Multi-view Clustering: A Survey

Yan Yang
the School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, China.

Hao Wang
the School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, China.

Recommended Citation

Multi-view Clustering: A Survey

Yan Yang* and Hao Wang

Abstract: In the big data era, the data are generated from different sources or observed from different views. These data are referred to as multi-view data. Unleashing the power of knowledge in multi-view data is very important in big data mining and analysis. This calls for advanced techniques that consider the diversity of different views, while fusing these data. Multi-view Clustering (MvC) has attracted increasing attention in recent years by aiming to exploit complementary and consensus information across multiple views. This paper summarizes a large number of multi-view clustering algorithms, provides a taxonomy according to the mechanisms and principles involved, and classifies these algorithms into five categories, namely, co-training style algorithms, multi-kernel learning, multi-view graph clustering, multi-view subspace clustering, and multi-task multi-view clustering. Therein, multi-view graph clustering is further categorized as graph-based, network-based, and spectral-based methods. Multi-view subspace clustering is further divided into subspace learning-based, and non-negative matrix factorization-based methods. This paper does not only introduce the mechanisms for each category of methods, but also gives a few examples for how these techniques are used. In addition, it lists some publically available multi-view datasets. Overall, this paper serves as an introductory text and survey for multi-view clustering.

Key words: multi-view clustering; co-training; multi-kernel learning; graph clustering; subspace clustering; subspace learning; non-negative matrix factorization; multi-task learning

1 Introduction

In many real-world applications of big data mining and analysis, data are collected from different sources in diverse domains or obtained from various feature collectors. For instance, pictures shared on websites often have corresponding textual tags and descriptions; specific news are reported by multiple news organizations; sensor signals decompose in the time and frequency domains; the same semantic meaning (e.g., hello) is represented with multilingual forms; an image is described by different types of features. All these are referred to as *multi-view data*. These data exhibit heterogeneous properties but hold potential connection. In other words, in such data, each individual view has its specific property for a particular knowledge discovery task; however, different views often contain complementary information that should be exploited. Therefore, a way to exploit this information, in order to uncover the potential values of multi-view data, is very significant in big data research. Applications in real-life data analysis also calls for advanced technologies that can deal with data objects with multiple views, in order to bring data mining and knowledge discovery to new heights.

In the past decade, plenty of machine learning technologies have been investigated for the scope of dealing with multi-view data. Good surveys on

• Yan Yang and Hao Wang are with the School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, China. E-mail: yyang@swjtu.edu.cn; cshaowang@gmail.com.
• Hao Wang is currently a PhD candidate and shares first authorship.
* To whom correspondence should be addressed.

multi-view learning have been conducted in Refs. [1–3]. Moreover, Zheng[4] provided an overview on the methodologies for multi-view (cross-domain) data fusion, in which some specific applications were discussed. Existing multi-view learning technologies are roughly divided into supervised learning and unsupervised learning. This paper focuses on one of the unsupervised learning techniques, namely, clustering. Clustering has emerged as a powerful alternative learning tool for exploring the underlying structure of data[5, 6], especially in the era of big data[7]. The basic idea of clustering algorithms is to partition a set of data objects according to some criteria, such that similar objects are grouped into the same cluster, and dissimilar objects are divided into different clusters.

Many advanced clustering algorithms have been investigated in the last few decades. Although these clustering algorithms have been very successful to some extent, most of them are only suitable to single view data. Even concatenating all views into a single view and then adopting state-of-the-art clustering algorithms on this single view may not improve the clustering performance, because such way is not physically meaningful due to each view having its specific statistical property. In comparison, Multi-view Clustering (MvC) performs effectively on multi-view data by considering the diversity and complementarity of different views. Early studies on MvC, such as reinforcement clustering for multi-type interrelated data[8], multi-view version of DBSCAN[9], and two-view version of EM-based and agglomerative algorithms[10], etc., began approximately in 2003. As an advanced clustering paradigm, MvC received increasing attention in recent years. Thus far, four workshops[11–14] and a mini-symposium[15] have been held in conjunction to related international conferences. In the context of MvC, an inherent problem (and also the goal) of all algorithms having to be dealt with elaborately, is to find a way to maximize clustering quality within each view, while taking clustering consistency across different views into consideration. Moreover, incomplete multi-view data, where some data objects could be missing their observation on one view (i.e., missing objects) or could be available only for their partial features on that view (i.e., missing feature), also pose challenges to MvC.

In this paper, we review a number of representative MvC methods. According to the mechanisms and principles on which these methods are based, we organize and summarize them in five categories:

- *Co-training style algorithms*: This category of methods treats multi-view data by using co-training strategy. It bootstraps the clustering of different views by using the prior or learning knowledge from one another. By iteratively carrying out this strategy, the clustering results of all views tend to each other and this leads to the broadest consensus across all views.
- *Multi-kernel learning*: This category of methods uses predefined kernels corresponding to different views, and then combines these kernels either linearly or non-linearly in order to improve clustering performance.
- *Multi-view graph clustering*: This category of methods seeks to find a fusion graph (or network) across all views and then uses graph-cut algorithms or other technologies (e.g., spectral clustering) on the fusion graph in order to produce the clustering result.
- *Multi-view subspace clustering*: This category learns a unified feature representation (to be input into a model for clustering) from all the feature subspaces of all views by assuming that all views share this representation. Typical models include subspace learning and Non-negative Matrix Factorization (NMF).
- *Multi-task multi-view clustering*: This category treads each view with one task or multiple related tasks, transfers the inter-task knowledge to one another, and exploits multi-task and multi-view relationships in order to improve clustering performance.

We will provide a specific introduction and a few examples for each category in the following sections. Moreover, we also list some widely used multi-view datasets in order to help researchers in this field.

To do that, the rest of this paper will be organized as follows. In Section 2, we illustrate two related principles that ensure the success of MvC. In Section 3, we provide an overview of earlier and more recent MvC methods with five categories, and enumerate a few examples for each category. Some publicly available datasets are covered in Section 4. Finally, in Section 5, we conclude this paper and discuss challenges and future trends for MvC.

Notations and Definitions: We begin with a description of the notations used in this paper. We state that matrices and vectors throughout this paper are

written in uppercase and lowercase letters, respectively. The common notations and corresponding definitions are summarized in Table 1.

2 Principles of MvC

This section deals with analyzing two significant principles of MvC, namely, *complementary* and *consensus* principles. These two principles partially answer why MvC is effective, what the underlying assumptions are, and above all how the MvC should be modeled and performed.

By referring to Ref. [16], we give an illustration on these two principles. Given a data object with two views, this data object is mapped into a latent data space as shown in Fig. 1. From Fig. 1, we can observe that: (1) some ingredients (part A and part C) exist in the individual view, such as part A in view 1 and part C in view 2, i.e., the complementarity of two views, and (2) some ingredients (part B) of the object are shared by both views, i.e., the consensus between two views.

Next, we analyze these two principles as follows:

Complementary principle: This principle states that multiple views should be employed in order to describe data objects more comprehensively and accurately. In the context of multi-view data, each single view is sufficient for a particular knowledge discovery task. However, different views often contain information complementary to each other. For instance, in the field of image processing, each image is described by different types of features, such as LBP, SIFT, and HOG, where LBP is a powerful texture feature, SIFT is robust to image illumination, noise, and rotation, while HOG is sensitive to marginal information. Therefore, it is necessary to exploit these mutually complementary information underlying multiple views in order to describe these data objects, and to provide deeper insights with regard to the internal clustering.

Consensus principle: This principle aims to maximize consistency across multiple distinct views. Based on probably approximately correct analysis, Dasgupta et al.[17] proposed a generalization error analysis for the consensus principle. Given a multi-view dataset X, this dataset has two views X^1 and X^2. Under some mild assumptions, Dasgupta et al.[17] demonstrated the connection between the consensus of two hypotheses on two views, respectively. The connection is formulated as the following inequality:

$$P(f^1 \neq f^2) \geqslant \max\{P_{\text{err}}(f^1), P_{\text{err}}(f^2)\} \quad (1)$$

From this inequality, we conclude that the error probability of the disagreement on two independent hypotheses is an upper bound on the error rate of either hypothesis. Thus, minimizing the disagreement of the two hypotheses will lead to the error rate of each minimized hypothesis; namely, maximizing the agreement (or consistence) of the two hypotheses will result in the error rate of each minimized hypothesis. This is called *maximizing-consistence policy*. Co-training[18] is a landmark technology, and is one of the most widely used schemes for multi-view learning. The standard co-training algorithm is trained alternately in order to maximize the mutual agreement of the two unlabeled data views by using learning or by providing labeled data to one another. In terms of clustering, De Sa[19] pioneered a two-view spectral clustering algorithm, which was inspired by the idea of minimizing-disagreement (the same concept as maximizing-consistency). There are also many co-training style MvC algorithms (see Section 3.1).

To sum up, both complementary and consensus principles play important roles in addressing the problem of MvC, and both of them should be considered in order to take full advantage of multi-view

Table 1   Notations and definitions.

Notation	Definition
k	User-specified number of clusters
m	Number of views
n	Total number of instances
n_v	Number of instances in the v-th view data
d_v	Dimensionality of the v-th view data
X	$X = \{X^1, ..., X^m\}$, data matrices with m views
X^v	Data matrix in the v-th view
x_j^v	j-th column vector of X^v
x_{ij}^v	(i, j)-th entry of X^v
I	Indentity matrix

Note: If multi-view data is complete, then $n = n_v$.
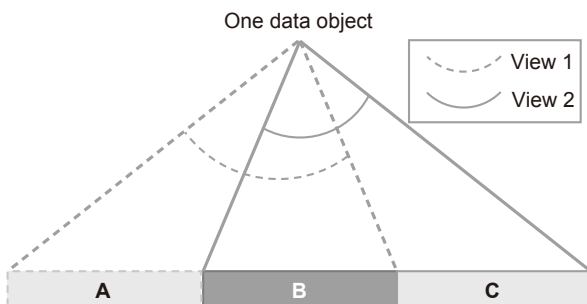

Fig. 1 Illustration of complementary and consensus principles[16].

data.

3 Multi-view Clustering

3.1 Co-training style algorithms

Co-training style algorithms are investigated under the consideration of multiple views consensus. This category of methods intends to maximize the mutual agreement across all views and arrive at their broadest consensus. The general procedure of the conventional co-training algorithms is shown in Fig. 2. According to the procedure, the algorithm is trained alternately in order to maximize the consistency of the two distinct views by using prior information or by learning knowledge from each other. Note that the success of co-training mainly relies on three assumptions: (1) Sufficiency: Each view is sufficient for the learning task on its own, (2) Compatibility: The objective functions export the same predictions for co-occurring features with high probability in both views, and (3) Conditional independence: All views provide learning labels in conditional independence. While, in practice, it is usually too hard to satisfy the conditional independence assumption. Thus, several weaker assumptions such as the weak conditional dependence assumption[20], much weaker "expansion" assumption[21], and the difference assumption[22] have been investigated. In addition, several extended versions of co-training such as co-EM[23], co-regularization[24], and co-clustering[25], have been studied.

Most of the above-mentioned methods are designed for multi-view data in a semi-supervised learning setting. In unsupervised learning (i.e., clustering), Bickel and Scheffer[10] first studied MvC with a co-training idea, and proposed two kinds of MvC algorithms for text data. One is a multi-view EM algorithm that works alternately between the views, while the other is an agglomerative algorithm, inspired by the co-training algorithm. As a result, Bickel and Scheffer[10] concluded that the multi-view EM algorithm significantly outperformed the single-view

Fig. 2   General procedure of co-training.

algorithm; however, the agglomerative algorithm led to negative results. Furthermore, they studied the estimation of mixture models with co-EM for multi-view data analysis[26], which contributes to adopting a mixture model estimation for multiple views by demonstrating that the co-EM algorithm is a special case of mixture model estimation. Moreover, Tzortzis and Likas[27] put forward a weighted multi-view convex mixture model that automatically assigns weights to views via EM. Assuming that the similar data objects are grouped into the same cluster, regardless of views, Kumar and Daumé III[28] proposed a co-training approach for multi-view spectral clustering, where the clusters are bootstrap to different views by using complementary information from one another. Kumar et al.[29] further proposed a co-regularized approach for multi-view spectral clustering, where the graph Laplacians are imposed on all views and the regularizations on the eigenvectors of the Laplacians in order to consistently lead the resulting clustering structures. Inspired by the work of Kumar et al.[29], Ye et al.[30] discussed co-regularized kernel K-means for MvC. This method automatically learns the weights of different view from data. In addition, a multiple view aware method with a co-training strategy was investigated in order to cluster process executions (traces)[31], in which it considers that the traces of an event log is described by multiple trace profiles, and adapts an iterative co-training strategy to the process mining setting. A co-regularized Probabilistic Latent Semantic Analysis (PLSA) model for MvC was developed in Ref. [32]. The central idea behind it is that the sample similarities in the topic space from one view should agree to another view. To address the challenge of partial mapping between the views (i.e., incomplete views), multi-view constrained clustering via co-EM with pairwise constraints propagation were studied in Refs. [33, 34]. In other words, the proposed methods in Refs. [33, 34] used co-EM in order to iteratively estimate the propagations within each view, transfer the given pairwise constraints across views, update the clustering model, and finally learn a unified clustering result for all views.

