciences 569, 582-595.

[72] Jian Dai, Zhenwen Ren, Yunzhi Luo, Hong Song and Jian Yang. 2021. Multi-view Clustering with Latent Low-rank Proxy Graph Learning. Cognitive Computation 13, 4, 1049-1060.

[73] Peiguang Jing, Yuting Su, Zhengnan Li and Liqiang Nie. 2021. Learning robust affinity graph representation for multi-view clustering. Information Sciences 544, 155-167.

[74] Hao Wang, Yan Yang and Bing Liu. 2019. GMC: Graph-based multi-view clustering. IEEE Transactions on Knowledge and Data Engineering 32, 6, 1116-1129.

[75] Hao Wang, Yan Yang, Bing Liu and Hamido Fujita. 2019. A study of graph-based system for multi-view clustering. Knowledge-Based Systems 163, 1009-1019.

[76] Huiling Xu, Xiangdong Zhang, Wei Xia, Quanxue Gao and Xinbo Gao. 2020. Low-rank tensor constrained co-regularized multi-view spectral clustering. Neural Networks 132, 245-252.

[77] 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, 257-264.

[78] Yongyong Chen, Xiaolin Xiao, Chong Peng, Guangming Lu, Yicong Zhou. 2022. Low-Rank Tensor Graph Learning for Multi-view Subspace Clustering. IEEE Transactions on Circuits and Systems for Video Technology 32, 1, 92-104.

[79] Sally El Hajjar, Fadi Dornaika and Fahed Abdallah. 2022. Multi-view spectral clustering via constrained nonnegative embedding. Information Fusion 78, 209-217.

[80] Shiping Wang, Shunxin Xiao, William Zhu and Yingya Guo. 2022. Multi-view fuzzy clustering of deep random walk and sparse low-rank embedding. *Information Sciences* 586, 224-238.

[81] Man-Sheng Chen, Ling Huang, Chang-Dong Wang and Dong Huang. 2019. Multi-view spectral clustering via multi-view weighted consensus and matrix-decomposition based discretization. *In Database Systems for Advanced Applications: 24th International Conference*, 175-190.

[82] 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*, 3513-3520.

[83] Mashaan Alshammari, John Stavrakakis and Masahiro Takatsuka. 2021. Refining a k-nearest neighbor graph for a computationally efficient spectral clustering. *Pattern Recognition* 114, 107869.

[84] Ziheng Li, Zhanxuan Hu, Feiping Nie, Rong Wang and Xuelong Li. 2022. Multi-view clustering based on generalized low rank approximation. *Neurocomputing* 471, 251-259.

[85] Zhanxuan Hu, Feiping Nie, Rong Wang and Xuelong Li. 2020. Multi-view spectral clustering via integrating nonnegative embedding and spectral embedding. *Information Fusion* 55, 251-259.

[86] Shaojun Shi, Feiping Nie, Rong Wang and Xuelong Li. 2020. Auto-weighted multi-view clustering via spectral embedding. *Neurocomputing* 399, 369-379.

[87] Guo Zhong, Ting Shu, Guoheng Huang and Xueming Yan. 2022. Multi-view spectral clustering by simultaneous consensus graph learning and discretization. *Knowledge-Based Systems* 235, 107632.

[88] Rong Wang, Feiping Nie, Zhen Wang, Haojie Hu and Xuelong Li. 2019. Parameter-free weighted multi-view projected clustering with structured graph learning. *IEEE Transactions on Knowledge and Data Engineering* 32, 10, 2014-2025.

[89] Zhao Kang, Zhiping Lin, Xiaofeng Zhu and Wenbo Xu. 2021. Structured graph learning for scalable subspace clustering: From single view to multiview. *IEEE Transactions on Cybernetics* 52, 9, 8976-8986.

[90] Hong Tao, Chenping Hou, Yuhua Qian, Jubo Zhu and Dongyun Yi. 2020. Latent complete row space recovery for multi-view subspace clustering. *IEEE Transactions on Image Processing* 29, 8083-8096.

[91] Xiaomeng Si, Qiyue Yin, Xiaojie Zhao and Li Yao. 2022. Consistent and diverse multi-View subspace clustering with structure constraint. *Pattern Recognition* 121, 108196.

[92] Yalan Qin, Hanzhou Wu, Xinpeng Zhang and Guorui Feng. 2021. Semi-Supervised Structured Subspace Learning for Multi-View Clustering. *IEEE Transactions on Image Processing* 31, 1-14.

[93] Xiaolan Liu, Gan Pan and Mengying Xie. 2021. Multi-view subspace clustering with adaptive locally consistent graph regularization. *Neural Computing and Applications* 33, 22, 15397-15412.

[94] Xiaobo Wang, Zhen Lei, Xiaojie Guo, Changqing Zhang, Hailin Shi and Stan Z Li. 2019. Multi-view subspace clustering with intactness-aware similarity. *Pattern Recognition* 88, 50-63.

[95] Zhiyong Yang, Qianqian Xu, Weigang Zhang, Xiaochun Cao and Qingming Huang. 2019. Split multiplicative multi-view subspace clustering. *IEEE Transactions on Image Processing* 28, 10, 5147-5160.

[96] Xiaosha Cai, Dong Huang, Guang-Yu Zhang and Chang-Dong Wang. 2023. Seeking commonness and inconsistencies: A jointly smoothed approach to multi-view subspace clustering. *Information Fusion* 91, 364-375.

[97] Tao Zhou, Changqing Zhang, Xi Peng, Harish Bhaskar and Jie Yang. 2019. Dual shared-specific multiview subspace clustering. *IEEE Transactions on Cybernetics* 50, 8, 3517-3530.

[98] Cao Rong-Wei, Zhu Ji-Hua, Hao Wen-Yu, Zhang Chang-Qing, Zhang Zhuo-Han and Li Zhong-Yu. 2022. Dual Weighted Multi-view Subspace Clustering. *Ruan Jian Xue Bao/Journal of Software* 33, 2, 585-597.

[99] Run-Kun Lu, Jian-Wei Liu and Xin Zuo. 2021. Attentive multi-view deep subspace clustering net. *Neurocomputing* 435, 186-196.

[100] Xiaoqian Zhang, Huaijiang Sun, Zhigui Liu, Zhenwen Ren, Qiongjie Cui and Yanmeng Li. 2019. Robust low-rank kernel multi-view subspace clustering based on the schatten p-norm and correntropy. *Information Sciences* 477, 430-447.

[101] Xiaoqian Zhang, Zhenwen Ren, Huaijiang Sun, Keqiang Bai, Xinghua Feng and Zhigui Liu. 2021. Multiple kernel low-rank representation-based robust multi-view subspace clustering. *Information Sciences* 551, 324-340.

[102] Guang-Yu Zhang, Yu-Ren Zhou, Xiao-Yu He, Chang-Dong Wang and Dong Huang. 2020. One-step kernel multi-view subspace clustering. *Knowledge-Based Systems* 189, 105126.

[103] Lele Fu, Zhaoliang Chen, Yongyong Chen and Shiping Wang. 2023. Unified low-rank tensor learning and spectral embedding for multi-view subspace clustering. *IEEE Transactions on Multimedia* 25,4972-4985.

[104] Haiyan Wang, Guoqiang Han, Junyu Li, Bin Zhang, Jiazhou Chen, Yu Hu, Chu Han and Hongmin Cai. 2021. Learning task-driving affinity matrix for accurate multi-view clustering through tensor subspace learning. *Information Sciences* 563, 290-308.

[105] Wentao Rong, Enhong Zhuo, Hong Peng, Jiazhou Chen, Haiyan Wang, Chu Han and Hongmin Cai. 2021. Learning a consensus affinity matrix for multi-view clustering via subspaces merging on Grassmann manifold. *Information Sciences* 547, 68-87.

[106] Yongyong Chen, Shuqin Wang, Xiaolin Xiao, Youfa Liu, Zhongyun Hua and Yicong Zhou. 2021. Self-paced enhanced low-rank tensor kernelized multi-view subspace clustering. *IEEE Transactions on Multimedia* 24, 4054-4066.

[107] Ruihuang Li, Changqing Zhang, Qinghua Hu, Pengfei Zhu and Zheng Wang. 2019. Flexible Multi-View Representation Learning for Subspace Clustering. *In Proceedings of The 28th International Joint Conference on Artificial Intelligence*, 2916-2922.

[108] Zhe Xue, Junping Du, Dawei Du and Siwei Lyu. 2019. Deep low-rank subspace ensemble for multi-view clustering. *Information Sciences* 482, 210-227.

[109] Pengfei Zhu, Binyuan Hui, Changqing Zhang, Dawei Du, Longyin Wen and Qinghua Hu. 2019. Multi-view Deep Subspace Clustering Networks. *arXiv preprint arXiv:1908.01978* abs/1908.01978.

[110] Jing-Hua Yang, Chuan Chen, Hong-Ning Dai, Meng Ding, Le-Le Fu and Zibin Zheng. 2022. Hierarchical Representation for Multi-view Clustering: From Intra-sample to Intra-view to Inter-view. *In Proceedings of The 31st ACM International Conference on Information & Knowledge Management*, 2362-2371.

[111] Qianqian Wang, Jiafeng Cheng, Quanxue Gao, Guoshuai Zhao and Licheng Jiao. 2020. Deep multi-view subspace clustering with unified and discriminative learning. *IEEE Transactions on Multimedia* 23, 3483-3493.

[112] Xiaomeng Si, Qiyue Yin, Xiaojie Zhao and Li Yao. 2022. Robust deep multi-view subspace clustering networks with a correntropy-induced metric. *Applied Intelligence* 52, 13, 14871-14887.

[113] Qinghai Zheng, Jihua Zhu, Yuanyuan Ma, Zhongyu Li and Zhiqiang Tian. 2021. Multi-view subspace clustering networks with local and global graph information. *Neurocomputing* 449, 15-23.