MvC based on co-clustering (simultaneously clustering the objects and features) has also been investigated. For instance, Meng et al.[35] proposed a heterogeneous data co-clustering approach, which does not only extend fusion from two views to multiple views, but also weight the features of multiple data sources. Based on matrix decomposition,
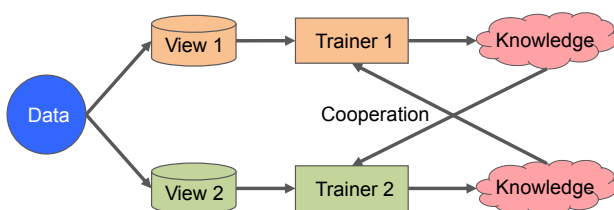
Sun et al.[36] presented a proximal alternating linearized minimization algorithm. This algorithm can simultaneously decompose multiple data matrices into sparse row and column vectors, and link different views of data with a binary vector, where the binary vector enforces consistency for the row clusters from all views. Simultaneously building similarity matrices, rather than a set of clusters, between the rows and columns of a data matrix, an architecture to learn co-similarities from multi-view datasets was designed in Ref. [37], and was subsequently parallelized in Ref. [38]. Assuming that transferring similarity values (generated from individual data) from one view to others would result in better data clustering, Hussain and Bashir[39] extended the co-similarity based architecture in order to handle multiple datasets with two kinds of integration schemes (i.e., intermediate integration and late integration). In addition, several collaborative MvC approaches have been investigated in Refs. [40, 41]. These approaches consist of two phases: the local phase and collaboration phase, where the local phase applies a clustering algorithm to each view, and the collaboration phase collaborates each view with the clustering results associated to the other views produced from the local phase.

Example 1: As illustrated in Fig. 3 (using two views for brevity), Kumar and Daumé III[28] first applied the co-training strategy to the problem of multi-view spectral clustering. Unlike semi-supervised learning, there are no labeled data in unsupervised learning settings; therefore, the prototypical co-training algorithms were not available directly for MvC. However, the motivation of co-training still remains the same as in unsupervised learning problems. In other words, it limits the search only to hypotheses (clusterings) that agree with those in other views. Assuming that the true underlying clustering would

---

1: Calculate the graph similarity matrix S^1 and S^2 for both views.

2: Initialize the graph Laplacian matrices L^1 and L^2, and the discriminative eigenvectors U^1 and U^2.

3: Perform spectral embedding on S^1 with U^2 to get new similarity matrix S^1.

4: Perform spectral embedding on S^2 with U^1 to get new similarity matrix S^2.

5: Compute the new Laplacians matrices L^1 and L^2, new eigenvectors U^1 and U^2.

6: Go to Step 3 and repeat for a number of iterations.

Fig. 3 Co-training approach for multi-view spectral clustering[28].

---

assign a point to the same cluster, irrespective of the views, as was done in most of those co-training based MVC approaches, Kumar and Daumé III[28] took the spectral embedding from one view in order to constrain the similarity graph of the other view. By carrying out this process iteratively, the clusterings of the two views tended to each other.

3.2 Multi-kernel learning

Multi-kernel learning was originally developed in order to boost the search space capacity of possible kernel functions, e.g., Linear kernel, Polynomial kernel, and Gaussian kernels, in order to achieve good generalization. As kernels in multi-kernel learning naturally correspond to different views, multi-kernel learning has been widely applied in order to deal with multi-view data. The general procedure of multi-kernel learning approaches is shown in Fig. 4, where different predefined kernels are used to deal with different views. Then these kernels are combined either linearly or non-linearly in order to arrive at a unified kernel. In an MvC setting, multi-kernel learning based MvC intends to optimally combine a group of predefined kernels in order to improve clustering performance. In such methods, an essential problem consists of finding a way to choose suitable kernel functions and combine these kernels optimally.

In a single view scenario, based on maximum margin clustering[42], Zhao et al.[43] presented a multiple kernel clustering algorithm, which can simultaneously find the maximum margin hyperplane, best clusterings, and the optimal kernels. Du et al.[44] performed a robust K-means (with $l_{2,1}$-norm) on kernel space, and proposed a multiple kernel K-means algorithm, which is able to simultaneously find the best clustering labels, the cluster membership, and the optimal combination of multiple kernels. It is worth stressing that this type of the above mentioned algorithms is available for dealing with multi-view data under the framework shown in Fig. 4. In a multi-view scenario, De Sa et al.[45] constructed a custom kernel combination method based on the minimizing-disagreement algorithm[46, 47]. Specifically,
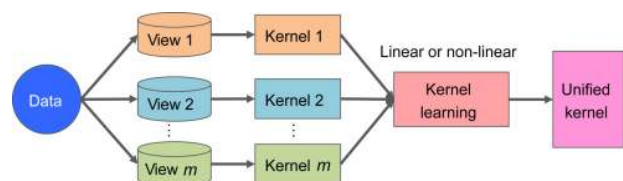

Fig. 4 General procedure of multi-kernel learning.

they generated a multi-partite graph in order to induce a kernel, which was then used for spectral clustering. In fact, this method could be regarded as a variant of kernel canonical correlation analysis, and a generalization of co-clustering and spectral clustering. Moreover, Yu et al.[48] extended the classical K-means clustering into Hilbert space, where multi-view data matrices were denoted as kernel matrices and were then combined automatically for data fusion. Similar work has also been done in Ref. [49]. The difference with Ref. [48] is that the kernels were combined in a localized way in order to better capture the sample characteristics of the data. Rather than extending the existing clustering algorithms with a multi-kernel learning setting, Lu et al.[50] studied multiple kernel clustering based on a centered kernel alignment (an effective kernel evaluation measure), which was employed in order to unify two clustering tasks and multi-kernel learning into a single optimization framework.

Methods with the weighted combination of kernels have also been studied by considering the difference of views (or kernels). For instance, kernel based weighted MvC was investigated in Ref. [51], where the weights to the kernels were assigned according to the information quality of the corresponding views. A systemic MvC approach was proposed in order to automatically assign weights for deriving the kernel matrix on each view through an optimization process in Ref. [52], where the kernel matrix learning was based on the kernel alignment in order to measure the similarity between two kernel matrices. In addition, Liu et al.[53] showed a weighted multiple kernel K-means clustering method with matrix-induced regularization, which could reduce the redundant kernels and enhance the diversity of the predefined kernels. Zhao et al.[54] provided a weighted MvC method with matrix-induced and low-rank regularization. Zhang et al.[55] also presented a weighted MvC algorithm based on improved Gaussian kernels with variable weights.

However, in many applications, it is common that data on some views is not available, or is only partially available, which leads to incomplete multi-view data, as we mentioned in the introduction. To address this issue, Trivedi et al.[56] presented a general approach that allows the MvC, in complete view settings, to be applicable in this scenario, where only one view was complete and the auxiliary views were incomplete. They took the kernel CCA based MvC as an example in order to illustrate their idea. De Sa et al.[45] (mentioned

above) also stated that their proposed algorithm could calculate sample affinities with missing views. In the setting where no view is complete, Shao et al.[57] proposed a collective kernel learning algorithm in order to infer the hidden sample similarity. The idea behind this approach was to collectively complete the kernel matrices of the incomplete views by optimizing the alignment of the shared instances of those views. In addition, unlike some existing methods, where incomplete kernels were first imputed and then an available multi-kernel clustering algorithm was applied to the inputting kernels, Liu et al.[58] integrated the kernel imputation and clustering into a unified learning procedure for incomplete MvC.

Example 2: One challenge of multi-kernel learning consists of choosing appropriate kernel functions (e.g., Linear kernel, Polynomial kernel, and Gaussian kernel), which map the original low-dimensional space to a high-dimensional space. The general method for multi-view data is to use a linear combination of several kernel functions, while the weights of different kernels should be taken into consideration. Moreover, the weights of different views are also an important factor for MvC. To these ends, Zhang et al.[55] developed an auto-weighted multi-kernel MvC algorithm that weights the views and kernels simultaneously. Figure 5 gives an illustration for the proposed algorithm. First, it employs Kernel Principal Component Analysis (KPCA) on each view in order to reduce the dimension of the original data, and results in low-dimensional multi-view data. Then, it applies the designed weighted Gaussian kernel on the low-dimensional multi-view data. This step drives the weight of each view and cluster centers. After finite iterations, it arrives at the final clustering result. It is worth noting that the designed weighted Gaussian kernel integrates the advantages of Gaussian kernel and Polynomial kernel. The designed weighted Gaussian
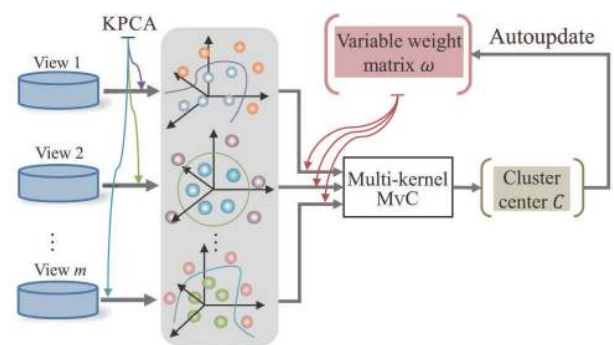

Fig. 5   The flow chart of auto-weighted multi-kernel MvC.

kernel[55] is formulated as

$$K(x, y) = \left[\exp\left(-\frac{\|x - y\|^2}{2\sigma^2} \right) + R \right]^p =$$

$$\sum_{q=0}^{p} \binom{p}{q} R^{p-q} \exp\left(-\frac{q\|x - y\|^2}{2\sigma^2} \right) =$$

$$R^p + \sum_{q=0}^{p} \binom{p}{q} R^{p-q} \exp\left(-\frac{\|x - y\|^2}{2\sigma^2/q} \right)$$

(2)

Given multi-view data with m views, n samples, and k clusters, the objective function[55] based on the K-means and designed kernel is formulated as

$$\min \sum_{v=1}^{m} \sum_{i=1}^{n} \omega_{i,v} \sum_{j=1}^{k} \delta_{ij} \|\phi^v(x_i^v) - \phi^v(c_j^v)\|^2,$$

$$\text{s.t., } \omega_{i,v} \geqslant 0, \prod_v \omega_{i,v} = 1 \quad (3)$$

where c_j^v is the cluster center, and δ_{ij} is the indicator variable with $\delta_{ij} = 1$ if $x_i \in c_j$, otherwise $\delta_{ij} = 0$.

By plugging the designed Gaussian kernel $K(x, y) = \phi(x) \cdot \phi(y)$ into Formula (3)[55], it is rewritten as

$$\min \sum_{v=1}^{m} \sum_{i=1}^{n} \omega_{i,v} \sum_{j=1}^{k} 2\delta_{ij} \left[(1 + R)^p - K(x_i^v, c_j^v) \right],$$

$$\text{s.t., } \omega_{i,v} \geqslant 0, \ \prod_v \omega_{i,v} = 1 \quad (4)$$

Note that Formula (4) inherits the properties of K-means and kernel, and the designed kernel integrates the advantages of Gaussian kernel and Polynomial kernel.

3.3 Multi-view graph clustering

Graphs (or networks) are widely used for representing the relationships between objects, where each node corresponds to a data object and each edge depicts the relationship between a pair of objects. In practice, the relationship is often denoted by the similarity or the affinity relationship; namely, the input graph matrix is generated from a data similarity matrix. In a multi-view scenario, data objects are captured by multiple graphs. A common assumption is that each individual graph can capture the partial information of the data; while, all graphs have the same underlying clustering structure of data. Thus, these graphs are able to mutually reinforce each other by consolidating the correlation between the data objects collectively. In general, the graph-based fusion procedure for multi-view data is similar to Fig. 6. Multi-view graph clustering aims to find a fusion graph across all views and then uses graph-cut algorithms or other technologies (e.g., spectral clustering) on the
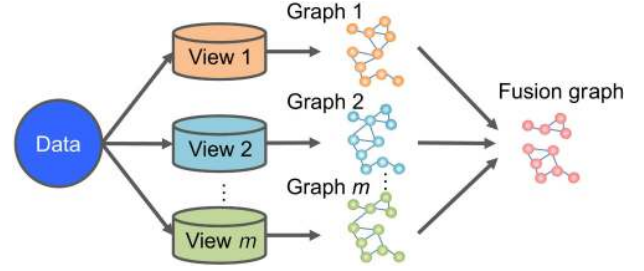

Fig. 6 General procedure of graph-based clustering.

fusion graph in order to produce the final clustering result.

In this category, the literature review is organized in three parts, namely, graph-based MvC, network-based MvC, and spectral-based MvC.

3.3.1 Graph-based MvC

Based on multiple similarity graphs, Tang et al.[59] discussed a general clustering problem. The proposed linked matrix factorization method extracted common factors from multiple graphs, which led to various graph-based clustering methods that could be naturally applied to multi-view data. Hussain et al.[60] proposed a multi-view document clustering algorithm, which first applies single-view clustering algorithms to the data matrix of each view, in order to generate multiple partitions. Then, it uses these partitions to generate a set of three different similarity matrices, namely, the affinity matrix, cluster based similarity matrix, and pair-wise dissimilarity matrix. Finally, it employs an ensemble technique in order to aggregate these matrices, and forms a unified similarity matrix for clustering. Moreover, the impact of different similarity measures (e.g., Pearson, and Spearman correlations, Euclidean, and Canberra Distances, etc.) on MvC has been studied in Ref. [61]. Later on, Xue et al.[62] proposed a group-aware multi-view fusion method, which adopts different weights to characterize the pairwise similarity between different groups. Furthermore, even though some MvC methods learned a weight for each graph, such methods have additional parameters. In order to address these challenges, Nie et al.[63] developed a parameter-free multiple graph framework to learn a set of weights automatically for all graphs. In addition, unsupervised feature selection for multi-view data has also been investigated. There, these selected features were used for a clustering task or other learning tasks. For instance, Wei et al.[64] proposed a method, named cross diffused matrix alignment based on feature selection, in order to select

features for each view by performing alignment on a cross diffused matrix. Then they applied co-regularized spectral clustering[29] on these selected features in order to produce the final results. Moreover, since traditional approaches such as Ref. [63] evaluate the similarity by a predefined or fixed graph Laplacian in each view, separately, and neglect the underlying common structures across different views, Hou et al.[65] presented a multi-view unsupervised feature selection algorithm with adaptive similarities and view weights. In total, this feature selection method employs three types of data information, i.e., data similarity, data clustering structure, and the correlation between different views.