[114] Guowang Du, Lihua Zhou, Kevin Lü, Hao Wu and Zhimin Xu. 2023. Multiview subspace clustering with multilevel representations and adversarial regularization. *IEEE Transactions on Neural Networks and Learning Systems* 34, 12, 10279-10293.

[115] Siwei Wang, Xinwang Liu, Xinzhong Zhu, Pei Zhang, Yi Zhang, Feng Gao and En Zhu. 2021. Fast Parameter-Free Multi-View Subspace Clustering With Consensus Anchor Guidance. *IEEE Transactions on Image Processing* 31, 556-568.

[116] Zhaoyang Li, Qianqian Wang, Zhiqiang Tao, Quanxue Gao and Zhaohua Yang. 2019. Deep Adversarial Multi-view Clustering Network. *In Proceedings of The 28th International Joint Conference on Artificial Intelligence*, 4. 2952-2958.

[117] Guowang Du, Lihua Zhou, Yudi Yang, Kevin Lü and Lizhen Wang. 2021. Deep Multiple Auto-Encoder-Based Multi-view Clustering. *Data Science and Engineering* 6, 3, 323-338.

[118] 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, 279-290.

[119] Ru Wang, Lin Li, Xiaohui Tao, Xiao Dong, Peipei Wang and Peiyu Liu. 2021. Trio-based collaborative multi-view graph clustering with multiple constraints. *Information Processing & Management* 58, 3, 102466.

[120] Yuan Xie, Bingqian Lin, Yanyun Qu, Cuihua Li, Wensheng Zhang, Lizhuang Ma, Yonggang Wen and Dacheng Tao. 2020. Joint deep multi-view learning for image clustering. *IEEE Transactions on Knowledge and Data Engineering* 33, 11, 3594-3606.

[121] Wei Xia, Sen Wang, Ming Yang, Quanxue Gao, Jungong Han and Xinbo Gao. 2022. Multi-view graph embedding clustering network: Joint self-supervision and block diagonal representation. *Neural Networks* 145, 1-9.

[122] Shunxin Xiao, Shide Du, Zhaoliang Chen, Yunhe Zhang and Shiping Wang. 2023. Dual fusion-propagation graph neural network for multi-view clustering. *IEEE Transactions on Multimedia*. *DOI 10.1109/TMM.2023.3248173.*

[123] Shiping Wang, Zhaoliang Chen, Shide Du and Zhouchen Lin. 2021. Learning deep sparse regularizers with applications to multi-view clustering and semi-supervised classification. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 44, 9, 5042-5055.

[124] Runwu Zhou and Yi-Dong Shen. 2020. End-to-end adversarial-attention network for multi-modal clustering. *In Proceedings of The IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 14619-14628.

[125] Zhenyu Huang, Peng Hu, Joey Tianyi Zhou, Jiancheng Lv and Xi Peng. 2020. Partially view-aligned clustering. *Advances in Neural Information Processing Systems* 33, 2892-2902.

[126] Zihan Fang, Shide Du, Xincan Lin, Jinbin Yang, Shiping Wang and Yiqing Shi. 2023. Dbo-net: Differentiable bi-level optimization network for multi-view clustering. *Information Sciences* 626, 572-585.

[127] 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.

[128] Uno Fang, Jianxin Li, Xuequan Lu, Ajmal Mian and Zhaoquan Gu. 2023. Robust image clustering via context-aware contrastive graph learning. *Pattern Recognition* 138, 109340.

[129] Guowang Du, Lihua Zhou, Zhongxue Li, Lizhen Wang and Kevin Lü. 2023. Neighbor-aware deep multi-view clustering via graph convolutional network. *Information Fusion* 93, 330-343.

[130] Fangfei Lin, Bing Bai, Kun Bai, Yazhou Ren, Peng Zhao and Zenglin Xu. 2022. Contrastive multi-view hyperbolic hierarchical clustering. *arXiv preprint arXiv:2205.02618*. 2022 *Joint Conference on Artificial Intelligence* 3250-3256.

[131] Jie Xu, Huayi Tang, Yazhou Ren, Liang Peng, Xiaofeng Zhu and Lifang He. 2022. Multi-level feature learning for contrastive multi-view clustering. *In Proceedings of The IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16030-16039.

[132] Dunqiang Liu, Shu-Juan Peng, Xin Liu, Lei Zhu, Zhen Cui and Taihao Li. 2022. Inconsistency Distillation For Consistency: Enhancing Multi-View Clustering via Mutual Contrastive Teacher-Student Leaning. *In 2022 IEEE International Conference on Data Mining (ICDM)*, 251-258.

[133] Xiaozhi Deng, Dong Huang, Ding-Hua Chen, Chang-Dong Wang and Jian-Huang Lai. 2023. Strongly augmented contrastive clustering. *Pattern Recognition* 139, 109470.

[134] Ru Wang, Lin Li, Xiaohui Tao, Peipei Wang and Peiyu Liu. 2022. Contrastive and attentive graph learning for multi-view clustering. *Information*

Processing & Management 59, 4, 102967.

[135] Erlin Pan and Zhao Kang. 2021. Multi-view contrastive graph clustering. *Advances in Neural Information Processing Systems* 34, 2148-2159.

[136] Shiping Wang, Xincan Lin, Zihan Fang, Shide Du and Guobao Xiao. 2022. Contrastive consensus graph learning for multi-view clustering. *IEEE/CAA Journal of Automatica Sinica* 9, 11, 2027-2030.

[137] Daniel J Trosten, Sigurd Løkse, Robert Jenssen and Michael C Kampffmeyer. 2023. On the Effects of Self-supervision and Contrastive Alignment in Deep Multi-view Clustering. *In Proceedings of The IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 23976-23985.

[138] Abhishek Kumar and Hal Daumé. 2011. A co-training approach for multi-view spectral clustering. *In Proceedings of The 28th International Conference on Machine Learning*, 393-400.

[139] Avrim Blum and Tom Mitchell. 1998. Combining labeled and unlabeled data with co-training. *In Proceedings of The 11th Annual Conference on Computational Learning Theory*, 92-100.

[140] Zhaohong Deng, Ruixiu Liu, Peng Xu, Kup-Sze Choi, Wei Zhang, Xiaobin Tian, Te Zhang, Ling Liang, Bin Qin and Shitong Wang. 2023. Multi-view clustering with the cooperation of visible and hidden views. *IEEE Transactions on Knowledge and Data Engineering* 34, 2, 803-815.

[141] Sayantan Mitra and Sriparna Saha. 2021. A Multi-task Multi-view based Multi-objective Clustering Algorithm. *In 2020 25th International Conference on Pattern Recognition (ICPR)*, 4720-4727.

[142] Xianchao Zhang, Xiaotong Zhang and Han Liu. 2015. Multi-task multi-view clustering for non-negative data. *In Proceedings of The 24th International Joint Conference on Artificial Intelligence*, 4055-4061.

[143] Xiaotong Zhang, Xianchao Zhang, Han Liu and Xinyue Liu. 2016. Multi-task multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering* 28, 12, 3324-3338.

[144] Zhe Xue, Guorong Li, Shuhui Wang, Jun Huang, Weigang Zhang and Qingming Huang. 2019. Beyond global fusion: A group-aware fusion approach for multi-view image clustering. *Information Sciences* 493, 176-191.

[145] Shudong Huang, Zenglin Xu, Ivor W Tsang and Zhao Kang. 2020. Auto-weighted multi-view co-clustering with bipartite graphs. *Information Sciences* 512, 18-30.

[146] Shizhe Hu, Xiaoqiang Yan and Yangdong Ye. 2020. Dynamic auto-weighted multi-view co-clustering. *Pattern Recognition* 99, 107101.

[147] Ling Zhao, Yunpeng Ma, Shanxiong Chen and Jun Zhou. 2022. Multi-view co-clustering with multi-similarity. *Applied Intelligence*, 1-12.

[148] Feiping Nie, Shaojun Shi and Xuelong Li. 2020. Auto-weighted multi-view co-clustering via fast matrix factorization. *Pattern Recognition* 102, 107207.

[149] Shide Du, Zhanghui Liu, Zhaoliang Chen, Wenyuan Yang and Shiping Wang. 2021. Differentiable Bi-Sparse Multi-View Co-Clustering. *IEEE Transactions on Signal Processing* 69, 4623-4636.

[150] Yazhou Ren, Shudong Huang, Peng Zhao, Minghao Han and Zenglin Xu. 2020. Self-paced and auto-weighted multi-view clustering. *Neurocomputing* 383, 248-256.

[151] Zongmo Huang, Yazhou Ren, Xiaorong Pu, Lili Pan, Dezhong Yao and Guoxian Yu. 2021. Dual self-paced multi-view clustering. *Neural Networks* 140, 184-192.

[152] Hong Tao, Chenping Hou, Xinwang Liu, Tongliang Liu, Dongyun Yi and Jubo Zhu. 2018. Reliable multi-view clustering. *In Proceedings of The AAAI Conference on Artificial Intelligence*, 4123-4130.

[153] Changan Yuan, Yonghua Zhu, Zhi Zhong, Wei Zheng and Xiaofeng Zhu. 2022. Robust self-tuning multi-view clustering. *World Wide Web* 25, 2, 489-512.

[154] Menglei Hu and Songcan Chen. 2018. Doubly aligned incomplete multi-view clustering. *In Proceedings of The 27th International Joint Conference on Artificial Intelligence* 2262-2268.

[155] Mouxing Yang, Yunfan Li, Peng Hu, Jinfeng Bai, Jiancheng Lv and Xi Peng. 2022. Robust multi-view clustering with incomplete information. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 45, 1, 1055-1069.

[156] Liang Zhao, Jie Zhang, Qiuhao Wang and Zhikui Chen. 2021. Dual Alignment Self-Supervised Incomplete Multi-View Subspace Clustering Network. *IEEE Signal Processing Letters* 28, 2122-2126.