On the other hand, methods with a combination of the nearest neighbor techniques have also been investigated. For instance, Hamzaoui et al.[66] designed a multi-source Shared Nearest Neighbors (SNN) scheme for multi-modal image clustering. The central idea was to extend the existing SNN-based similarity measures to the case of multiple sources, and then introduce an original automatic source selection step in order to build candidate clusterings. With consideration of the generative and manifold data structure, Wang et al.[67, 68] developed a generative model with an ensemble manifold regularization for MvC. Specifically, they constructed a nearest neighbor graph for each view in order to encode the corresponding manifold information, and a multiple graph ensemble regularization framework was designed in order to learn the optimal intrinsic manifold. Then, the manifold regularization term was incorporated into a multi-view topic model based on PLSA, which resulted in a unified objective function. Unlike the above two methods, Zhang and Mao[69] focused on the task of efficiently selecting clustering consistent neighbors for MvC. The proposed approach used jointly sparse

weights in order to filter unreliable neighbors in the union of view-specific neighborhoods by representing each object in a weighted sum of its neighbors under each view. The learning sparse weights were employed in order to generate a similarity graph, and this graph was further utilized for MvC. In addition, Nie et al.[70] introduced a novel multi-view learning model with adaptive neighbors. This model performs semi-supervised classification and local manifold learning and clustering, simultaneously. It modifies the similarity matrix during each iteration until it arrives at the optimal one. Moreover, it automatically allocates the weight coefficient for each view without penalty parameters.

Example 3: It should be noted that most of the existing methods adopt a globally uniform similarity measure over the entire data space. However, in real-world objects such as images, different images have different visual appearance and the visual distribution is also complex. It is difficult to capture the similarity of different objects accurately only by using a globally uniform measure. To solve this problem, Xue et al.[62] presented a group-aware multi-view fusion approach for image clustering. This approach can partition images into different groups with more compact visual cohesiveness, and assign diverse fusion weights for images between and within groups. In comparison to the global fusion methods, this group-aware fusion model provides a more flexible fusion strategy and more effective similarity measures among images. The framework of this method is shown in Fig. 7. Concretely, multiple features such as LBP, GIST, and Centrist, were first extracted from images, which constituted three different features (views). Then, a graph was constructed for each view. Next, all images were divided into different groups, and a fused graph was constructed with the proposed fusion strategy.
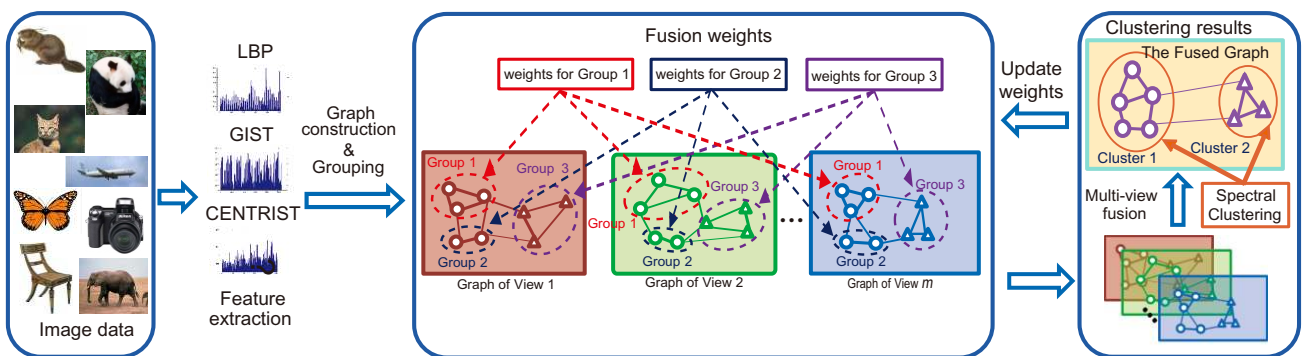

Fig. 7   The framework of group-aware multi-view fusion approach[62].

The intent was to assign different fusion weights for images belonging to different groups and the same fusion weights for the images within the same group. Finally, the fusion weights were learned by solving the formulated objective function, and the clustering findings are obtained by performing spectral clustering on this fused graph.

3.3.2 Network-based MvC

Despite previous successes, most graph-based MvC approaches usually assume that the same set of data objects are available for different views. Thus, the relationship between data objects in different views is one-to-one relationship. However, in many real-life applications, such as social networks, literature citation networks, and biology interaction networks, data are collected from different domains and an object in one domain may correspond to multiple objects in another domain, which results in many-to-many mapping relationships. Representing these relationships with networks rather than with graphs may be more appropriate. This is the main reason for distinguishing network-based MvC from graph-based MvC.

Related work on network-based MvC starts from Ref. [71], in which a network based multi-view graph clustering framework is developed, and termed co-regularized graph clustering. This framework illustrates several key properties, namely, many-to-many mapping relationships, mapping associated with weights, and partial mapping among different networks. However, different networks may have different data distributions, leading to assumptions such as the one in Ref. [71], by which all networks admit that a common clustering structure no longer holds. To relax this assumption, Ni et al.[72] presented a robust and flexible framework, which allows multiple underlying clustering structures across different networks. It treats domain similarity as the main network, and formulates the clustering problem via NMF on the designed network of network settings. Similar work has also been presented in Ref. [73], where the network grouping and underlying clustering detection are coupled and mutually enhanced during the learning process. Furthermore, Liu et al.[74] stated that existing network-based methods tend to focus on the network clustering task itself, but ignore any associations that may be exhibited between clustering findings from different domains. Given this, they offered a robust clustering approach that

can detect network clusters in multiple domains, and their cross-domain associations. In addition, Yu and Zhang[75] studied community detection in multiple social networks, and attempted to find the communities for multiple net-works involving both anchor and non-anchor users simultaneously. Wang et al.[76] proposed an MvC algorithm, named multi-view affinity propagation, based on max-product belief propagation. The key point was to establish an MvC model consisting of two components that measure the within-view quality and the explicit clustering consistency across different views, respectively.

Example 4: Most multi-view graph clustering approaches usually assume that data objects in different views have a strict one-to-one mapping relationship, while, in many actual applications, data are collected from diverse domains (views), where the cross-domain mapping relationship is many-to-many rather than one-to-one. In other words, an object in one domain may correspond to multiple objects in another domain. Moreover, different domains have their inherent data distributions. This breaks another assumption by which all views share a common clustering structure. To address the above mentioned two challenges, Ni et al.[72] developed a robust and flexible multi-network clustering framework that allows many-to-many relationship and multiple underlying clustering structures.

Specially, Ni et al.[72] modeled each domain similarity as a network, and also modeled the similarity among different domains as a network to regularize the clustering structures in different networks. They defined a global network, named *Network of Networks* (NoN) as shown in Fig. 8, where the dashed network represents the *main network* among six domains {A,
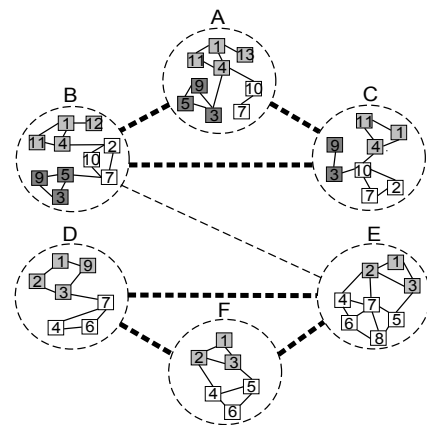

Fig. 8   Network of networks[72].

B, C, D, E, F}, and each node in the main network corresponds to a *domain-specific network* denoted by the solid lines. Correspondingly, the clusterings in the main network and domain-specific networks are referred to as *main clusterings* and *domain clusterings*, respectively. Given these concepts, they sought to partition the NoN by using a two-phase approach. By this approach, first, the main network is partitioned, and then the learning information is incorporated from the main network in order to cluster domain-specific networks. The objective function[72] is formulated as follows:

$$O_M = \|G - HH^{\mathrm{T}}\|_{\mathrm{F}}^2 \qquad (5)$$

where $\| \cdot \|_{\mathrm{F}}^2$ is the *Frobenius norm*, $G \in \mathbb{R}^{g \times g}$ (g is the number of nodes in the main network) is the main network, and $H \in \mathbb{R}_+^{g \times k}$ is the factor matrix of the main network.

The domain-specific network clustering[72] is formulated as follows:

$$\min_{\substack{\forall i, U^i \geqslant 0 \\ \forall j, V^j \geqslant 0}} O_S = \underbrace{\sum_{i=1}^{g} \|A^i - U^i (U^i)^{\mathrm{T}}\|_{\mathrm{F}}^2}_{\text{Domain-specific network clustering}} +$$

$$\alpha \underbrace{\sum_{i=1}^{g} \sum_{j=1}^{k} h_{ij} \|U^i - V^j\|_{\mathrm{F}}^2}_{\text{Main cluster guided regularization}} \qquad (6)$$

where A^i is the domain-specific network corresponding to node i ($i = 1, ..., g$) in the main network G, U^i is the factor matrix of A^i, V^j is the j-th hidden factor matrix, h_{ij} (an element of H) indicates to which degree the main node i belongs to the j-th main clustering, and α is a regularization parameter.

Furthermore, based on the above proposed model, they designed a general model that allowed partially aligned domain-specific networks to have different node sizes and a different number of clusterings. This model[72] is formulated as the following function:

$$\min_{\substack{\forall i, U^i \geqslant 0 \\ \forall j, V^j \geqslant 0}} O_S = \sum_{i=1}^{g} J_A + \alpha \sum_{i=1}^{g} \sum_{j=1}^{k} J_R \qquad (7)$$

where

$J_A = \|A^i - U^i (U^i)^T\|_{\mathrm{F}}^2$,

$J_R = \|(Q^{ij} U^i)(Q^{ij} U^i)^{\mathrm{T}} - (P^{ij} V^j)(P^{ij} V^j)^{\mathrm{T}}\|_{\mathrm{F}}^2$,

where P^{ij} is the mapping matrix between U^i and V^j, and Q^{ij} is denoted by $Q^{ij} = P^{ij}(P^{ij})^{\mathrm{T}}$. In the end, the clustering results are driven on the factor matrices $U^i, ..., U^g$.

3.3.3 Spectral-based MvC

Spectral clustering is a classic data clustering paradigm. The basic idea is to form a pairwise affinity matrix between any pairs of objects, normalize this affinity matrix, and compute eigenvectors of this normalized affinity matrix (i.e., graph Laplacian). It has been shown that the second eigenvector of the normalized graph Laplacian is a relaxation of a binary vector solution. This solution can minimize the normalized cut on a graph, which is the relationship between the spectral and graph. In Ref. [19], De Sa developed a spectral clustering algorithm on two independent views, each of which could be fed into a clustering model. This spectral-based MvC algorithm created a bipartite graph with a minimizing-disagreement criterion[46, 47] in order to connect the two-view features and then perform available spectral clustering algorithms on this bipartite graph. Zhou and Burges[77] investigated multi-view spectral clustering by generalizing a normalized cut from the single view to multiple views, by considering how to learn a clustering close to the optimal solution for all graphs, and to further develop a multi-view transductive inference on the basis of multi-view spectral clustering. Similar work has been carried out in Ref. [78], where it also intended to find a balance cut that could separate all similarity graphs well. In addition, Long et al.[79] developed a general model for multi-view unsupervised learning under a distributed framework, with the aim of detecting hidden patterns individually from each representation of multiple views, and to seek optimal hidden patterns from these finding patterns. The authors put forward the concept of mapping function in order to make the patterns from various pattern spaces comparable in this general model. Hence, an optimal pattern was achieved from these various patterns of multiple representations. Instead of committing to one clustering solution, Niu et al.[80] proposed a method that can provide several non-redundant clustering solutions. This method learns non-redundant subspaces for multiple views, and produces a clustering solution simultaneously for each view. To address the issue that data may have considerable noise, Xia et al.[81] investigated the Markov chain in order to formulate a multi-view spectral clustering model. This model has the flavor of low-rank and sparse decomposition. It first draws a transition probability matrix from each single view, and then uses these matrices in order to form a shared low-

rank transition probability matrix. Finally, this shared matrix is input to a standard Markov chain model for clustering. To handle a large-scale problem and improve computational efficiency, Li et al.[82] offered a multi-view spectral clustering algorithm for large-scale multi-view data. This algorithm uses a local manifold fusion in order to fuse heterogeneous features, and bipartite graphs so as to approximate the similarity graphs. Moreover, Chikhi[83] presented a multi-view normalized cuts approach, a parameter free multi-view spectral clustering algorithm, based on spectral partitioning and local refinement. Lu et al.[84] studied convex sparse spectral clustering with sparse regularization for single view data, and proposed a pairwise sparse spectral clustering for handling multi-view data. However, as the number of views grows, it is scarcely possible to avoid dependencies among views, and these dependencies often delude correct predictions. To address these issues, Son et al.[85] extended traditional spectral clustering in order to deal with the dependencies among views. Especially, they designed a brainstorming process in order to force the information of each view to be shared among them.