[157] Jiaqi Jin, Siwei Wang, Zhibin Dong, Xinwang Liu and En Zhu. 2023. Deep Incomplete Multi-view Clustering with Cross-view Partial Sample and Prototype Alignment. *In Proceedings of The IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11600-11609.

[158] Guoqing Chao, Songtao Wang, Shiming Yang, Chunshan Li and Dianhui Chu. 2022. Incomplete multi-view clustering with multiple imputation and ensemble clustering. *Applied Intelligence* 52, 13, 14811-14821.

[159] Jie Xu, Chao Li, Yazhou Ren, Liang Peng, Yujie Mo, Xiaoshuang Shi and Xiaofeng Zhu. 2022. Deep incomplete multi-view clustering via mining cluster complementarity. *In Proceedings of The AAAI Conference on Artificial Intelligence*, 8761-8769.

[160] Guoqing Chao, Jiangwen Sun, Jin Lu, An-Li Wang, Daniel D Langleben, Chiang-Shan Li and Jinbo Bi. 2019. Multi-view cluster analysis with incomplete data to understand treatment effects. *Information sciences* 494, 278-293.

[161] Jie Wen, Zheng Zhang, Zhao Zhang, Lunke Fei and Meng Wang. 2020. Generalized incomplete multiview clustering with flexible locality structure diffusion. *IEEE Transactions on Cybernetics* 51, 1, 101-114.

[162] Naiyao Liang, Zuyuan Yang, Lingjiang Li, Zhenni Li and Shengli Xie. 2021. Incomplete multiview clustering with cross-view feature transformation. *IEEE Transactions on Artificial Intelligence* 3, 5, 749-762.

[163] Naiyao Liang, Zuyuan Yang, Zhenni Li and Wei Han. 2022. Incomplete multi-view clustering with incomplete graph-regularized orthogonal non-negative matrix factorization. *Applied Intelligence* 52, 13, 14607-14623.

[164] Xiangyu Liu and Peng Song. 2022. Incomplete multi-view clustering via virtual-label guided matrix factorization. *Expert Systems with Applications* 210, 118408.

[165] Zhenjiao Liu, Zhikui Chen, Yue Li, Liang Zhao, Tao Yang, Reza Farahbakhsh, Noel Crespi and Xiaodi Huang. 2023. IMC-NLT: Incomplete multi-view clustering by NMF and low-rank tensor. *Expert Systems with Applications* 221, 119742.

[166] Wei Zhou, Hao Wang and Yan Yang. 2019. Consensus graph learning for incomplete multi-view clustering. *In Advances in Knowledge Discovery and Data Mining: 23rd Pacific-Asia Conference*, 529-540.

[167] Jianlun Liu, Shaohua Teng, Lunke Fei, Wei Zhang, Xiaozhao Fang, Zhuxiu Zhang and Naiqi Wu. 2021. A novel consensus learning approach to incomplete multi-view clustering. *Pattern Recognition* 115, 107890.

[168] Jun Yin and Shiliang Sun. 2023. Incomplete multi-view clustering with reconstructed views. *IEEE Transactions on Knowledge and Data Engineering* 35(3): 2671-2682.

[169] Menglei Hu and Songcan Chen. 2019. One-pass incomplete multi-view clustering. *In Proceedings of The AAAI Conference on Artificial Intelligence*, 3838-3845.

[170] Jun Yin and Shiliang Sun. 2022. Incomplete multi-view clustering with cosine similarity. *Pattern Recognition* 123, 108371.

[171] Xinwang Liu, Xinzhong Zhu, Miaomiao Li, Lei Wang, En Zhu, Tongliang Liu, Marius Kloft, Dinggang Shen, Jianping Yin and Wen Gao. 2020. Multiple Kernel $ k $ k-Means with Incomplete Kernels. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 42, 05, 1191-1204.

[172] Xinwang Liu, Miaomiao Li, Chang Tang, Jingyuan Xia, Jian Xiong, Li Liu, Marius Kloft and En Zhu. 2020. Efficient and effective regularized incomplete multi-view clustering. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 43, 8, 2634-2646.

[173] Xingfeng Li, Quansen Sun, Zhenwen Ren and Yinghui Sun. 2022. Dynamic incomplete multi-view imputing and clustering. *In Proceedings of The 30th ACM International Conference on Multimedia*, 3412-3420.

[174] Dongxue Xia, Yan Yang, Shuhong Yang and Tianrui Li. 2023. Incomplete multi-view clustering via kernelized graph learning. *Information Sciences* 625, 1-19.

[175] Jie Wen, Zheng Zhang, Yong Xu, Bob Zhang, Lunke Fei and Hong Liu. 2019. Unified embedding alignment with missing views inferring for incomplete multi-view clustering. *In Proceedings of The AAAI Conference on Artificial Intelligence*, 5393-5400.

[176] Jing-Hua Yang, Le-Le Fu, Chuan Chen, Hong-Ning Dai and Zibin Zheng. 2023. Cross-view graph matching for incomplete multi-view clustering. *Neurocomputing* 515, 79-88.

[177] Jie Wen, Zheng Zhang, Zhao Zhang, Lei Zhu, Lunke Fei, Bob Zhang and Yong Xu. 2021. Unified tensor framework for incomplete multi-view clustering and missing-view inferring. *In Proceedings of The AAAI Conference on Artificial Intelligence*, 10273-10281.

[178] Shuping Zhao, Lunke Fei, Jie Wen, Jigang Wu and Bob Zhang. 2023. Intrinsic and complete structure learning based incomplete multiview clustering. *IEEE Transactions on Multimedia* 25, 1098-1110.

[179] Hao Wang, Linlin Zong, Bing Liu, Yan Yang and Wei Zhou. 2019. Spectral perturbation meets incomplete multi-view data. *In Proceedings of The 28th International Joint Conference on Artificial Intelligence* 3677-3683.

[180] Jun Yin and Jianwei Jiang. 2023. Incomplete Multi-view Clustering Based on Self-representation. *Neural Processing Letters*, 1-15.

[181] Kaiwu Zhang, Baokai Liu, Shiqiang Du, Yao Yu and Jinmei Song. 2023. Incomplete multi-view clustering based on low-rank representation with adaptive graph regularization. *Soft Computing* 27, 11, 7131-7146.

[182] Xiao Zheng, Xinwang Liu, Jiajia Chen and En Zhu. 2022. Adaptive partial graph learning and fusion for incomplete multi - view clustering. *International Journal of Intelligent Systems* 37, 1, 991-1009.

[183] Ziyu Lv, Quanxue Gao, Xiangdong Zhang, Qin Li and Ming Yang. 2022. View-consistency learning for incomplete multiview clustering. *IEEE Transactions on Image Processing* 31, 4790-4802.

[184] Naiyao Liang, Zuyuan Yang and Shengli Xie. 2022. Incomplete multi-view clustering with sample-level auto-weighted graph fusion. *IEEE Transactions on Knowledge and Data Engineering* 35, 6, 6504-6511.

[185] Jie Wen, Ke Yan, Zheng Zhang, Yong Xu, Junqian Wang, Lunke Fei and Bob Zhang. 2020. Adaptive graph completion based incomplete multi-view clustering. *IEEE Transactions on Multimedia* 23, 2493-2504.

[186] Lusi Li, Zhiqiang Wan and Haibo He. 2021. Incomplete multi-view clustering with joint partition and graph learning. *IEEE Transactions on*

Knowledge and Data Engineering 35, 1, 589-602.

[187] Jiye Liang, Xiaolin Liu, Liang Bai, Fuyuan Cao and Dianhui Wang. 2022. Incomplete multi-view clustering via local and global co-regularization. *Science China Information Sciences* 65, 5, 152105.

[188] Qianli Zhao, Linlin Zong, Xianchao Zhang, Xinyue Liu and Hong Yu. 2020. Incomplete multi-view spectral clustering. *Journal of Intelligent & Fuzzy Systems* 38, 3, 2991-3001.

[189] Jie Wen, Huijie Sun, Lunke Fei, Jinxing Li, Zheng Zhang and Bob Zhang. 2021. Consensus guided incomplete multi-view spectral clustering. *Neural Networks* 133, 207-219.

[190] Wenzhang Zhuge, Chenping Hou, Xinwang Liu, Hong Tao and Dongyun Yi. 2019. Simultaneous Representation Learning and Clustering for Incomplete Multi-view Data. *In Proceedings of The 28th International Joint Conference on Artificial Intelligence*, 4482-4488.

[191] Jie Wen, Yong Xu and Hong Liu. 2018. Incomplete multiview spectral clustering with adaptive graph learning. *IEEE Transactions on Cybernetics* 50, 4, 1418-1429.

[192] Jie Wen, Zheng Zhang, Zhao Zhang, Lei Zhu, Lunke Fei, Bob Zhang and Yong Xu. 2021. Unified tensor framework for incomplete multi-view clustering and missing-view inferring. *In Proceedings of The AAAI Conference on Artificial Intelligence*, 10273-10281.

[193] Shijie Deng, Jie Wen, Chengliang Liu, Ke Yan, Gehui Xu and Yong Xu. 2023. Projective incomplete multi-view clustering. *IEEE Transactions on Neural Networks and Learning Systems*.

[194] Xiao Yu, Hui Liu, Yuxiu Lin, Yan Wu and Caiming Zhang. 2022. Auto-weighted sample-level fusion with anchors for incomplete multi-view clustering. *Pattern Recognition* 130, 108772.

[195] Senhong Wang, Jiangzhong Cao, Fangyuan Lei, Jianjian Jiang, Qingyun Dai and Bingo Wing-Kuen Ling. 2023. Multiple kernel-based anchor graph coupled low-rank tensor learning for incomplete multi-view clustering. *Applied Intelligence* 53, 4, 3687-3712.

[196] Jun Guo and Jiahui Ye. 2019. Anchors bring ease: An embarrassingly simple approach to partial multi-view clustering. *In Proceedings of The AAAI Conference on Artificial Intelligence*, 118-125.

[197] Jun Yin, Runcheng Cai and Shiliang Sun. 2022. Anchor-based incomplete multi-view spectral clustering. *Neurocomputing* 514, 526-538.

[198] Wenjue He, Zheng Zhang, Yongyong Chen and Jie Wen. 2023. Structured anchor-inferred graph learning for universal incomplete multi-view clustering. *World Wide Web* 26, 1, 375-399.

[199] Xia Ji, Lei Yang, Sheng Yao, Peng Zhao and Xuejun Li. 2023. Fast and General Incomplete Multi-view Adaptive Clustering. *Cognitive Computation* 15, 2, 683-693.

[200] Zhenglai Li, Chang Tang, Xiao Zheng, Xinwang Liu, Wei Zhang and En Zhu. 2022. High-order correlation preserved incomplete multi-view subspace clustering. *IEEE Transactions on Image Processing* 31, 2067-2080.

[201] Ming Yin, Xiaohua Liu, Liuyang Wang and Guoliang He. 2023. Learning latent embedding via weighted projection matrix alignment for incomplete multi-view clustering. *Information Sciences* 634, 244-258.

[202] Xuemei Han, Zhenwen Ren, Chuanyun Zou and Xiaojian You. 2022. Incomplete multi-view subspace clustering based on missing-sample recovering and structural information learning. *Expert Systems with Applications* 208, 118165.

[203] Chao Zhang, Huaxiong Li, Caihua Chen, Xiuyi Jia and Chunlin Chen. 2022. Low-rank tensor regularized views recovery for incomplete

multiview clustering. *IEEE Transactions on Neural Networks and Learning Systems*.

[204] Yongli Hu, Cuicui Luo, Boyue Wang, Junbin Gao, Yanfeng Sun and Baocai Yin. 2021. Complete/incomplete multi - view subspace clustering via soft block - diagonal - induced regulariser. *IET Computer Vision* 15, 8, 618-632.

[205] Zhenglai Li, Chang Tang, Xinwang Liu, Xiao Zheng, Wei Zhang and En Zhu. 2021. Tensor-based multi-view block-diagonal structure diffusion for clustering incomplete multi-view data. *In 2021 IEEE International Conference on Multimedia and Expo (ICME)*, 1-6.

[206] Guoli Niu, Youlong Yang and Liqin Sun. 2021. One-step multi-view subspace clustering with incomplete views. *Neurocomputing* 438, 290-301.

[207] Zhiqi Yu, Mao Ye, Siying Xiao and Liang Tian. 2022. Learning missing instances in latent space for incomplete multi-view clustering. *Knowledge-Based Systems* 250, 109122.

[208] Guo Zhong and Chi-Man Pun. 2023. Simultaneous Laplacian embedding and subspace clustering for incomplete multi-view data. *Knowledge-Based Systems* 262, 110244.

[209] Jiyuan Liu, Xinwang Liu, Yi Zhang, Pei Zhang, Wenxuan Tu, Siwei Wang, Sihang Zhou, Weixuan Liang, Siqi Wang and Yuexiang Yang. 2021. Self-representation subspace clustering for incomplete multi-view data. *In Proceedings of The 29th ACM International Conference on Multimedia*, 2726-2734.

[210] Xiang Fang, Yuchong Hu, Pan Zhou and Dapeng Oliver Wu. 2021. Unbalanced incomplete multi-view clustering via the scheme of view evolution: Weak views are meat; strong views do eat. *IEEE Transactions on Emerging Topics in Computational Intelligence* 6, 4, 913-927.

[211] Liang Zhao, Jie Zhang, Tao Yang and Zhikui Chen. 2022. Incomplete multi-view clustering based on weighted sparse and low rank representation. *Applied Intelligence* 52, 13, 14822-14838.

[212] Shaowei Wei, Jun Wang, Guoxian Yu, Carlotta Domeniconi and Xiangliang Zhang. 2020. Deep incomplete multi-view multiple clusterings. *In 2020 IEEE International Conference on Data Mining (ICDM)*, 651-660.

[213] Zhe Xue, Junping Du, Changwei Zheng, Jie Song, Wenqi Ren and Meiyu Liang. 2021. Clustering-Induced Adaptive Structure Enhancing Network for Incomplete Multi-View Data. *In Proceedings of The 30th International Joint Conference on Artificial Intelligence*, 3235-3241.

[214] Qianqian Wang, Zhengming Ding, Zhiqiang Tao, Quanxue Gao and Yun Fu. 2021. Generative partial multi-view clustering with adaptive fusion and cycle consistency. *IEEE Transactions on Image Processing* 30, 1771-1783.

[215] Jie Wen, Zheng Zhang, Zhao Zhang, Zhihao Wu, Lunke Fei, Yong Xu and Bob Zhang. 2020. Dimc-net: Deep incomplete multi-view clustering network. *In Proceedings of The 28th ACM international conference on multimedia*, 3753-3761.

[216] Jie Xu, Chao Li, Liang Peng, Yazhou Ren, Xiaoshuang Shi, Heng Tao Shen and Xiaofeng Zhu. 2023. Adaptive feature projection with distribution alignment for deep incomplete multi-view clustering. *IEEE Transactions on Image Processing* 32, 1354-1366.

[217] Mengying Xie, Zehui Ye, Gan Pan and Xiaolan Liu. 2021. Incomplete multi-view subspace clustering with adaptive instance-sample mapping and deep feature fusion. *Applied Intelligence* 51, 5584-5597.

[218] Jie Wen, Zhihao Wu, Zheng Zhang, Lunke Fei, Bob Zhang and Yong Xu. 2021. Structural deep incomplete multi-view clustering network. *In Proceedings of The 30th ACM International Conference on Information & Knowledge Management*, 3538-3542.

[219] Cai Xu, Ziyu Guan, Wei Zhao, Hongchang Wu, Yunfei Niu and Beilei Ling. 2019. Adversarial incomplete multi-view clustering. *In Proceedings of The 28th International Joint Conference on Artificial Intelligence*, 3933-3939.

[220] Yiming Wang, Dongxia Chang, Zhiqiang Fu, Jie Wen and Yao Zhao. 2022. Incomplete Multiview Clustering via Cross-View Relation Transfer. *IEEE Transactions on Circuits and Systems for Video Technology* 33, 1, 367-378.

[221] Huayi Tang and Yong Liu. 2022. Deep safe incomplete multi-view clustering: Theorem and algorithm. *In International Conference on Machine Learning*, 21090-21110.

[222] Yijie Lin, Yuanbiao Gou, Zitao Liu, Boyun Li, Jiancheng Lv and Xi Peng. 2021. Completer: Incomplete multi-view clustering via contrastive prediction. *In Proceedings of The IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11174-11183.

[223] Yijie Lin, Yuanbiao Gou, Xiaotian Liu, Jinfeng Bai, Jiancheng Lv and Xi Peng. 2022. Dual contrastive prediction for incomplete multi-view representation learning. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 45, 4, 4447-4461.

[224] Yiming Wang, Dongxia Chang, Zhiqiang Fu, Jie Wen and Yao Zhao. 2022. ACTIVE: Augmentation-Free Graph Contrastive Learning for Partial Multi-View Clustering. *CoRR*. DOI 10 1109 TMM. 022.3210376.

[225] Zhe Xue, Junping Du, Hai Zhu, Zhongchao Guan, Yunfei Long, Yu Zang and Meiyu Liang. 2022. Robust diversified graph contrastive network for incomplete multi-view clustering. *In Proceedings of The 30th ACM International Conference on Multimedia*, 3936-3944.

[226] Yi Zhang, Xinwang Liu, Siwei Wang, Jiyuan Liu, Sisi Dai and En Zhu. 2021. One-stage incomplete multi-view clustering via late fusion. *In Proceedings of The 29th ACM International Conference on Multimedia*, 2717-2725.

[227] Wei Zhang, Zhaohong Deng, Kup-Sze Choi and Shitong Wang. 2023. End-to-end Incomplete Multiview Fuzzy Clustering with Adaptive Missing View Imputation and Cooperative Learning. *IEEE Transactions on Fuzzy Systems* 31, 5, 1445-1459.

[228] Longqi Yang, Liangliang Zhang and Yuhua Tang. 2021. Online binary incomplete multi-view clustering. *In Machine Learning and Knowledge Discovery in Databases: European Conference*, 75-90.

[229] Xiang Fang, Yuchong Hu, Pan Zhou and Dapeng Oliver Wu. 2020. V $^ 3$ H: View variation and view heredity for incomplete multiview clustering. *IEEE Transactions on Artificial Intelligence* 1, 3, 233-247.

[230] Krishna Kumar Sharma and Ayan Seal. 2021. Outlier-robust multi-view clustering for uncertain data. *Knowledge-Based Systems* 211, 106567.

[231] Leo L Duan. 2020. Latent simplex position model: High dimensional multi-view clustering with uncertainty quantification. *The Journal of Machine Learning Research* 21, 1, 1411-1435.

[232] Yale Chang, Junxiang Chen, Michael H Cho, Peter J Castaldi, Edwin K Silverman and Jennifer G Dy. 2017. Multiple clustering views from multiple uncertain experts. *In International Conference on Machine Learning*, 674-683.

[233] Hongwei Yin, Wenjun Hu, Zhao Zhang, Jungang Lou and Minmin Miao. 2021. Incremental multi-view spectral clustering with sparse and connected graph learning. *Neural Networks* 144, 260-270.

[234] Peng Zhou, Yi-Dong Shen, Liang Du, Fan Ye and Xuejun Li. 2019. Incremental multi-view spectral clustering. *Knowledge-Based Systems* 174, 73-86.

[235] Yanyong Huang, Kejun Guo, Xiuwen Yi, Zhong Li and Tianrui Li. 2023. Incremental unsupervised feature selection for dynamic incomplete multi-view data. *Information Fusion* 96, 312-327.