Several MvC methods have also been studied by combining spectral clustering and other technologies. For instance, Huang et al.[86] developed an affinity aggregation spectral clustering algorithm by extending spectral clustering to a setting with multiple available affinities. Shao et al.[87] designed a multi-source MvC framework based on collective spectral clustering with a discrepancy penalty across sources. Note that this method is applicable to incomplete multi-view data. Moreover, Feng et al.[88] introduced a multi-view spectral clustering via roust local subspace learning by considering that all views are noisy and derived from a robust unified subspace and noisy. Wang et al.[89] proposed an iterative low-rank based structured optimization method for multi-view spectral clustering, which encodes the local manifold structure of the data from each view-dependent feature space, and arrives at a multi-view agreement based on an iterative process. Zhao et al.[90] discussed the semi-supervised MvC, and presented a multi-view matrix completion method with a pairwise similarity matrix in order to utilize side information; namely, must-link and cannot-link.

In addition, there are also spectral-based MvC methods for multi-type relational data[91], multi-modal image data[92], social media data[93], as well as some applications with spectral-based MvC, such as multimodal brain network inferences[94], social circle detection[95], and human microbiome data analysis[96].

Example 5: In general, spectral clustering methods involve two time-consuming steps. The first step, constructing the similarity (or affinity) graph, takes $O(n^2 d)$ time, while the second step, computing the eigen-decomposition, takes $O(k n^2)$ time. Moreover, another drawback of spectral clustering is that most spectral clustering methods usually do not provide a natural extension for dealing with the out-of-sample problem. To overcome the above two issues, Li et al.[82] proposed a multi-view spectral clustering approach for large-scale multi-view data. In summary, the designed algorithm uses local manifold regularization to fuse heterogeneous features, and approximates the similarity graphs with bipartite graphs in order to improve efficiency. It is also easily extended in order to deal with the out-of-sample problem. First, it generates a few consensus salient points for all views. These salient points are employed in order to capture the manifold of the original views. Then a bipartite graph is constructed between the raw data points and the salient points for each view. The graphs of all views are fused together with a local manifold regularization item. Finally, it applies a spectral clustering algorithm on the resulting fused graph and outputs the clustering indicator of the salient points, in order to deal with the out-of-sample problem efficiently.

Here, the two important questions that need to be stressed are how to reach consensus across all views, and how to express their relationship. With local manifold learning, the two questions mentioned above are formulated with the following function[82]:

$$\min_{F^{\mathrm{T}}F=I,\alpha^v} \sum_{i=1}^{m}(\alpha^i)^r \mathrm{tr}(F^{\mathrm{T}}L^i F),$$
$$\text{s.t., } \sum_i \alpha^i = 1, \alpha^i \geqslant 0 \qquad (8)$$

where α^i is the non-negative normalized weight factor for the i-th view, $\mathrm{tr}(\cdot)$ is the trace of a matrix, r is a scalar to control the distribution of different weights among different views, $F \in \mathbb{R}^{n \times k}$ is the class indicator matrix, and L^i is the normalized graph Laplacian matrix for the i-th view. The normalized graph Laplacian matrix[82] is formulated as

$$L = I - D^{-1/2}WD^{-1/2} \qquad (9)$$

where $W \in \mathbb{R}^{n \times n}$ is the adjacent matrix of the graph, $D \in \mathbb{R}^{b \times n}$ is the degree matrix whose i-th diagonal element is $d_{ii} = \sum_{j=1}^{n} w_{ij}$.

Formula (8) aims to provide a consensus result F among all views. This unique consensus eliminates the requirement for computing the local results for each view, and the computational overhead of communicating back and forth between the local results and the global result. To further uncover inter-view relationships, Formula (8)[82] is rewritten as

$$\min_{F^{\mathrm{T}}F=I,\alpha^v} \mathrm{tr}(F^{\mathrm{T}}LF),$$

$$\text{s.t., } \sum_i^m \alpha^i = 1, \alpha^i \geqslant 0 \qquad (10)$$

where $L = \sum_i^m (\alpha^i)^r L^i$. Denote that L is regarded as the local manifold fusion of all views. Formula (10) is solved by iterative optimization techniques. The total computational complexity is approximately $O(Tkn^2 + \sum_i^m mn(d^i)^2)$, where T is the number of iterations.

3.4 Multi-view subspace clustering

Multi-view subspace clustering, i.e., learning a new and unified representation for all view data, from multiple subspaces, or a latent space that makes it easier to deal with high-dimensional data when building clustering models, has become a hot topic in the field of MvC. The general procedure of multi-view subspace clustering is illustrated as Fig. 9; there it obtains such a unified feature representation in two ways: ① learn a unified representation from multiple subspaces directly, or ② first learn a latent space and then arrive at this unified representation. Finally, this unified representation was fed into an off-the-shelf clustering model in order to produce the clustering results. After reviewing the literature on MvC, we divided the multi-view subspace clustering methods into two major types, namely, subspace learning-based and NMF-based (a special case in subspace learning) methods.

3.4.1 Subspace learning-based MvC

Subspace learning-based MvC seeks to find a latent space from multiple low-dimensional subspaces by assuming that data points are drawn from this latent subspace. Here, we extend this concept in order to make its role more general. In this paper,
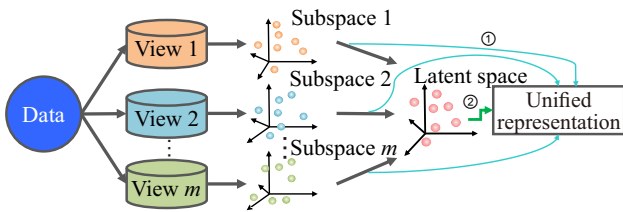

Fig. 9   General procedure of multi-view subspace clustering.

the technologies involved in subspace learning-based MvC include subspace learning, subspace clustering, subspace projection, low-rank approximation, and tensor decomposition.

Assuming that all views are conditionally independent, given the clustering labels, Chaudhuri et al.[97] presented a multi-view subspace learning method based on canonical correlation analysis. This method provides auxiliary results for Gaussian mixtures and log concave distribution mixtures. Guo[98] proposed a convex subspace representation learning method for MvC. The key idea is to detect a shared subspace representation across multiple views, and then adopt standard clustering algorithms on this shared representation. Zhao et al.[99] developed a co-training framework for multi-view subspace clustering. It combined classical K-means and linear discriminant analysis under a co-training scheme, which utilized labels learned automatically in one view in order to generate discriminative subspaces in another. Deng et al.[100] put forward a feature weighting method based on subspace learning, where it locally adapted the feature weighting of each group automatically according to the tightness of views. Moreover, Cao et al.[101] stated that exploiting the specific independently constructed matrices is insufficient for the success of MvC, and exploring the underlying complementarity is of great importance. To this end, they designed a framework, named diversity-induced multi-view subspace clustering framework. Concretely, they extended the existing single-view subspace clustering to the multi-view domain, and utilized the Hilbert Schmidt Independence Criterion as a diversity term in order to explore the complementarity of multi-view representations. However, many studies have usually focused on the combining information rather than on improving the feature representation capability of each view. To solve this problem, Wang et al.[102] presented a framework with an extreme learning machine, and implemented three algorithms on this framework. Unlike the methods in Refs. [98, 102] that perform subspace clustering on a common view, Gao et al.[103] performed subspace clustering on every view, simultaneously, while guaranteeing the clustering consistence among different views by adopting a common indicator. In addition, Xu et al.[104, 105] proposed an MvC method called discriminatingly embedded K-means, which embedded the synchronous learning of multiple

discriminative subspaces into multi-view K-means clustering in order to formulate a unified framework, while controlling the inter coordination between these subspaces adaptively. To effectively exploit data correlation consensus among multi-views, subspace clustering with a similarity matrix for multi-view data was studied in Refs. [106, 107], where the authors intended to find a correlation or similarity consensus among all views, which was inspired by the idea that data objects within the same subspace have large similarity, while having small similarity for data objects within the distinct subspaces for each view. Rather than using a similarity matrix, Fan et al.[108] drew global low-rank constraints and local cross topology preserving constraints into subspace clustering for the purpose of characterizing data correlations. There have also been several methods investigated by combing with other technologies, sparse subspace clustering[109, 110], low-rank approximation[111], and tensor decomposition[112–115], to name a few.

Unlike the above mentioned methods or frameworks, which output just a single clustering, Cui et al.[116, 117] presented pioneering work in order to find alternative and multiple clustering solutions based subspace learning. They designed an MvC framework in order to find all non-redundant data clustering views, and suggested two methods within this framework, i.e., orthogonal clustering, and clustering in orthogonal subspaces. The difference is that the former seeks orthogonality in the cluster space, while the latter does it in the feature space. Similar work has also been carried out in Ref. [118], where multiple generalizations of the data are provided by using multiple mixture models. Each mixture describes a specific view on the data by using a mixture of the Beta distributions in subspace projections. Moreover, Muller et al.[119] presented a short tutorial on this topic.

In a semi-supervised clustering setting, Günnemann et al.[120] developed a new Bayesian framework for semi-supervised MvC based on their previous work[118], which also sought to the detections of multiple and alternative clusterings. This new framework treated multi-view data with several multivariate mixture distributions located in subspace projections, and handled prior knowledge in a form of sample-level constraints in order to indicate which objects should or should not be grouped together. Moreover, Yin et al.[121] presented a pairwise sparse subspace representation model for MvC. The designed model

harnessed the prior information in order to obtain the view-specific sparse representation, while utilizing the correlation between different views. Moreover, Cao et al.[122] put forward a constrained multi-view video face clustering method, which considers both the video face pairwise constraints and the multi-view consistence simultaneously. Unlike some existing clustering methods that only employ these constraints in the clustering phase, this method strengthens the pairwise constraints through the entire framework, namely, in sparse subspace representation and spectral clustering.

An incomplete MvC based on subspace learning has also been investigated. For instance, Yin et al.[123, 124] proposed an incomplete MvC method, which unified subspace learning, feature selection, and inter-view and intra-view similarity into a single objective function. It learns a latent representation for incomplete multi-view data, where this latent representation serves as an approximation of the normalized indicator matrix. Xu et al.[125] suggested that the key to deal with incomplete view problem is to exploit the connections between different views. This enables incomplete views to be restored with the assistance of complete views. They investigated the estimation of incomplete views with the help of information from other observed views through this subspace.

Example 6: Multi-view data is often incomplete, namely, data objects have incomplete feature sets. Based on subspace learning, Yin et al.[123, 124] studied incomplete multi-view learning for incomplete and unlabeled multi-view data. Figure 10 shows the presented subspace learning model, which learns a unified latent representation for incomplete multi-view data. This model directly optimizes the class indicator matrix, which establishes a bridge for incomplete feature sets. Moreover, feature selection is considered to deal with high dimensional and noisy features. Moreover, the structures of the inter-view data and intra-view data are preserved in order to enhance learning performance. To this end, an objective function was developed along with an efficient optimization algorithm.

Let $X = [X_c^v, \hat{X}^v] \in \mathbb{R}^{d_v \times (n_c^v + \hat{n}^v)}$ denote the v-th view data matrix, where $X_c^v \in \mathbb{R}^{d_v \times n_c^v}$ and $\hat{X} \in \mathbb{R}^{d_v \times \hat{n}^v}$ represent the data matrix in the v-th view for complete and partial instances, respectively. Similarly, $Y = [Y_c^v, \hat{Y}^v] \in \mathbb{R}^{(n_c^v + \hat{n}^v) \times k}$ represents the class indicator of the v-th view. To learn the class indicator
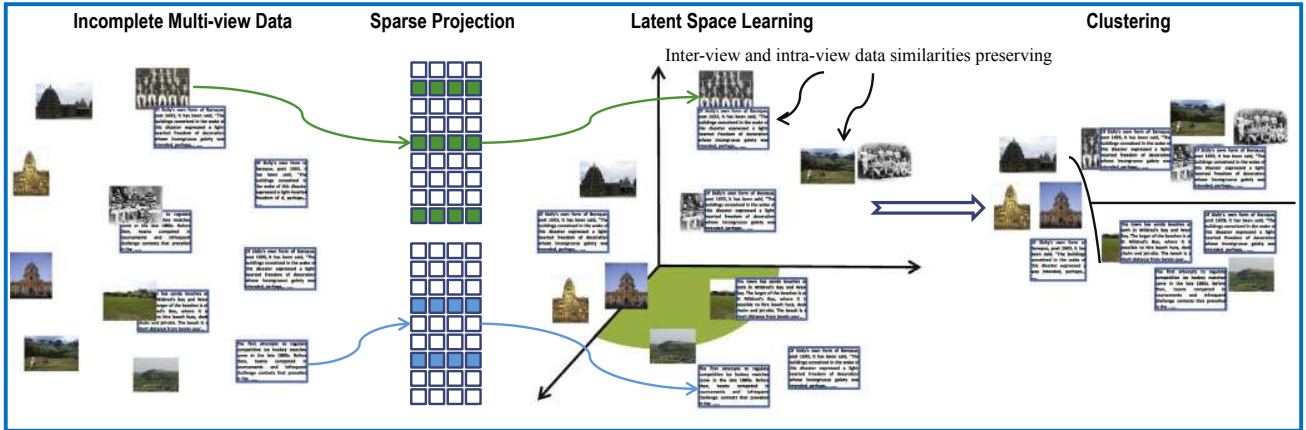
Fig. 10 The overview of the proposed model with two views, i.e., text and image[124]. For the incomplete multi-view dataset, it uses a projection matrix in order to project the original features (text and image) to a latent space, which explicitly captures the clustering structure. Moreover, group sparsity is imposed on projection matrices for feature selection. Moreover, inter-view and intra-view data similarities are preserved in order to enhance the model. Finally, the features on this latent space are applied to the clustering task.