[236] Amir Rosenfeld and John K Tsotsos. 2018. Incremental learning through deep adaptation. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 42, 3, 651-663.

[237] Adil Fahad, Najlaa Alshatri, Zahir Tari, Abdullah Alamri, Ibrahim Khalil, Albert Y Zomaya, Sebti Foufou and Abdelaziz Bouras. 2014. A survey of clustering algorithms for big data: Taxonomy and empirical analysis. *IEEE Transactions on Emerging Topics in Computing* 2, 3, 267-279.

[238] Chang-Dong Wang, Jian-Huang Lai and S Yu Philip. 2015. Multi-view clustering based on belief propagation. *IEEE Transactions on Knowledge and Data Engineering* 28, 4, 1007-1021.

[239] Zhenfan Wang, Xiangwei Kong, Haiyan Fu, Ming Li and Yujia Zhang. 2015. Feature extraction via multi-view non-negative matrix factorization with local graph regularization. *In 2015 IEEE International Conference on Image Processing (ICIP)*, 3500-3504.

[240] Kun Zhan, Jinhui Shi, Jing Wang, Haibo Wang and Yuange Xie. 2018. Adaptive structure concept factorization for multiview clustering. *Neural Computation* 30, 4, 1080-1103.

[241] Kun Zhan, Feiping Nie, Jing Wang and Yi Yang. 2018. Multiview consensus graph clustering. *IEEE Transactions on Image Processing* 28, 3, 1261-1270.

[242] Rongkai Xia, Yan Pan, Lei Du and Jian Yin. 2014. Robust multi-view spectral clustering via low-rank and sparse decomposition. *In Proceedings of The AAAI Conference on Artificial Intelligence*, 2149--2155.

[243] Feiping Nie, Lai Tian and Xuelong Li. 2018. Multiview clustering via adaptively weighted procrustes. *In Proceedings of The 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2022-2030.

[244] Xiaochun Cao, Changqing Zhang, Huazhu Fu, Si Liu and Hua Zhang. 2015. Diversity-induced multi-view subspace clustering. *In Proceedings of The IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 586-594.

[245] Xiaobo Wang, Xiaojie Guo, Zhen Lei, Changqing Zhang and Stan Z Li. 2017. Exclusivity-consistency regularized multi-view subspace clustering. *In Proceedings of The IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 923-931.

[246] Shirui Luo, Changqing Zhang, Wei Zhang and Xiaochun Cao. 2018. Consistent and specific multi-view subspace clustering. *In Proceedings of 32nd AAAI Conference on Artificial Intelligence* 3730-3737.

[247] Mahdi Abavisani and Vishal M. Patel. 2018. Deep Multimodal Subspace Clustering Networks. *IEEE Journal of Selected Topics in Signal Processing* 12, 6, 1601-1614.

[248] Jie Xu, Yazhou Ren, Huayi Tang, Zhimeng Yang, Lili Pan, Yang Yang and Xiaorong Pu. 2023. Self-supervised discriminative feature learning for multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering* 35, 7, 7470-7482.

[249] Guang-Yu Zhang, Chang-Dong Wang, Dong Huang, Wei-Shi Zheng and Yu-Ren Zhou. 2018. TW-Co-k-means: Two-level weighted collaborative k-means for multi-view clustering. *Knowledge-Based Systems* 150, 127-138.

[250] Man-Sheng Chen, Ling Huang, Chang-Dong Wang, Dong Huang and Jian-Huang Lai. 2021. Relaxed multi-view clustering in latent embedding space. *Information Fusion* 68, 8-21.

[251] Youwei Liang, Dong Huang, Chang-Dong Wang and Philip S. Yu. 2022. Multi-view graph learning by joint modeling of consistency and inconsistency. *IEEE Transactions on Neural Networks and Learning Systems*.

APPENDIX (E-PUB ONLY)

Table 1 the characteristics of representative approaches for MVC

Category	Approach name	Motivation	Model structure	NMF objects	Fusion strategy	Weighting strategy	Clustering routine	Clustering method	Time complexity	Memory cost	Peculiarity
NMF	MVOCNMF [48]	high-quality	shallow	original data	early-fusion	parameter weighted-free	two-step routine	k-means	$O(n^2)$	$O\left(n\sum_{v=1}^{m} d_v\right)$	label information, orthogonality constraint, co-regularization
	MVC-DMF-PA[59]	hierarchy and non-linearity	deep	original data	late-fusion	parameter-weighted	one-step routine	-	$O(n^2)$	$O\left(m\sum_{v=1}^{m} d_v n\right)$	deep NMF, view-specific structure, partition alignment
Multiple kernel learning	MVCMK [66]	robustness	shallow	-	direct-fusion	parameter-free weighted	one-step routine	-	$O(n^3)$	$O\left(\sum_{v=1}^{m} (d_v)^2\right)$	learning in kernel spaces, joint optimization, direct partition,
	JLMVC [67]	non-linearity	shallow	-	early-fusion	parameter-free weighted	two-step routine	spectral clustering	$O(n^2 \log n)$	$O(mn^2)$	kernel representation tensor, tensor nuclear norm, kernel trick, joint optimization
Graph	PCGL [70]	high-quality	shallow	-	direct-fusion	parameter weighted-free	one-step routine	-	$O(n^2)$	$O\left(n^2+\sum_{v=1}^{m} d_v\right)$	dimensionality reduction, manifold structure learning, feature selection, joint optimization
	CNESE [79]	efficiency	shallow	-	direct-fusion	parameter-free weighted	one-step routine	-	$O(n^2)$	$O\left(\sum_{v=1}^{m} nd_v + mn^2\right)$	integration of nonnegative embedding and spectral embedding, cluster indices smoothness, orthogonality constraint
Subspace	SSSL-M [92]	robustness	deep	-	early-fusion	equal-weighted	one-step routine	-	$O(n^3)$	$O\left(mn^2\right)$	anti-block-diagonal indicator matrix, semi-supervision, block-diagonal structure
	MvSC-MRAR [114]	non-linearity	deep	-	early-fusion	equal-weighted	two-step routine	spectral clustering	$O(n^2)$	$O\left(\sum_{v=1}^{m} nd_v + p+\sum_{v=1}^{m} n^2\right)$	multiple self-expressive layers, multiple deep auto-encoders, multi-level representations, universal adversarial discriminator
Deep learning	DMJC [120]	high-quality	deep	-	early-fusion	parameter-free weighted	one-step routine	-	$O(mB^2)$	$O\left(\sum_{v=1}^{m} nd_v + Kmd_v\right)$	implicit and explicit fusion, joint learning
	MVGC [121]	high-quality	deep	-	early-fusion	equal-weighted	two-step routine	spectral clustering	$O(n^2 + nlogn)$	$O\left(\sum_{v=1}^{m} nd_v + mn^2\right)$	Euler transform, augmentation of the node attribute, block diagonal representation constraint, self-expression coefficient matrix
Contrastive learning	NMvC-GCN [129]	high-quality	deep	-	early-fusion	parameter-weighted-free	one-step routine	-	$O(n\log n + Kn)$	$O(mn)$	neighbor information, unsupervised optimization of GCN, joint optimization, consensus regularization,
	MFLVC [69]	robustness	deep	-	fusion-free	equal-weighted	one-step routine	-	$O(n)$	$O(Kmn)$	multi-level feature learning, fusion-free
Co-	MTMV-MO [141]	high-quality	shallow	-	direct-fusion	equal-weighted	one-step routine	-	$O(K^2n^2)$	$O(mKn^2)$	within-view task relation, within-task view relation,

learning											clustering quality, multi-objective optimization
	DBMC and EDBMC [149]	interpretability	deep	-	early-fusion	equal-weighted	one-step routine		$O(X)$	$O\left(\sum_{v=1}^{m} nmd_v + n^2 m\right)$	interpretable and consistent collaborative representations, sparsity in the dual space of features and samples, deep differentiable network
Self-paced learning	SAMVC [150]	robustness	shallow	-	late-fusion	parameter-free weighted	one-step routine	-	$O(mnK)$	$O\left(\sum_{v=1}^{m} nd_v\right)$	non-convexity, noises, soft weighting scheme, auto-weighted
	DSMVC [151]	robustness	shallow		late-fusion	parameter-free weighted	one-step routine	-	$O(n)$	$O\left(\sum_{v=1}^{m} nd_v + n^2\right)$	non-convexity, noises, soft-weighting scheme, dual self-paced learning for instances and feature selection

Table 2 the characteristics of representative approaches for IMVC

Category	Approach name	Motivation	Missing	Imputing	Model structure	Fusion strategy	Weighting strategy	Clustering routine	Clustering method	Time complexity	Memory cost	Peculiarity
NMF	IMCRV [48]	high-quality	view	reconstruct views	shallow	early-fusion	parameter weighted	two-step routine	k-means	$O(mn^3)$	$O(mn^2)$	Laplacian regularization, global property of latent representation
	IMCCS [170]	efficiency	view	-	shallow	early-fusion	equal-weighted	two-step routine	k-means	$O(\sum_{v=1}^{m} d_v n^2)$	$O(mn^2)$	direct calculation of the cosine similarity in the original multi-view space to preserve manifold structure
Multiple kernel learning	KGIMC [174]	non-linearity	instance	kernel completion	shallow	direct-fusion	parameter-free weighted	one-step routine	-	$O(n^2)$	$O(n^2)$	joint optimization, self-expression learning in kernel space, multi-kernel learning
	EE-IMVC [172]	efficiency	view	each base clustering matrix	shallow	late-fusion	parameter-free weighted	two-step routine	k-means	$O(K^3 + (n-n_p)K^2)$	$O(mnK)$	prior knowledge regularization, maximal alignment between the consensus clustering matrix and an adaptively weighted base clustering matrices with an optimal permutation, SVD
Graph	([183])	stability	view, value	view-specific partial graph	shallow	direct-fusion	parameter weighted	one-step routine	-	$O(n^3)$	$O(\sum_{v=1}^{m} d_v n)$	integration of graph learning and spectral clustering, tensor Schatten p-norm, view-(un)important-content
	APGLF [182]	high-quality	feature	-	shallow	late-fusion	equal-weighted	one-step routine	-	$O(n^3 + n_p^3)$	$O(mn^3)$	Within-view partial graph learning, cross-view graph fusion, rank constraint on the graph Laplacian matrix, joint optimization
Subspace	MVC-SBD/IMVC-SBD) [204]	robustness	view	-	shallow	early-fusion	equal-weighted	two-step routine	normalized cut (NCuts) or k-means	$O(n^6)$	$O(n^2)$	in(complete) multi-view data, soft block-diagonal-induced regulariser, introduction of multiple indicator matrices into self-representation model
	LNRLSD [24]	high-quality	view	-	shallow	early-fusion	equal-weighted	two-step routine	spectral clustering	$O(n^3)$	$O\left(\sum_{v=1}^{m} nd_v + n^2\right)$	data description in local manner, block-diagonal structure

Deep learning	APADC[216]	high-quality	view	-	deep	early-fusion	parameter weighted-free	two-step routine	k-means	$O(n_p / n^3)$	$O(Mnm)$	imputation-free via adaptive feature projection, view-specific features via autoencoders, common cluster information via maximal mutual information, feature distributions alignment via minimal the mean discrepancy
	GP-MVC[214]	high-quality	view	recovery data	deep	early-fusion	equal-weighted	two stage routine	-	$O(n^2)$	$O(n^2)$	high-level features and high-order geometric structure via view-specific graph convolutional encoder networks,
Contrastive learning	COMPLETER [222]	high-quality	view	recovery data	deep	early-fusion	equal-weighted	two-step routine	k-means	$O(\sum_{v=1}^{m} nd_v)$	$O(nm)$	incorporation consistent representation learning and cross-view data recovery, consistent representation via maximal the mutual information, missing views recovery via minimal the conditional entropy
	SURE[155]	robustness	sample	recovery feature	deep	early-fusion	equal-weighted	two-step routine	k-means	$O(nB^2)$	$O(nm)$	partially view-unaligned problem (PVP/partially sample-missing problem (PSP), noise-robust contrastive loss, noisy correspondence problem via noisy labels (false-negative pairs, FNPs)

Table 3 the characteristics of representative approaches for uncertain MVC

Approach name	Motivation	Model structure	Fusion strategy	Weighting strategy	Clustering routine	Clustering method	Time complexity	Memory cost	Peculiarity
MSCUO[26]	high-quality	shallow	late-fusion	equal-weighted	two-step routine	randomized *k*-means	$O(n^3)$	$O(\sum_{v=1}^{m} nd_v K)$	fusion of induced kernel distance and Jeffrey-divergence w.r.t overlap degree, self-adaptive mixture similarity measure (SAM), pairwise co-regularization
OMVC[230]	robustness	shallow	late-fusion	parameter weighted	one-step routine	-	$O(\hat{r}Em)$	$O(\sum_{v=1}^{m} nd_v + nK)$	self-adaptive mixture similarity function based on geometric distance and S-divergence, threshold-based residual objective function in k-medoids

Table 4 the characteristics of representative approaches for dynamic MVC

Approach name	Motivation	Model structure	Fusion strate	Weighting strategy	Clustering routine	Clustering method	Time complexity	Memory cost	Peculiarity

			gy						
SCGL[233]	scalability	shallow	early-fusion	parameter-weighted-free	two-step routine	spectral clustering	$O(mn^3)$	$O(mn^2)$	store of only one consensus similarity matrix, integration of the sparse graph learning and the connected graph learning
IMSC[234]	scalability	shallow	early-fusion	equal-weighted	two-step routine	spectral rotation	$O(nkr + m(nr_{\max}^2 + r_{\max}^3))$	$O(mnr_{\max} + n(K+k))$	updating of initial model, low-rank approximation via random Fourier features, base kernels construction, low rank SVD decompositions

Note: n : Number of samples, m : Number of views, $\hat{k}$: Number of neighbors, $\hat{r}$: Number of uncertain data, d_v : Dimension of the v-th view, n_p : Number of missing samples, r ($r_{\max}$): rank of matrix (maximal rank), B : Batch size, K : Number of clusters, E : the complexity of evaluating similarity measure between two uncertain data, M : Ratio of positive and negative samples, $X = \sum_{v=1}^{m} \left(\max(n^2 d_v, (d_v)^3 + + n^3 \right)$.

Table 5~9 summarize the main statistics of different datasets, and gives the feature types and feature dimensions of each view, where "#views", "#classes", "#instances", and "F-Type(#view v)" denotes the number of views, the number of clusters, the number of instances, and the type as well as dimension of feature in the v-th view, respectively. For example, "intensity (4096)" indicates the feature type is "intensity" and the corresponding feature dimension of the current view is 4096. It can be seen from Table 20~23 that sometimes the same domain has multiple datasets, such as handwritten digits have datasets UCI, Digits, HW2source, Handwritten, MNIST-USPS, MNIST-10000, Noisy MNIST-Rotated MNIST. These datasets may use different feature types, different numbers of views, or different numbers of instances. "x/y" indicates two values of a variable, for example, the entry of "#views" of BBCSport in Table 5, "2/3", indicates that some datasets contain 2 views while some datasets contain 3 views. As observed, the numbers of instances, views, and clusters vary within large intervals, which supply a good platform to compare the performance of different clustering algorithms.

Table 5 Statistics and multi-view features (# dimensions) of the text datasets

	Dataset	#vie	#class	#instan	F-Type	F-Type	F-Type	F-Type	F-	F-

		ws	es	ces	(#View1)	(#View2)	(#View3)	(#View4)	Type (#View5)	Type (#View6)
news	3Sources[16]	3	6	169	BBC(3560)	Reuters(3631)	Guardian (3068)			
	BBC[75]	4	5	685	seg1(4659)	seg2(4633)	seg3(4665)	seq4(4684)		
	BBCSport[250]	2/3	5	544/282	seq1 (3183/2582)	seg2(3203/2544)	/seq3(2465)			
	Newsgroup(text) [60, 74]	3	5	500	(2000)	(2000)	(2000)			
multilingual documents	Reuters [68, 250]	3/5	6	600/1200	English (9749/2000)	French (9109/2000)	German (7774/2000)	/ Italian (2000)	/ Spanish (2000)	
	Reuters-21578 [251]	5	6	1500	English (21531)	French (24892)	German (34251)	Italian (15506)	Spanish (11547)	
citations	Citeseers(text) [250]	2	6	3312	Citations (4732)	word vector (3703)				
webpage	Web KB [151] Cornell	2	5	195	Citation (195)	Content (1703)				
	Web KB [151] Texas	2	5	187	Citation (187)	Content (1398)				
	Web KB [151] Washington	2	5	230	Citation (230)	Content (2000)				
	Web KB [151] Wisconsin	2	5	265	Citation (265)	Content (1703)				
articles	Wikipedia [44]	2	10	693						
diseases	Derm [90]	2	6	366	Clinical (11)	Histopathological (22)				

Table 6 Statistics and multi-view features (# dimensions) of the image datasets

	Dataset	#views	#classes	#instances	F-Type (#View1)	F-Type (#View2)	F-Type (#View3)	F-Type (#View4)	F-Type (#View5)	F-Type (#View6)
face	Yale [250]	3	15	165	Intensity (4096)	LBP(33040	Gabor (6750)			
	Yale-B [68]	3	10	650	Intensity(2500)	LBP(3304)	Gabor(6750)			
	Extended-Yale [79]	2	28	1774	LBP(900)	COV(45)				
	VIS/NIR [68]	2	22	1056	VL(10000)	NIRI(10000)				
	ORL_v1 [250]	3	40	400	Intensity(4096)	LBP(3304)	Gabor(6750)			
	ORL_v2 [86]	4	40	400	GIST(512)	LBP (59)	HOG(864)	Centrist(254)		
	Notting-Hill [68]	3	5	550	Intensity(2000)	LBP(3304)	Gabor(6750)			
	YouTube Faces [56]	3	66	152,549	CH(768)	GIST(1024)	HOG(1152)			
Handwritten digits	UCI [64]	3	10	2000	FAC (216)	FOU (76)	KAR(64)			
	Digits [90]	3	10	2000	FAC(216)	FOU(76)	KAR (64)			
	HW2sources [75]	2	10	2000	FOU (76)	PIX (240)				
	Handwritten [148]	6	10	2000	FOU(76)	FAC(216)	KAR(64)	PIX(240)	ZER(47)	MOR(6)
	MNIST-USPS [118]	2	10	5000	MNIST(28×28)	USPS(16×16)				
	MNIST-10000_v1 [79]	2	10	10000	VGG16 FC1(4096)	Resnet50(2048)				
	MNIST-10000_v2 [148]	3	10	10000	ISO(30)	LDA(9)	NPE(30)			
	Noisy MNIST-	2	10	70000	Noisy	Rotated				