matrix, it drives a projection matrix $U^v \in \mathbb{R}^{d_v \times k}$ for each view in order to project their original spaces to a unified space. The objective function[124] is formulated as

$$\min_{U,Y} \underbrace{\sum_{i=1}^{m} \|[X_c^v, \hat{X}^v]^{\mathrm{T}} U^i - [Y_c^v, \hat{Y}^v]\|_{\mathrm{F}}^2}_{\text{Feature projection}} + \beta \underbrace{\sum_{i=1}^{m} \|U^i\|_{21}}_{\text{Feature learning}} +$$

$$\gamma \underbrace{\sum_{p=1}^{m} \sum_{q=1}^{m} \mathrm{tr}((U^p)^{\mathrm{T}} X^i L_{pq}(X)^{\mathrm{T}} U^q)}_{\text{Similarity preserving}},$$

$$\text{s.t., } Y \in \{0,1\}^{n \times k}, \ Y\mathbf{1}_k = \mathbf{1}_n \qquad (11)$$

where it has three terms: using the projection matrix to project each incomplete view to the latent space defined by Y; feature selection for each view based on l_{21}-norm regularization; inter-view and intra-view data similarity, which preserves the term defined by the Laplacian matrix L_{pq}. Moreover, the constraints imposed on Y guarantee that each example belongs to one group only.

3.4.2 NMF-based MvC

NMF, which was originally investigated as a dimensionality reduction technique[126], has emerged as an effective latent feature learning method. The non-negative constraint leads to the parts-based representation of samples, which accords with the cognitive process of the human brain, from the psychological and physiological evidences. Given an input non-negative data matrix $X \in \mathbb{R}^{d \times n}$, each column of X is a feature vector of one sample.

The NMF aims to find two non-negative matrices $W \in \mathbb{R}^{d \times p}$ and $H \in \mathbb{R}^{p \times n}$, whose product can adequately approximate the original matrix X. Here, the former matrix W is termed as the basis matrix (basic space), while the latter matrix H represents the coefficient matrix (representation feature), and p (in general, $p \ll \min\{n, d\}$) denotes the desired reduced dimension. The reconstruction processes[126] can be formulated as a *Frobenius norm* optimization problem, defined as

$$\min_{W,H} \|X - WH\|_{\mathrm{F}}^2, \quad \text{s.t., } W \geqslant 0, H \geqslant 0 \qquad (12)$$

Moreover, many variants of NMF have also been put forward, such as G-orthogonal NMF[127], Regularized NMF[128, 129], Convex and Semi-NMF[130], and Multi-layer NMF[131]. Li and Ding[132] proposed a survey on NMF for clustering, in which more details of NMF can be found.

In a multi-view scenario, a late integration approach via NMF was studied in Ref. [133]. The proposed approach takes the clustering results generated independently on each available view, constructs an intermediate matrix representation for these clustering results, and performs NMF on this representation in order to reconcile the groups arising from individual views. Unlike the approach presented in Ref. [133], which plugs the clustering results into NMF, Liu et al.[134] developed a new NMF-based MvC framework, which feeds the data directly into NMF and drives a fused representation. Reference [134] formulates a joint matrix factorization with normalization strategy

that pushes the representative result of each view toward a common consensus. Moreover, it provides a new insight into applying NMF to MvC. Inspired by this framework, some improved work has also been investigated. For instance, a semi-supervised MvC algorithm based on NMF with weight for each view was studied in Refs. [16, 135], where it discovers a partially shared latent representation. With this learned representation of multi-view data, a robust sparse regression model was introduced in order to predict the clustering results. Embedding the similarity matrices of the data points into NMF, He et al.[136] also focused on learning a shared latent representation for MvC. Chang et al.[137] developed a multi-view NMF algorithm and applied it to clothing image clustering, by taking a new regularization term in order to advocate the structural incoherence between the representing result of each view. Considering the local geometric structure of each view, and penalizing the disagreement of different views at the same time, Ou et al.[138] proposed another types of multi-view NMF with a patch alignment strategy. Instead of treating the objects or views as distinct positions, Xu et al.[139] introduced a new self-paced learning algorithm with a smoothed weighting scheme, which inherits the merits of logistic function and provides probabilistic weights.

Many other methods based on the NMF variants have also been investigated. Based on G-orthogonal NMF[127], Cai et al.[140] presented a robust multi-view K-means clustering algorithm for large-scale multi-view data. Qian and Zhai[141] proposed a multi-view unsupervised method for text-image web news data, where image local learning regularized orthogonal NMF was adopted in order to learn pseudo labels, and robust joint $l_{2,1}$-norm was performed in order to select discriminative features. Zhao et al.[142] presented a deep matrix factorization framework via Multi-layer NMF[131] for MvC, where Semi-NMF[130] was used to learn a hierarchical representation for multi-view data in a layer-wise manner. Multiple sparse views clustering approaches with $l_{2,1}$-norm and group l_1-norm have also been investigated in Refs. [143–145]. Moreover, graph (or manifold) regularized NMF for MvC has also attracted attention. Graph regularized NMF[129] is an extension of NMF, which has been shown to improve the quality of the X factorization based on a manifold assumption, i.e., if two data points are close in the intrinsic geometry, then the representation of these two data points in the new

basis space are also close to each other. Motivated by Refs. [129, 134], Hidru and Goldenberg[146], and Wang et al.[147] investigated graph-regularized multi-view NMF-based clustering respectively, both of which had little difference. Rather than taking graph regularization from unlabeled data, Guan et al.[148] constructed the graph embedding framework through partial label information and considered the sparseness constraints at the same time. Multi-manifold regularized NMF goes into Refs. [149, 150], in which multiple manifolds were combined linearly and two kinds of MvC methods were led by different strategies. Based on SymNMF[151], Zhang et al.[152] introduced a graph regularized symmetric NMF framework for MvC. Furthermore, graph regularized MvC approaches via concept factorization[153] were discussed in Refs. [154, 155].

In the research field of incomplete MvC, Shao et al.[156, 157] made some attempts via NMF. The main idea was to incorporate weighted NMF in order to handle the missing objects in each incomplete view, and pushing the learned latent representation feature towards a consensus. Shao et al.[158] also proposed a general framework for incomplete data via tensor modeling and factorization. This framework first uses the kernel matrices in order to generate an initial tensor across all views, and then formulates a joint tensor factorization process with the sparsity constraint. This process is used to iteratively push the initial tensor towards an exploration of the latent factors. Moreover, the later fusion method based on NMF[133] can also handle incomplete views as author notes. In addition, Li et al.[159] presented a partial multi-view clustering algorithm (named PVC), which is specifically designed for two-view datasets. It employs NMF in order to learn a latent subspace, in which the samples belonging to the same group are close to each other and the similar samples from the same view should be grouped well. Later on, Qian et al.[160] improved PVC by considering the cluster similarity and manifold preserving constraints. Furthermore, Rai et al.[161] extended PVC to support multi-views and view-specific graph Laplacian regularization. With the help of inter-view constraints (i.e., must-link and cannot-link constraints), Zhang et al.[162] defined a disagreement between each pair of views in order to guide the factorization process.

Example 7: Wang et al.[154] proposed an auto-weighted Multi-view Concept Clustering (MvCC)

based on concept factorization with local manifold regularization. The MvCC framework is shown in Fig. 11. In brief, concept factorization[153] is a variant of NMF, which is available for handling the data containing negative, and is also easily performed in the kernel space. Furthermore, the local manifold regularization is incorporated into the concept factorization process in order to preserve the locally geometrical structure of the original data space. Both weights of each view are determined automatically and the given co-normalized scheme makes fusion meaningful in terms of driving the common consensus representation. In addition, the clustering results are driven directly from the common consensus representation, without requiring additional clustering steps. This is due to that the consensus matrix being sparse.

Given a multi-view data $X = \{X^1, ..., X^m\}$, the objective function of MvCC[154] is formulated as

$$\min_{H^*} \sum_{v=1}^{m} \left\{ \begin{array}{c} \underbrace{\|X^v - X^v W^v (H^v)^{\mathrm{T}}\|_{\mathrm{F}}^2}_{\text{Concept factorization}} + \\ \alpha \underbrace{\mathrm{tr}((H^v)^{\mathrm{T}} L^v H^v)}_{\text{Local manifold regularization}} + \\ \beta \underbrace{\omega_v \|H^v - H^*\|_{\mathrm{F}}^2}_{\text{Consensus representation}} \end{array} \right\},$$

$$\text{s.t.,} \quad W^v \geqslant 0, H^v \geqslant 0, \omega_v \geqslant 0, \sum_v \omega_v = 1 \tag{13}$$

where W^v is the association matrix, H^v is the representation matrix, H^* is the consensus representation matrix, L^v is the Laplacian matrix, ω_v is the weight of the v-th view, α and β are the trade-off parameters. It is worth stressing that the weight ω is determined automatically, and the parameters α and β are suggested with empirical values, while, as the objective function Formula (13) is not convex over

all the variables simultaneously, and an alternating iterative algorithm based on the multiplicative update rules was developed to optimize it. More details can be found in Ref. [154].

3.5 Multi-task multi-view clustering

MvC exploits the consistency and complementarity among different views in order to achieve better clustering quality, as mentioned above. Another concept, namely, multi-task clustering (belonging to the field of multi-task learning[163]), performs multiple related tasks together and utilizes the relationship between these tasks in order to enhance clustering performance for single-view data. By inheriting the property of both MvC and multi-task clustering, the Multi-task Multi-view Clustering (M^2vC) treats each individual view data with one task or multiple tasks as shown in Fig. 12. This has received some attention in recent years. The main challenges of M^2vC consist of finding a way to model the intra-task (within-task) clustering on each view, and a way to exploit the multi-task and multi-view relationship, while transferring the inter-task (between-task) knowledge to one another. Here, we provide a review on M^2vC in order to attract further attention and promote research in this area.

By assuming that a common underlying subspace is shared by multiple related tasks, Gu and Zhou[164] proposed a cross-domain multi-task clustering method by treading each view with a task. This method aims to learn such a subspace, and through it, the knowledge of one task can be transferred to another. Note that the authors also assumed that the dimensionality of the feature vector for each task is the same, and the number of clusters in each task is also the same. Later on, Zhang and Zhou[165] relaxed these assumptions, and introduced an improved
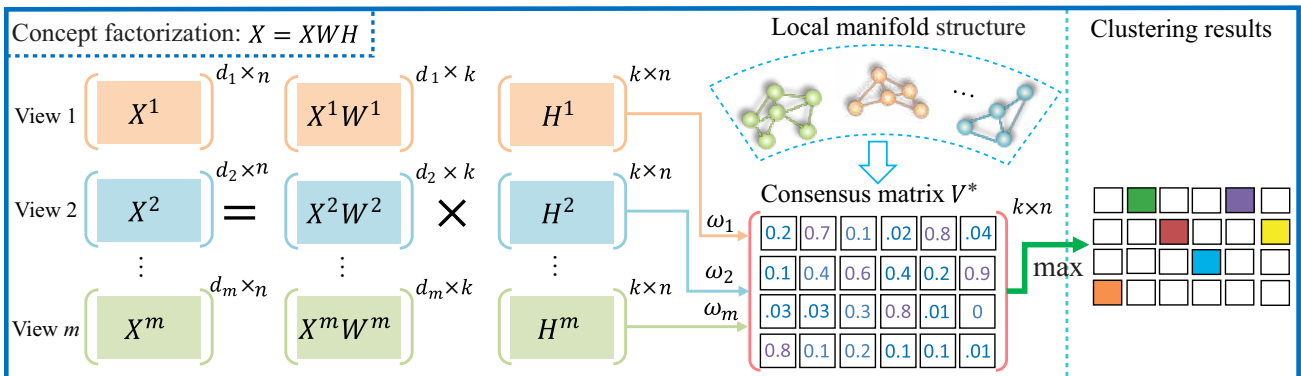

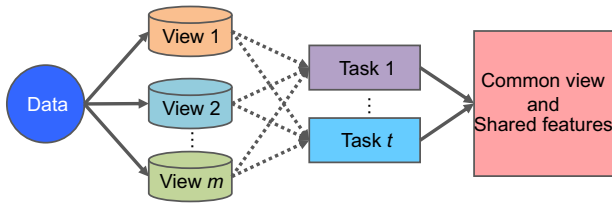**Fig. 11 The graphic presentation of the MvCC model.**

Fig. 12 General procedure of multi-task multi-view clustering.

cross-domain multi-task clustering, which can performs multiple related clustering tasks simultaneously through domain adaptation. Besides, Xie et al.[166] presented a multi-task co-clustering based on 3-factor NMF. The objective function of this method consisted of two parts, i.e., task-specific co-clustering and cross-task feature space regularization. The multi-task clustering method via SymNMF[151] for multi-view data has also been studied in Ref. [167], where several tasks were performed simultaneously with a geometric affine transformation in order to control intra-task and inter-task knowledge sharing. In addition, Wang et al.[168] explored multi-view spectral clustering by using a multi-objective formulation (seen as a multi-task problem), which is solved by Pareto optimization. Wahid et al.[169, 170] studied a multi-objective MvC ensemble method based on an evolutionary approach. Zhang et al.[171] developed a multi-task clustering algorithm by transferring the knowledge of instances. The proposed algorithm learns a shared subspace, and constructs a shared nearest neighbor similarity matrix for each individual task. Then, it applies a traditional spectral clustering method on the shared nearest neighbor similarity matrix of each task. Shi et al.[172] incorporated spectral clustering and discriminative analysis into a unified framework by exploiting the correlation information between multiple views, where spectral clustering aims to discover the cluster structure, and the discriminative analysis aims to preserve the structure. Moreover, Zhang et al.[173, 174] presented an M^2vC framework, which integrates within-view-task clustering, multi-view relationship learning, and multi-task relationship learning. Under this framework, they proposed two M^2vC algorithms, i.e., the bipartite graph based M^2vC algorithm, and the semi-NMF based M^2vC algorithm.