	Rotated MNIST [118]				MNIST(28×28)	MNIST(28×28)				
	NUSWIDEObj [56, 68]	5	31	30,000	CH(65)	CM(226)	CORR(145)	ED(74)	WT(129)	
	NUSWIDE [148]	6	7	2400	CH(64)	CC(144)	EDH(73)	WAV(128)	BCM(255)	SIFT(500)
object	MSRC [152]	5	7	210	CM(48)	LBP(256)	HOG(100)	SIFT(200)	GIST(512)	
	MSRCv1 [250]	5	7	210	CM(24)	HOG(576)	GIST(512)	LBP(256)	GENT(254)	
	COIL-20 [71]	3	20	1440	Intensity(1024)	LBP (3304)	Gabor (6750)			
	Caltech101-7/20/102 [56, 118, 148]	6	7/20/102	1474/2386/9144	Gabor(48)	WM(40)	Centrist (254)	HOG(1984)	GIST(512)	LBP(928)
Multimedia	MSRA-MM2.0 [51]	4	25	5000	HSV-CH(64)	CORRH(144)	EDH(75)	WT(128)		
	Scene [86]	4	8	2688	GIST(512)	CM(432)	HOG(256)	LBP(48)		
Scene	scene-15 [65]	3	15	4485	GIST(1800)	PHOG(1180)	LBP(1240)			
	Out-Scene [79]	4	8	2688	GIST(512)	LBP(48)	HOG(256)	CM(432)		
	Indoor [152]	6	5	621	SURF(200)	SIFT(200)	GIST(512)	HOG(680)	WT(32)	
plant species	100leaves[61]	3	100	1600	TH(64)	FSM(64)	SD(64)			
Animal	Animal with attributes[115, 148]	6	50	4000/30475	CH(2688)	LSS(2000)	PHOG(252)	SIFT(2000)	RGSIFT (2000)	(2000)
multi-temporal remote sensing	Forest [90]	2	4	524	RS(9)	GWSV(18)				
Fashion	Fashion-10K [118]	2	10	70000	Test set(28×28)	sampled set (28×28)				
sports event	Event [152]	6	8	1579	SURF(500)	SIFT(500)	GIST(512)	HOG(680)	WT(32)	LBP(256)
	ALOI [74]	4	100	110250	RGB-CH(77)	HSV-CH(13)	CS(64)	Haralick (64)		
	ImageNet [40]	3	50	12000	HSV-CH(64)	GIST(512)	SIFT(1000)			
	Corel [152]	3	50	5000	CH (9)	EDH(18)	WT (9)			
image	Cifar-10 [56]	3	10	60,000	CH(768)	GIST (1024)	HOG(1152)			
	SUN1k [152]	3	10	1000	SIFT(6300)	HOG(6300)	TH(10752)			
	Sun397 [56]	3	397	108,754	CH(768)	GIST (1024)	HOG(1152)			

Note: BCM: block-wise color moment, CC: color correlogram, CH: color histogram, CORR: color correlation, CORRH: color correlogram, COV: covariance descriptor, CS: color similarity, CM: color moment, ED: edge distribution, EDH: edge direction histogram, FAC: profile correlations, FOU: Fourier coefficients of the character shapes, FSM: fine-scale margin, GENT: Centrist feature, GIST: abstract representation of the scene, GWSV: geographically weighted similarity variables, HOG: histogram of oriented gradients, HSV-CH: HSV color histograms, ISO: isometric projection, KAR/KLC: Karhunen-Love coefficients, LBP: local binary pattern, LDA: linear discriminant analysis, LSS: local self-similarity, MOR: morphological features, NPE: neighborhood preserving embedding, PHOG: pyramid HOG, PIX: pixel averages in 2 ×3 windows, Resnet50: Residual neural network, RGB-CH: RGB color histograms, RGSIFT: color SIFT, SIFT: Scale-invariant feature transform, SURF: speeded up robust features, TH: texture histogram, VGG16 FC1: Visual Geometry Group from Oxford, VL: visible light, NIRI: near-IR illumination, RS: reflected spectral, SD: shape descriptor, WAV: wavelet texture, WM: wavelet moments, WT: wavelet texture, ZER: Zernike moment.

Table 7 Statistics and multi-view features (# dimensions) of the text-gene datasets

	Dataset	#views	#classes	#instances	F-Type (#View1)	F-Type (#View2)	F-Type (#View3)	F-Type (#View4)	F-Type (#View5)	F-Type (#View6)
Prokaryotic Species	Prok [90]	3	4	551	Textual TF-IDF re-weighted word frequencies	Genomic-the proteome composition	Genomic - the gene repertoire			

Note: TF-IDF: term frequency–inverse document frequency.

Table 8 Statistics and multi-view features (# dimensions) of the image-text datasets

	Dataset	#views	#classes	#instances	F-Type (#View1)	F-Type (#View2)	F-Type (#View 3)	F-Type (#View 4)	F-Type (#View 5)	F-Type (#View 6)
articles	Wikipedia [44, 90]	2	10	693/2866	image	article				
drosophila embryos	BDGP [118]	5		2500	Visual(1750)	Textual(79)				
NBA-NASCAR Sport	NNSpt [152]	2	2	840	image(1024)	TF-IDF(296)				
indoor scenes	SentencesNYU v2 (RGB-D) [124]	2	13	1449	image (2048)	text (300)				
Pascal	VOC [124]	2	20	5,649	Image: Gist (512)	Text (399)				
object	NUS-WIDE-C5(NWC) [124]	2	5	4000	visual codeword vector(500)	annotation vector(1000)				
photographic images	MIR Flickr 1M [73]	4	10	2000	HOG(300)	LBP(50)	HSV CORRH (114)	TF-IDF(60)		

Table 9 Statistics and multi-view features (# dimensions) of the video datasets

	Dataset	#views	#classes	#instances	F-Type (#View1)	F-Type (#View2)	F-Type (#View 3)	F-Type (#View 4)	F-Type (#View 5)	F-Type (#View 6)
actions of passengers	DTHC [90]	3 cameras	Dispersing from the center quickly	3 video sequences	151 frames/video	resolution 135 × 240				
pedestrian video shot	Lab [90]	4 cameras	4 people	16 video sequences	3915 frames/video	resolution 144 ×180				
Motion of Body	CMU Mobo [73]	4 videos	24 objects	96 video sequences	about 300 frames/video	resolution 40×40				
face video	Honda/UCSD [73]	at least 2	20 objects	59 video sequenc	12 to 645 frames/vid	resolution 20×20				

sequences		videos / person		es	eo				
	YouTubeFace_sel [115]	5	31	101499	64	512	64	647	838
Columbia Consumer video	CCV[124]	3	20	6773 YouTube videos	SIFT(5000)	STIP(5000)	MFCC (4000)		

Note: MFCC: Mel-scale frequency cepstral coefficients, STIP: spatial-temporal interest points.

Table 10 The formulas of evaluation indicators

	Category	Formulas	Symbols
Internal indices [237]	Compactness (CP)	$\overline{CP} = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{\mid \Omega_k \mid}\sum_{x_k \in \Omega_k} \parallel x_k - c_k \parallel$	K : number of clusters Ω_k : set of samples in the k-th cluster c_k : centroid of the k-th cluster S_i : average distance from samples within the i-th cluster to cluster centroid C_i
	Davies-Bouldin Index (DBI)	$DBI = \frac{1}{K}\sum_{i=1}^{K}\max_{j \neq i}\left(\frac{\overline{S_i} + \overline{S_j}}{\left\|c_i - c_j\right\|_2}\right)$	
	Dunn Validity Index (DVI)	$DVI = \frac{\min\limits_{0<m\neq n<K}\left\{\min\limits_{\substack{\forall x_i \in \Omega_m \\ \forall x_j \in \Omega_n}}\left\{\left\|x_i - x_j\right\|\right\}\right\}}{\max\limits_{0<m\leq K}\max\limits_{\forall x_i, x_j \in \Omega_m}\left\{\left\|x_i - x_j\right\|\right\}}$	
	Separation (SP)	$\overline{SP} = \frac{2}{K^2 - K}\sum_{i=1}^{K}\sum_{j=i+1}^{K}\left\|c_i - c_j\right\|_2$	
External indexes [51, 238]	clustering accuracy (ACC)	$ACC = \sum_{i=1}^{K}\frac{\max(\Omega_i \mid Y_i)}{\mid \Omega \mid}$	Ω : set of clusters L_i : class labels for all samples in the i-th cluster $\max(\Omega_i \mid Y_i)$: number of samples with the majority label in the i-th cluster $H(\Omega)$: entropy of cluster set Ω $MI(\Omega, \Omega')$: mutual information between Ω and Ω' Y_i : set of classes n_{11} : number of pairs of samples that are in the same cluster in both n_{00} : number of pairs of samples that are indifferent cluster n_{10} : number of pairs of samples that are in the same cluster in A, but in different clusters in B n_{01} : number of pairs of samples that are indifferent clusters in A, but in the same cluster in B TP/TN: number of samples predicted as 1/0 and correctly predicted FP/FN: number of samples predicted as 1/0 and incorrectly predicted
	normalized mutual information (NMI)	$NMI = \frac{MI(\Omega, \Omega')}{\max(H(\Omega), H(\Omega'))}$	
	Purity	$Purity = \frac{1}{n}\sum_{k}\max_{j}\left\|\Omega_k \cap Y_j\right\|$	
	Adjusted Rand Index (ARI)	$ARI = \frac{n_{11} + n_{00}}{n_{00} + n_{01} + n_{10} + n_{11}}$	
	Precision	$Precision = \frac{TP}{TP + FP}$	
	Recall	$Recall = \frac{TP}{TP + FN}$	
	F-score (F1)	$F1 = \frac{Precision \times Recall}{Precision + Recall}$	

Table 11 ACC and NMI with different MVC approaches

Approach	Dataset													
	BBCSport		Reuters		Reuters-21578		NUSWIDE		Caltech101-7		RGB-D		NUSWIDE30K	
	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI
SCBest	0.953	0.875	0.332	0.175	0.549	0.335	0.143	0.173	0.658	0.427	0.462	0.391	0.131	0.146
SCCat	0.966	0.906	0.280	0.145	0.518	0.320	0.166	0.189	0.449	0.329	0.455	0.39	0.152	0.147
MultiNMF [16]	0.860	0.742	0.424	0.216	NAN	NAN	0.124	0.076	0.502	0.428	0.405	0.373	OV	OV
MultiGNMF [239]	0.445	0.127	0.256	0.133	NAN	NAN	0.124	0.076	0.685	0.653	0.496	0.410	OOM	OOM
MVCC[44]	0.822	0.733	0.435	0.256	0.463	0.264	0.171	0.143	0.468	0.505	0.411	0.289	OOM	OOM
MVCF [240]	0.664	0.460	NAN	NAN	0.472	0.279	NA	NA	0.413	0.524	0.436	0.351	OOM	OOM
MvCDMF[63]	0.683	0.510	NAN	NAN	0.313	0.129	NA	NA	0.567	0.506	0.173	0.063	OV	OV
MCDCF[58]	0.801	0.739	0.274	0.171	0.493	0.346	0.159	0.205	0.652	0.610	0.434	0.330	OOM	OOM
RMSC [242]	0.877	0.815	0.479	0.283	0.488	0.323	0.141	0.169	0.639	0.374	0.332	0.269	OOM	OOM
AMGL [32]	0.359	0.145	0.181	0.029	0.301	0.029	0.163	0.181	0.661	0.561	0.533	0.440	OOM	OOM
SwMC [30]	0.362	0.155	0.188	0.045	0.301	0.037	0.151	0.071	0.443	0.161	0.333	0.135	OOM	OOM
MCGC [241]	0.958	0.895	0.273	0.127	0.331	0.132	0.161	0.172	0.662	0.522	0.38	0.247	OOM	OOM
AWP [243]	0.918	0.844	0.269	0.136	0.482	0.283	0.147	0.165	0.571	0.467	0.370	0.221	0.152	0.115
DiMSC [244]	0.859	0.707	0.396	0.181	0.533	0.379	0.092	0.099	0.462	0.427	0.388	0.311	OOM	OOM
ECMSC [245]	0.320	0.026	0.230	0.143	0.298	0.035	0.155	0.184	0.521	0.519	0.382	0.349	OOM	OOM
CSMSC [246]	0.827	0.686	0.465	0.226	0.542	0.252	**0.241**	0.163	0.728	0.648	0.554	0.381	OOM	OOM
DMSCN [247]	0.888	0.813	0.574	0.324	0.518	0.337	0.154	0.179	0.669	0.586	0.548	0.385	OOM	OOM
FMR [107]	0.886	0.755	0.508	0.302	0.578	0.388	0.172	0.211	0.789	0.405	0.387	0.292	OV	OV
SM2VC [95]	0.881	0.761	0.498	0.277	0.382	0.148	0.151	0.181	0.451	0.541	0.283	0.175	OOM	OOM
DSS-MSC [97]	NAN	NAN	0.529	0.343	0.500	0.353	0.152	0.175	0.613	0.636	0.482	0.384	OOM	OOM
MvDSCN[109]	0.931	0.835	0.611	0.359	0.586	0.364	0.159	0.181	0.711	0.623	0.561	0.398	OOM	OOM
MvSC-MRAR [114]	0.958	0.874	**0.625**	0.394	**0.629**	**0.395**	0.201	**0.219**	**0.892**	**0.743**	**0.618**	**0.471**	OOM	OOM
SDMVC [248]	0.729	0.498	0.569	0.327	0.584	0.384	0.175	0.197	0.558	0.570	0.412	0.376	0.161	0.151
CoMVC [127]	0.615	0.385	0.447	0.223	0.482	0.329	0.173	0.205	0.521	0.447	0.468	0.329	0.152	0.117
MVC-MAE [117]	0.931	0.806	0.475	0.263	0.501	0.289	0.177	0.217	0.446	0.619	0.374	0.321	**0.172**	**0.169**
NMvC-GCN [129]	**0.978**	**0.952**	0.583	**0.401**	0.610	0.391	0.185	0.207	0.713	0.508	0.476	0.393	OOM	OOM
CoregSC [15]	0.433	0.225	0.274	0.149	0.523	0.324	0.147	0.189	0.588	0.495	0.461	0.316	OOM	OOM
TW-Co-k-means [249]	0.424	0.126	0.301	0.106	0.373	0.085	0.158	0.192	0.500	0.364	0.351	0.206	0.116	0.071
DWMVC [37]	0.858	0.759	0.500	0.302	0.401	0.180	0.179	0.111	0.510	0.313	0.427	0.389	NAN	NAN

OOM: Out of memory. This term denotes that the methods in the current row cannot deal with big dataset.

OV: OverTime. The algorithm of the current row does not output clustering results after running a day on the dataset of the corresponding column.

NAN: An exception occurred when the algorithm of the current row was run on the dataset for the current column.

Table 12 ACC and NMI with different IMVC approaches

Approach	Dataset

Missing-Rate		BBCSport		Reuters		Reuters-21578		NUSWIDE		Caltech101-7		RGB-D		NUSWIDE30K	
		ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI
0.1	COMPLETER[222]	0.389	0.083	0.268	0.108	0.294	0.049	0.167	0.165	**0.805**	**0.719**	0.345	0.250	0.114	0.070
	DIMVC[159]	0.479	0.165	**0.469**	**0.257**	0.404	0.212	0.136	0.161	0.673	0.588	0.532	0.433	0.113	0.073
	DSIMVC[221]	0.603	0.484	0.430	0.257	0.445	0.281	**0.465**	**0.279**	0.438	0.275	0.314	0.155	**0.445**	**0.276**
	LATER[203]	**0.978**	**0.926**	NAN	NAN	**0.524**	**0.343**	0.164	0.137	0.603	0.689	**0.590**	**0.530**	0.176	0.164
	PIMVC[193]	0.897	0.822	0.385	0.260	0.503	0.295	0.169	0.208	0.522	0.457	0.443	0.348	0.158	0.157
	APADC[216]	0.513	0.336	0.345	0.142	0.396	0.128	0.164	0.167	0.798	0.526	0.563	0.321	OOM	OOM
0.3	COMPLETER[222]	0.359	0.061	0.261	0.105	0.287	0.044	0.186	0.155	0.778	**0.697**	0.346	0.155	0.120	0.071
	DIMVC[159]	0.358	0.061	0.351	0.183	0.431	0.169	0.118	0.130	0.488	0.578	0.425	0.130	.0119	0.082
	DSIMVC[221]	0.598	0.457	**0.441**	**0.269**	0.455	0.289	**0.452**	**0.282**	0.452	0.280	0.321	**0.282**	**0.438**	**0.275**
	LATER[203]	**0.937**	**0.864**	NAN	NAN	0.502	**0.356**	0.168	0.131	0.601	0.688	**0.543**	0.131	0.170	0.165
	PIMVC[193]	0.854	0.801	0.380	0.236	**0.510**	0.291	0.166	0.206	0.514	0.492	0.336	0.206	0.156	0.153
	APADC[216]	0.472	0.321	0.380	0.112	0.421	0.171	0.154	0.162	**0.793**	0.566	0.513	0.162	OOM	OOM
0.5	COMPLETER[222]	0.343	0.054	0.213	0.092	0.265	0.011	0.179	0.152	0.593	0.604	0.360	0.176	0.123	0.067
	DIMVC[159]	0.344	0.054	0.309	0.132	0.393	0.145	0.122	0.146	0.640	0.616	0.339	0.299	0.120	0.104
	DSIMVC[221]	0.553	0.441	**0.462**	**0.280**	0.442	0.262	**0.463**	**0.290**	0.449	0.273	0.322	0.169	**0.450**	**0.274**
	LATER[203]	**0.905**	**0.843**	NAN	NAN	**0.526**	**0.351**	0.157	0.134	0.606	**0.690**	**0.521**	**0.445**	0.161	0.161
	PIMVC[193]	0.775	0.795	0.371	0.214	0.506	0.292	.0173	0.205	0.499	0.449	0.349	0.247	0.147	0.151
	APADC[216]	0.463	0.279	0.371	0.110	0.439	0.167	0.123	0.147	**0.840**	0.603	0.472	0.279	OOM	OOM
0.7	COMPLETER[222]	0.356	0.065	0.194	0.083	0.270	0.013	0.171	0.141	0.497	0.374	0.326	0.155	0.127	0.087
	DIMVC[159]	0.337	0.074	0.327	0.115	0.363	0.115	0.117	0.104	0.536	0.285	0.263	0.155	0.117	0.099
	DSIMVC[221]	0.572	0.435	**0.444**	**0.275**	0.453	0.280	**0.437**	**0.269**	0.455	0.278	0.338	0.161	**0.461**	**0.287**
	LATER[203]	**0.886**	**0.801**	NAN	NAN	**0.520**	**0.373**	0.134	0.095	0.690	0.597	**0.436**	**0.348**	0.154	0.156
	PIMVC[193]	0.762	0.615	0.376	0.233	0.496	0.288	0.152	0.201	0.526	0.488	0.307	0.207	0.156	0.151
	APADC[216]	0.391	0.169	0.369	0.109	0.420	0.140	0.113	0.121	**0.788**	**0.696**	0.391	0.169	OOM	OOM
0.9	COMPLETER[222]	0.337	0.045	0.189	0.056	0.324	0.072	0.134	0.120	0.533	0.288	0.247	0.107	0.120	0.069
	DIMVC[159]	0.325	0.049	0.221	0.034	0.345	0.102	0.093	0.074	0.453	0.201	0.278	0.127	0.119	0.095
	DSIMVC[221]	0.563	0.449	**0.460**	**0.284**	0.464	0.277	**0.451**	**0.278**	0.446	0.269	0.328	0.159	**0.452**	**0.269**
	LATER[203]	**0.782**	**0.641**	NAN	NAN	**0.502**	0.287	0.132	0.091	0.606	**0.696**	**0.361**	**0.296**	0.151	0.152
	PIMVC[193]	0.667	0.547	0.383	0.226	**0.502**	**0.289**	0.155	0.193	0.582	0.481	0.250	0.168	0.149	0.148
	APADC[216]	0.321	0.091	0.371	0.094	0.383	0.110	0.089	0.084	**0.721**	0.483	0.321	0.091	OOM	OOM

Missing-rate: Missing rate of samples