Upon reviewing the literature on MvC, we found that some existing work is hard to be assigned to any of the five above-mentioned categories. Here, we provide a brief summary, including multi-modal clustering based

on Markov random field[175], multi-view clustering ensembles based on multi-view spectral clustering and multi-view kernel K-means clustering with ensemble technology[176], bi-level weighted MvC based on K-means[177–180], and multi-view fuzzy clustering[181–184].

Example 8: In multi-task multi-view settings, the tasks are related through common views. The key step of M^2vC is to link the features in the common view in order to integrate the related tasks. In the field of M^2vC, Zhang et al.[174] developed a typical M^2vC framework based on co-clustering. An illustration of this framework is shown in Fig. 13, where the square region represents the set of data samples, and the circular region represents the set of data features under a view in each task as described in Ref. [174]. Note that the samples of task 1 and task 2 have a common view, which contains task shared features (denoted by the light gray overlapping area) and task specific features (denoted by the light gray non-overlapping area). This framework consists of three components: within-view-task clustering, multi-view relationship learning, and multi-task relationship learning. Under this framework, they proposed two M^2vC algorithms. One is the bipartite graph based M^2vC algorithm, which only handles the data containing non-negative values. Another one is the semi-non-negative matrix tri-factorization based M^2vC algorithm, which is a general M^2vC method, i.e., it can deal with the data containing negative or non-negative values.

Given T clustering tasks, each task is covered with m_t views. S is the index collection of the common views; T_v is the index collection of all tasks under this common view. For within-view-task clustering, it treats the data objects in each view of each task with co-clustering, which accomplishes the essential part of the whole algorithm and ensures the preservation of the knowledge available locally at each view of each task in order to avoid negative transfer. For multi-view relationship learning, it minimizes the disagreement between the clusters of data under each pair of views
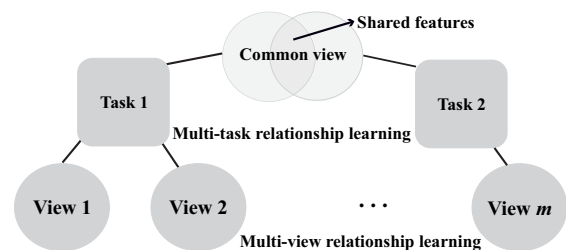

Fig. 13 Co-clustering based M^2vC framework[174].

in each task. For multi-task relationship learning, it uses co-clustering in order to drive a shared subspace among the related tasks under each common view by assuming that related tasks should share some common or relevant features. The M^2vC clustering framework[174] is formalized as

$$\min \sum_{t=1}^{T} \left(\sum_{i=1}^{m_t} H_1 + \lambda \sum_{i=1}^{m_t} \sum_{j \neq i}^{m_t} H_2 \right) + \mu \sum_{i \in S} \sum_{t \in T_i} H_3$$

$$(14)$$

where $\sum_{i=1}^{m_t} H_1$ is to co-cluster data objects and features of all the views in each task t, $\sum_{i=1}^{m_t} \sum_{j \neq i}^{m_t} H_2$ is to minimize the disagreement between the clustering assignments of any two different views in each task t, $\sum_{i \in S} \sum_{t \in T_i} H_3$ is to obtain the shared subspace under each common view by the same co-clustering method as the first component. λ and μ are trade-off parameters. Under this framework, two specific clustering algorithms were investigated in Ref. [174].

4 Publically Available Datasets

To support researchers working in the field of MvC, we summarize some widely used multi-view datasets. For these publically available datasets, we provide their URLs.

3Sources Dataset[*]: A multi-view text corpus, constructed from news articles from three online news services. This repository also has **Multi-View Twitter Datasets**, a collection of Twitter datasets for social networks discovery, and **BBC** and **BBCSport Datasets**, two synthetic text datasets originating from BBC News.

WebKb Datasets[†]: These datasets contain web-page data collected from the computer science departments of four universities, namely, four multi-view datasets.

Newsgroup Datasets[‡]: There are subsets of the NG20 dataset with 3 different pre-processings. The description of the subsets, and details on the preprocessing steps can be found in Ref. [185]. Moreover, this repository also has the **Reuters Multilingual Dataset**, **Cora Dataset**, **CiteSeer Dataset**, **Movies617 Dataset**, and mini **Wekb Datasets**.

Wikipedia Article Dataset[§]: The collected datasets are selected sections from Wikipedia's featured articles collection. They are available in full or small versions.

Handwritten Digit Dataset[¶]: It consists of features of handwritten numerals (0–9) from the UCI repository.

100leaves Dataset[‖]: It contains sixteen different kinds of plant leafs, where each kind has one-hundred samples. For each sample, the shape descriptor, fine scale margin, and the texture histogram are given.

Corel Images Dataset[**]: This dataset consists of image features extracted from a Corel image collection. It provides four sets of features, namely, color histogram, color histogram layout, color moments, and co-occurence texture.

NUS-WIDE Dataset[††]: A web image dataset with six types of low-level features extracted from these images.

YouTube Video Dataset[‡‡]: This dataset contains approximately 1.2×10^5 instances, where each instance is described by 13 types of features, and also has its class information.

5 Conclusion and Discussion

The proliferation of multi-view data calls for advanced clustering technologies that can discover knowledge from multi-view datasets. This paper surveyed most of the existing algorithms and technologies of MvC, and classified these MvC algorithms into five categories, i.e., co-training style algorithm, multi-kernel learning, multi-view graph clustering, multi-view subspace clustering, and multi-task multi-view clustering. For each category of MvC, we did not only review the existing algorithms, but also introduced the ideas and technologies behind them, while giving specific illustrative examples.

Although MvC was proposed around 2003, as we mentioned in the introduction section, there is no criterion to decide which MvC algorithm is the best, since different methods have their own advantages and

---

[*]http://mlg.ucd.ie/datasets/bbc.html,
[†]http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-20/www/data/.
[‡]http://lig-membres.imag.fr/grimal/data.html.
[§]http://www.svcl.ucsd.edu/projects/crossmodal/

[¶]http://archive.ics.uci.edu/ml/datasets/Multiple+Features
[‖]https://archive.ics.uci.edu/ml/datasets/One-hundred+plant+species+leaves+data+set
[**]https://archive.ics.uci.edu/ml/datasets/Corel+Image+Features
[††]http://lms.comp.nus.edu.sg/research/NUS-WIDE.htm
[‡‡]https://archive.ics.uci.edu/ml/datasets/YouTube+Multiview+Video+Games+Dataset

disadvantages. In brief, co-training style algorithms can enhance the clusters of different views interactively by exchanging information. However, they are intractable when the number of views is more than three. Kernel based MvC inherits the advantage of kernel, while bringing about high computational complexity. Multi-view graph clustering introduces spectral graph theory, while relying on the constructed affinity (or similarity) matrices. Multi-view subspace clustering methods have straightforward interpretability, and also have initialization dependence. Multi-task multi-view inherits both properties of multi-task clustering and multi-view clustering; however, this is still in infancy. Hopefully, these technologies have close relationship to one another. For example, subspace learning can be performed on the kernel space, therefore, it is valuable in developing the general framework of MvC, which inherits the merits of different categories.

Below, we would like to highlight a number of challenging problems and future directions in order to encourage more research in MvC. Their solutions will have a fundamental impact on MvC, specifically, on multi-view data fusion, machine learning, and artificial intelligence in general.

- *Correctness of views*: Finding a way of knowing whether a view is correct, is crucial for MvC. Since MvC exploits all available views in order to help clustering performance, incorrect views are very harmful. Although some work leverages these views with weights, errors could be propagated from a misleading view to other views. Thus, this problem must be solved or mitigated to a great extent in order to ensure that MvC is effective.
- *The opportune moment of fusion*: Existing MvC adopts three fusion strategies for multi-view data in the clustering process, namely, fusion in the data, fusion in the projected features, and fusion in the results. Most of the current research works of MvC focus on the second fusion strategy. However, there is no theoretical foundation to decide which one is the best. Theoretical and methodological research is required in order to uncover their essence.
- *Incomplete MvC*: Although some attempts have been made for incomplete multi-view data, as we mentioned in each section of the category, incomplete MvC is still a challenging problem. In real-life, data loss occurs frequently, while the research in incomplete MvC has not been

extensive. Effort is expected to be put into the investigation of the incomplete MvC.
- *Multi-task multi-view clustering*: This direction is a new trend in the research of MvC; however, this trend is accompanied by a few challenges such as finding a way to explore the relationships of different tasks and different views, and finding a way to transfer the knowledge between each other views.

In addition, several widely used datasets were listed in order to provide convenience for future researchers. In summary, this paper serves as a bridge for readers in order to further promote the research of MvC.

Acknowledgment

References

[1] C. Xu, D. C. Tao, and C. Xu, A survey on multi-view learning, arXiv preprint arXiv: 1304.5634, 2013.

[2] S. L. Sun, A survey of multi-view machine learning, *Neural Comput. Appl.*, vol. 23, nos. 7&8, pp. 2031–2038, 2013.

[3] J. Zhao, X. J. Xie, X. Xu, and S. L. Sun, Multi-view learning overview: Recent progress and new challenges, *Information Fusion*, vol. 38, pp. 43–54, 2017.

[4] Y. Zheng, Methodologies for cross-domain data fusion: An overview, *IEEE Trans. Big Data*, vol. 1, no. 1, pp. 16–34, 2015.

[5] R. Xu and D. Wunsch, Survey of clustering algorithms, *IEEE Trans. Neural Netw.*, vol. 16, no. 3, pp. 645–678, 2005.

[6] C. C. Aggarwal and C. K. Reddy, *Data Clustering: Algorithms and Applications*. Boca Raton, FL, USA: Chapman and Hall/CRC, 2013.

[7] A. Fahad, N. Alshatri, Z. Tari, A. Alamri, I. Khalil, A. Y. Zomaya, S. Foufou, and A. Bouras, A survey of clustering algorithms for big data: Taxonomy and empirical analysis, *IEEE Trans. Emerg. Top. Comput.*, vol. 2, no. 3, pp. 267–279, 2014.

[8] J. D. Wang, H. J. Zeng, Z. Chen, H. J. Lu, L. Tao, and W. Y. Ma, ReCoM: Reinforcement clustering of multi-type interrelated data objects, in *Proc. 26th Annu. Int. ACM SIGIR Conf. RESEARCH and Development in Information Retrieval*, 2003, pp. 274–281.

[9] K. Kailing, H. P. Kriegel, A. Pryakhin, and M. Schubert, Clustering multi-represented objects with noise, in *Proc. 8th Pacific-Asia Conf. Advances in Knowledge Discovery and Data Mining*, Sydney, Australia, 2004, pp. 394–403.

[10] S. Bickel and T. Scheffer, Multi-view clustering, in *Proc. 4th IEEE Int. Conf. Data Mining*, Brighton, UK, 2004, pp. 19–26.

[11] MultiClust: 1st international workshop on discovering, summarizing and using multiple clusterings held in

conjunction with KDD-2010, http://eecs.oregonstate.edu/research/multiclust/, 2010.

[12] 2ed MultiClust Workshop: Discovering, summarizing and using multiple clusterings held in conjunction with ECML/PKDD 2011, http://dme.rwth-aachen.de/en/MultiClust2011, 2011.

[13] 3rd MultiClust Workshop: Discovering, summarizing and using multiple clusterings held in conjunction with ICDM 2012, http://www.dbs.ifi.lmu.de/research/MultiClust2012/, 2012.

[14] MultiClust: 4th international workshop on discovering, 20 big data mining and analytics, summarizing and using multiple clusterings held in conjunction KDD, http://cs.au.dk/research/data-intensive-systems/projects/multiclust2013/, 2013.

[15] MultiClust 2014: Mini-symposium on multiple clusterings, multi-view data, and multi-source knowledge-driven held in conjunction with SDM, http://www.wikicfp.com/cfp/servlet/event.showcfp?eventid=33931, 2014.

[16] J. Liu, Y. Jiang, Z. C. Li, Z. H. Zhou, and H. Q. Lu, Partially shared latent factor learning with multiview data, *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 6, pp. 1233–1246, 2015.

[17] S. Dasgupta, M. L. Littman, and D. A. McAllester, PAC generalization bounds for co-training, in *Proc. 14th Int. Conf. Neural Information Processing Systems: Natural and Synthetic*, Vancouver, Canada, 2001, pp. 375–382.

[18] A. Blum and T. Mitchell, Combining labeled and unlabeled data with co-training, in *Proc. 11th Annu. Conf. Computational Learning Theory*, Madison, WI, USA, 1998, pp. 92–100.

[19] V. R. De Sa, Spectral clustering with two views, in *Proc. 22nd Workshop on Learning with Multiple Views*, Bonn, Germany, 2005, pp. 20–27.

[20] S. Abney, Bootstrapping, in *Proc. 40th Annu. Meeting of the Association for Computational Linguistics*, Philadelphia, PA, USA, 2002, pp. 360–367.

[21] M. F. Balcan, A. Blum, and K. Yang, Co-training and expansion: Towards bridging theory and practice, in *Proc. 17th Int. Conf. Neural Information Proc. Systems*, Vancouver, Canada, 2004, pp. 89–96.

[22] W. Wang and Z. H. Zhou, Analyzing co-training style algorithms, in *Proc. 18th European Conf. Machine Learning*, Warsaw, Poland, 2007, pp. 454–465.

[23] K. Nigam and R. Ghani, Analyzing the effectiveness and applicability of co-training, in *Proc. 9th Int. Conf. Information and Knowledge Management*, McLean, VA, USA, 2000, pp. 86–93.

[24] V. Sindhwani, P. Niyogi, and M. Belkin, A co-regularization approach to semi-supervised learning with multiple views, in *Proc. 22nd Workshop on Learning with Multiple Views*, Bonn, Germany, 2005, pp. 74–79.

[25] Y. Z. Cheng and G. M. Church, Biclustering of expression data, in *Proc. 8th Int. Conf. Intelligent Systems for Molecular Biology*, Menlo Park, CA, USA, 2000, pp. 93–103.

[26] S. Bickel and T. Scheffer, Estimation of mixture models using Co-EM, in *Proc. 16th European Conf. Machine Learning*, Porto, Portugal, 2005, pp. 35–46.

[27] G. F. Tzortzis and A. C. Likas, Multiple view clustering using a weighted combination of exemplar-based mixture models, *IEEE Trans. Neural Netw.*, vol. 21, no. 12, pp. 1925–1938, 2010.

[28] A. Kumar and H. Daumé III, A co-training approach for multi-view spectral clustering, in *Proc. 28th Int. Conf. Machine Learning*, Bellevue, WA, USA, 2011, pp. 393–400.

[29] A. Kumar, P. Rai, and H. Daumé III, Co-regularized multi-view spectral clustering, in *Proc. 24th Int. Conf. Neural Information Processing Systems*, Granada, Spain, 2011, pp. 1413–1421.

[30] Y. K. Ye, X. W. Liu, J. P. Yin, and E. Zhu, Co-regularized kernel k-means for multi-view clustering, in *Proc. 23rd Int. Conf. Pattern Recognition*, Cancun, Mexico, 2016, pp. 1583–1588.

[31] A. Appice and D. Malerba, A co-training strategy for multiple view clustering in process mining, *IEEE Trans. Serv. Comput.*, vol. 9, no. 6, pp. 832–845, 2016.

[32] Y. Jiang, J. Liu, Z. C. Li, P. Li, and H. Q. Lu, Co-regularized PLSA for multi-view clustering, in *Proc. 11th Asian Conf. Computer Vision-Volume Part II*, Daejeon, Korea, 2013, pp. 202–213.

[33] E. Eaton, M. desJardins, and S. Jacob, Multi-view clustering with constraint propagation for learning with an incomplete mapping between views, in *Proc. 19th ACM Int. Conf. Information and Knowledge Management*, Toronto, Canada, 2010, pp. 389–398.

[34] E. Eaton, M. desJardins, and S. Jacob, Multi-view constrained clustering with an incomplete mapping between views, *Knowl. Inf. Syst.*, vol. 38, no. 1, pp. 231–257, 2014.

[35] L. Meng, A. H. Tan, and D. Xu, Semi-supervised heterogeneous fusion for multimedia data co-clustering, *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 9, pp. 2293–2306, 2014.

[36] J. W. Sun, J. Lu, T. Y. Xu, and J. B. Bi, Multi-view sparse co-clustering via proximal alternating linearized minimization, in *Proc. 32nd Int. Conf. Machine Learning*, Lille, France, 2015, pp. 757–766.

[37] G. Bisson and C. Grimal, An architecture to efficiently learn co-similarities from multi-view datasets, in *Proc. 19th Int. Conf. Neural Information Proc.*, Doha, Qatar, 2012, pp. 184–193.

[38] G. Bisson and C. Grimal, Co-clustering of multi-view datasets: A parallelizable approach, in *Proc. IEEE 12th Int. Conf. Data Mining*, Brussels, Belgium, 2012, pp. 828–833.

[39] S. F. Hussain and S. Bashir, Co-clustering of multi-view datasets, *Knowl. Inf. Syst.*, vol. 47, no. 3, pp. 545–570, 2016.

[40] Y. Jiang, J. Liu, Z. C. Li, and H. Q. Lu, Collaborative PLSA for multi-view clustering, in *Proc. 21st Int. Conf. Pattern Recognition*, Tsukuba, Japan, 2012, pp. 2997–3000.

[41] M. Ghassany, N. Grozavu, and Y. Bennani, Collaborative multi-view clustering, in *Proc. 2013 Int. Joint Conf. Neural Networks*, Dallas, TX, USA, 2013, pp. 1–8.

[42] L. L. Xu, J. Neufeld, B. Larson, and D. Schuurmans, Maximum margin clustering, in *Proc. 18th Annu. Conf. Advances in Neural Information Processing Systems 17*, Cambridge, MA, USA, 2005, pp. 1537–1544.

[43] B. Zhao, J. T. Kwok, and C. S. Zhang, Multiple kernel clustering, in *Proc. 2009 SIAM Int. Conf. Data Mining*, Sparks, NV, USA, 2009, pp. 638–649.

[44] L. Du, P. Zhou, L. Shi, H. M. Wang, M. Y. Fan, W. J. Wang, and Y. D. Shen, Robust multiple kernel k-means using $\ell_{2,1}$-norm, in *Proc. 24th Int. Conf. Artificial Intelligence*, Buenos Aires, Argentina, 2015, pp. 3476–3482.

[45] V. R. De Sa, P. W. Gallagher, J. M. Lewis, and V. L. Malave, Multi-view kernel construction, *Mach. Learn.*, vol. 79, nos. 1&2, pp. 47–71, 2010.

[46] V. R. De Sa, Learning classification with unlabeled data, in *Advances in Neural Information Processing Systems 6*, Cambridge, MA, USA, 1993, pp. 112–119.

[47] V. R. De Sa and D. H. Ballard, Category learning through multimodality sensing, *Neural Comput.*, vol. 10, no. 5, pp. 1097–1117, 1998.

[48] S. Yu, L. C. Tranchevent, X. H. Liu, W. Glanzel, J. A. K. Suykens, B. De Moor, and Y. Moreau, Optimized data fusion for kernel k-means clustering, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 5, pp. 1031–1039, 2012.

[49] M. Gönen and A. A. Margolin, Localized data fusion for kernel k-means clustering with application to cancer biology, in *Proc. 28th Annu. Conf. Neural Information Proc. Systems*, Montreal, Canada, 2014, pp. 1305–1313.

[50] Y. T. Lu, L. T. Wang, J. F. Lu, J. Y. Yang, and C. H. Shen, Multiple kernel clustering based on centered kernel alignment, *Pattern Recognit.*, vol. 47, no. 11, pp. 3656–3664, 2014.

[51] G. Tzortzis and A. Likas, Kernel-based weighted multi-view clustering, in *Proc. 12th Int. Conf. Data Mining*, Brussels, Belgium, 2012, pp. 675–684.

[52] D. Y. Guo, J. Zhang, X. W. Liu, Y. Cui, and C. X. Zhao, Multiple kernel learning based multi-view spectral clustering, in *Proc. 22nd Int. Conf. Pattern Recognition*, Stockholm, Sweden, 2014, pp. 3774–3779.

[53] X. W. Liu, Y. Dou, J. P. Yin, L. Wang, and E. Zhu, Multiple kernel k-means clustering with matrix-induced regularization, in *Proc. 30th AAAI Conf. Artificial Intelligence*, Phoenix, AZ, USA, 2016, pp. 1888–1984.

[54] Y. Zhao, Y. Dou, X. W. Liu, and T. Li, A novel multi-view clustering method via low-rank and matrix-induced regularization, *Neurocomputing*, vol. 216, pp. 342–350, 2016.

[55] P. R. Zhang, Y. Yang, B. Peng, and M. J. He, Multi-view clustering algorithm based on variable weight and MKL, in *Proc. Int. Joint Conf. Rough Sets*, Olsztyn, Poland, 2017, pp. 599–610.

[56] A. Trivedi, P. Rai, H. Daumé III, and S. L. DuVall, Multiview clustering with incomplete views, in *Proc. Workshop on Machine Learning for Social Computing*, Whistler, Canada, 2010.

[57] W. X. Shao, X. X. Shi, and P. S. Yu, Clustering on multiple incomplete datasets via collective kernel learning, in *Proc. 13th Int. Conf. Data Mining*, Dallas, TX, USA, 2013, pp. 1181–1186.

[58] X. W. Liu, M. M. Li, L. Wang, Y. Dou, J. P. Yin, and E. Zhu, Multiple kernel k-means with incomplete kernels, in *Proc. 31st AAAI Conf. Artificial Intelligence*, San Francisco, CA, USA, 2017, pp. 2259–2265.

[59] W. Tang, Z. D. Lu, and I. S. Dhillon, Clustering with multiple graphs, in *Proc. 9th IEEE Int. Conf. Data Mining*, Miami, FL, USA, 2009, pp. 1016–1021.

[60] S. F. Hussain, M. Mushtaq, and Z. Halim, Multi-view document clustering via ensemble method, *J. Intell. Inf. Syst.*, vol. 43, no. 1, pp. 81–99, 2014.

[61] A. Serra, D. Greco, and R. Tagliaferri, Impact of different metrics on multi-view clustering, in *Proc. 2015 Int. Joint Conf. Neural Networks*, Killarney, Ireland, 2015, pp. 1–8.

[62] Z. Xue, G. R. Li, S. H. Wang, C. J. Zhang, W. G. Zhang, and Q. M. Huang, GOMES: A group-aware multi-view fusion approach towards real-world image clustering, in *Proc. 2015 IEEE Int. Conf. Multimedia and Expo*, Turin, Italy, 2015, pp. 1–6.

[63] F. P. Nie, J. Li, and X. L. Li, Parameter-free auto-weighted multiple graph learning: A framework for multiview clustering and semi-supervised classification, in *Proc. 25th Int. Joint Conf. Artificial Intelligence*, New York, NY, USA, 2016, pp. 1881–1887.

[64] X. K. Wei, B. K. Cao, and P. S. Yu, Multi-view unsupervised feature selection by cross-diffused matrix alignment, in *Proc. 2017 Int. Joint Conf. Neural Networks*, Anchorage, AK, USA, 2017, pp. 494–501.

[65] C. P. Hou, F. P. Nie, H. Tao, and D. Y. Yi, Multi-view unsupervised feature selection with adaptive similarity and view weight, *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 9, pp. 1998–2011, 2017.

[66] A. Hamzaoui, A. Joly, and N. Boujemaa, Multi-source shared nearest neighbours for multi-modal image clustering, *Multimed. Tools Appl.*, vol. 51, no. 2, pp. 479–503, 2011.

[67] S. K. Wang, Y. M. Ye, and R. Y. K. Lau, A generative model with ensemble manifold regularization for multi-view clustering, in *Proc. 11th Int. Conf. Advanced Intelligent Computing Theories and Applications*, Fuzhou, China, 2015, pp. 109–114.

[68] S. K. Wang, E. K. Wang, X. T. Li, Y. M. Ye, R. Y. K. Lau, and X. L. Du, Multi-view learning via multiple graph regularized generative model, *Knowl.-Based Syst.*, vol. 121, pp. 153–162, 2017.

[69] Z. Y. Zhang and J. Y. Mao, Jointly sparse neighborhood graph for multi-view manifold clustering, *Neurocomputing*, vol. 216, pp. 28–38, 2016.

[70] F. P. Nie, G. H. Cai, and X. L. Li, Multi-view clustering and semi-supervised classification with adaptive neighbours, in *Proc. 31st AAAI Conf. Artificial Intelligence*, San Francisco, CA, USA, 2017, pp. 2408–2414. t

[71] W. Cheng, X. Zhang, Z. S. Guo, Y. B. Wu, P. F. Sullivan, and W. Wang, Flexible and robust co-regularized multi-domain graph clustering, in *Proc. 19th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, Chicago, IL, USA, 2013, pp. 320–328.

[72] J. C. Ni, H. H. Tong, W. Fan, and X. Zhang, Flexible and robust multi-network clustering, in *Proc. 21th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, Sydney, Australia, 2015, pp. 835–844.

[73] J. C. Ni, W. Cheng, W. Fan, and X. Zhang, Self-grouping multi-network clustering, in *Proc. 16th Int. Conf. Data Mining*, Barcelona, Spain, 2016, pp. 1119–1124.

[74] R. Liu, W. Cheng, H. H. Tong, W. Wang, and X. Zhang, Robust multi-network clustering via joint cross-domain cluster alignment, in *Proc. 2015 IEEE Int. Conf. Data Mining*, Atlantic City, NJ, USA, 2015, pp. 291–300.

[75] P. S. Yu and J. W. Zhang, MCD: Mutual clustering across multiple social networks, in *Proc. 2015 IEEE Int. Conf. Big Data*, New York, NY, USA, 2015, pp. 762–771.

[76] C. D. Wang, J. H. Lai, and P. S. Yu, Multi-view clustering based on belief propagation, *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 4, pp. 1007–1021, 2016.

[77] D. Y. Zhou and C. J. C. Burges, Spectral clustering and transductive learning with multiple views, in *Proc. 24th Int. Conf. Machine Learning*, Corvalis, OR, USA, 2007, pp. 1159–1166.

[78] Y. Cheng and R. L. Zhao, Multiview spectral clustering via ensemble, in *Proc. 2009 IEEE Int. Conf. Granular Computing*, Nanchang, China, 2009, pp. 101–106.

[79] B. Long, P. S. Yu, and Z. F. Zhang, A general model for multiple view unsupervised learning, in *Proc. 2008 SIAM Int. Conf. Data Mining*, Philadelphia, PA, USA, 2008, pp. 822–833.

[80] D. L. Niu, J. G. Dy, and M. I. Jordan, Multiple non-redundant spectral clustering views, in *Proc. 27th Int. Conf. Int. Conf. Machine Learning*, Haifa, Israel, 2010, pp. 831–838.

[81] R. K. Xia, Y. Pan, L. Du, and J. Yin, Robust multi-view spectral clustering via low-rank and sparse decomposition, in *Proc. 28th AAAI Conf. Artificial Intelligence*, Québec City, Canada, 2014, pp. 2149–2155.

[82] Y. Q. Li, F. P. Nie, H. Huang, and J. Z. Huang, Large-scale multi-view spectral clustering via bipartite graph, in *Proc. 29th AAAI Conf. Artificial Intelligence*, Austin, TX, USA, 2015, pp. 2750–2756.

[83] N. F. Chikhi, Multi-view clustering via spectral partitioning and local refinement, *Information Processing and Management*, vol. 52, no. 4, pp. 618–627, 2016.

[84] C. Y. Lu, S. C. Yan, and Z. C. Lin, Convex sparse spectral clustering: Single-view to multi-view, *IEEE Trans. Image Proc.*, vol. 25, no. 6, pp. 2833–2843, 2016.

[85] J. W. Son, J. Jeon, A. Lee, and S. J. Kim, Spectral clustering with brainstorming process for multi-view data, in *Proc. 31st AAAI Conf. Artificial Intelligence*, San Francisco, CA, USA, 2017, pp. 2548–2554.

[86] H. C. Huang, Y. Y. Chuang, and C. S. Chen, Affinity aggregation for spectral clustering, in *Proc. 2012 IEEE Conf. Computer Vision and Pattern Recognition*, Providence, RI, USA, 2012, pp. 773–780.

[87] W. X. Shao, J. W. Zhang, L. F. He, and P. S. Yu, Multi-source multi-view clustering via discrepancy penalty, in *Proc. 2016 Int. Joint Conf. Neural Networks*, Vancouver, Canada, 2016, pp. 2714–2721.

[88] L. Feng, L. Cai, Y. Liu, and S. L. Liu, Multi-view spectral clustering via robust local subspace learning, *Soft Comput.*, vol. 21, no. 8, pp. 1937–1948, 2017.

[89] Y. Wang, W. J. Zhang, L. Wu, X. M. Lin, M. Fang, and S. R. Pan, Iterative views agreement: An iterative low-rank based structured optimization method to multi-view spectral clustering, in *Proc. 25th Int. Joint Conf. Artificial Intelligence*, New York, NY, USA, 2016, pp. 2153–2159.

[90] P. Zhao, Y. Jiang, and Z. H. Zhou, Multi-view matrix completion for clustering with side information, in *Proc. 21st Pacific-Asia Conf. Advances in Knowledge Discovery and Data Mining*, Jeju, South Korea, 2017, pp. 403–415.

[91] B. Long, Z. F. Zhang, X. Y. Wú, and P. S. Yu, Spectral clustering for multi-type relational data, in *Proc. 23rd Int. Conf. Machine Learning*, Pittsburgh, PA, USA, 2006, pp. 585–592.

[92] X. Cai, F. P. Nie, H. Huang, and F. Kamangar, Heterogeneous image feature integration via multi-modal spectral clustering, in *Proc. 2011 IEEE Conf. Computer Vision and Pattern Recognition*, Colorado Springs, CO, USA, 2011, pp. 1977–1984.

[93] J. L. Tang, X. Hu, H. J. Gao, and H. Liu, Unsupervised feature selection for multi-view data in social media, in *Proc. 2013 SIAM Int. Conf. Data Mining*, Austin, TX, USA, 2013, pp. 270–278.

[94] H. B. Chen, K. M. Li, D. J. Zhu, X. Jiang, Y. X. Yuan, P. L. Lv, T. Zhang, L. Guo, D. G. Shen, and T. M. Liu, Inferring group-wise consistent multimodal brain networks via multi-view spectral clustering, *IEEE Trans. Med. Imag.*, vol. 32, no. 9, pp. 1576–1586, 2013.

[95] Y. H. Yang, C. Lan, X. L. Li, B. Luo, and J. Huan, Automatic social circle detection using multi-view clustering, in *Proc. 23rd ACM Int. Conf. Conf. Information and Knowledge Management*, Shanghai, China, 2014, pp. 1019–1028.

[96] Y. Zhang, X. H. Hu, and X. P. Jiang, Multi-view clustering of microbiome samples by robust similarity network fusion and spectral clustering, *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 14, no. 2, pp. 264–271, 2017.

[97] K. Chaudhuri, S. M. Kakade, K. Livescu, and K. Sridharan, Multi-view clustering via canonical correlation analysis, in *Proc. 26th Annu. Int. Conf. Machine Learning*, Montreal, Canada, 2009, pp. 129–136.

[98] Y. H. Guo, Convex subspace representation learning from multi-view data, in *Proc. 27th AAAI Conf. Artificial Intelligence*, Bellevue, WA, USA, 2013, pp. 387–393.

[99] X. R. Zhao, N. Evans, and J. L. Dugelay, A subspace co-training framework for multi-view clustering, *Pattern Recogn. Lett.*, vol. 41, pp. 73–82, 2014.

[100] Q. Deng, Y. Yang, M. He, and H. Xing, Locally adaptive feature weighting for multiview clustering, in *Proc. 12th Int. FLINS Conf. Uncertainty Modelling in Knowledge Engineering and Decision Making*, Roubaix, France, 2016, pp. 139–145.

[101] X. C. Cao, C. Q. Zhang, H. Z. Fu, S. Liu, and H. Zhang, Diversity-induced multi-view subspace clustering, in *Proc. 2015 IEEE Conf. Computer Vision and Pattern Recognition*, Boston, MA, USA, 2015, pp. 586–594.

[102] Q. Wang, Y. Dou, X. W. Liu, Q. Lv, and S. J. Li, Multi-view clustering with extreme learning machine, *Neurocomputing*, vol. 214, pp. 483–494, 2016.

[103] H. C. Gao, F. P. Nie, X. L. Li, and H. Huang, Multi-view subspace clustering, in *Proc. 2015 IEEE Int. Conf. Computer Vision*, Santiago, Chile, 2016, pp. 4238–4246.

[104] J. L. Xu, J. W. Han, and F. P. Nie, Discriminatively embedded k-means for multi-view clustering, in *Proc. 2016 IEEE Conf. Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 2016, pp. 5356–5364.

[105] J. L. Xu, J. W. Han, F. P. Nie, and X. L. Li, Re-weighted discriminatively embedded k-means for multi-view clustering, *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 3016–3027, 2017.

[106] Y. Wang, X. M. Lin, L. Wu, W. J. Zhang, and Q. Zhang, Exploiting correlation consensus: Towards subspace clustering for multi-modal data, in *Proc. 22nd ACM Int. Conf. Multimedia*, Orlando, FL, USA, 2014, pp. 981–984.

[107] Y. Wang, X. M. Lin, L. Wu, W. J. Zhang, Q. Zhang, and X. D. Huang, Robust subspace clustering for multi-view data by exploiting correlation consensus, *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3939–3949, 2015.

[108] Y. B. Fan, R. He, and B. G. Hu, Global and local consistent multi-view subspace clustering, in *Proc. 3rd IAPR Asian Conf. Pattern Recognition*, Kuala Lumpur, Malaysia, 2015, pp. 564–568.

[109] X. Zhang, D. Phung, S. Venkatesh, D. S. Pham, and W. Q. Liu, Multi-view subspace clustering for face images, in *Proc.2015 Int. Conf. Digital Image Computing*: *Techniques and Applications*, Adelaide, Australia, 2015, pp. 1–7.

[110] D. Wang, R. He, L. Wang, and T. N. Tan, Adaptive multi-view clustering via cross trace lasso, in *Proc. 3rd IAPR Asian Conf. Pattern Recognition*, Kuala Lumpur, Malaysia, 2015, pp. 559–563.

[111] D. Wang, Q. Y. Yin, R. He, L. Wang, and T. N. Tan, Multi-view clustering via structured low-rank representation, in *Proc. 24th ACM Int. on Conf. Information and Knowledge Management*, Melbourne, Australia, 2015, pp. 1911–1914.

[112] X. H. Liu, S. W. Ji, W. Glänzel, and B. De Moor, Multiview partitioning via tensor methods, *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 5, pp. 1056–1069, 2013.

[113] C. Q. Zhang, H. Z. Fu, S. Liu, G. C. Liu, and X. C. Cao, Low-rank tensor constrained multiview subspace clustering, in *Proc. 2015 IEEE Int. Conf. Computer Vision*, Santiago, Chile, 2015, pp. 1582–1590.

[114] Y. Xie, D. C. Tao, W. S. Zhang, and L. Zhang, Multi-view subspace clustering via relaxed L_1-norm of tensor multi-rank, arXiv preprint arXiv: 1610.07126, 2016.

[115] M. Yin, J. B. Gao, S. L. Xie, and Y. Guo, Low-rank multi-view clustering in third-order tensor space, arXiv preprint arXiv: 1608.08336, 2016.

[116] Y. Cui, X. Z. Fern, and J. G. Dy, Non-redundant multi-view clustering via orthogonalization, in *Proc. 7th IEEE Int. Conf. Data Mining*, Omaha, NE, USA, 2007, pp. 133–142.

[117] Y. Cui, X. Z. Fern, and J. G. Dy, Learning multiple nonredundant clusterings, *ACM Trans. Knowl. Discov.Data*, vol. 4, no. 3, p. 15, 2010.

[118] S. Günnemann, I. Färber, and T. Seidl, Multi-view clustering using mixture models in subspace projections, in *Proc. 18th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, Beijing, China, 2012, pp. 132–140.

[119] E. Muller, S. Günnemann, I. Farber, and T. Seidl, Discovering multiple clustering solutions: Grouping objects in different views of the data, in *Proc. 10th Int. Conf. Data Mining*, Sydney, Australia, 2012, p. 1220.

[120] S. Günnemann, I. Färber, M. Rüdiger, and T. Seidl, SMVC: Semi-supervised multi-view clustering in subspace projections, in *Proc. 20th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, New York, NY, USA, 2014, pp. 253–262.

[121] Q. Y. Yin, S. Wu, R. He, and L. Wang, Multi-view clustering via pairwise sparse subspace representation, *Neurocomputing*, vol. 156, pp. 12–21, 2015.

[122] X. C. Cao, C. Q, Zhang, C. J. Zhou, H. Z. Fu, and H. Foroosh, Constrained multi-view video face clustering, *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4381–4393, 2015.

[123] Q. Y. Yin, S. Wu, and L. Wang, Incomplete multi-view clustering via subspace learning, in *Proc. 24th ACM Int. on Conf. Information and Knowledge Management*, Melbourne, Australia, 2015, pp. 383–392.

[124] Q. Y. Yin, S. Wu, and L. Wang, Unified subspace learning for incomplete and unlabeled multi-view data, *Pattern Recogn.*, vol. 67, pp. 313–327, 2017.

[125] C. Xu, D. C. Tao, and C. Xu, Multi-view learning with incomplete views, *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5812–5825, 2015.

[126] D. D. Lee and H. S. Seung, Learning the parts of objects by non-negative matrix factorization, *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[127] C. Ding, X. F. He, and H. D. Simon, Nonnegative lagrangian relaxation of k-means and spectral clustering, in *Proc. 16th European Conf. Machine Learning*, Porto, Portugal, 2005, pp. 530–538.

[128] V. P. Pauca, J. Piper, and R. J. Plemmons, Nonnegative matrix factorization for spectral data analysis, *Linear Algebra Appl.*, vol. 416, no. 1, pp. 29–47, 2006.

[129] D. Cai, X. F. He, J. W. Han, and T. S. Huang, Graph regularized nonnegative matrix factorization for data representation, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, 2011.

[130] C. H. Q. Ding, T. Li, and M. I. Jordan, Convex and semi-nonnegative matrix factorizations, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 45–55, 2010.

[131] A. Cichocki and R. Zdunek, Multilayer nonnegative matrix factorisation, *Electron. Lett.*, vol. 42, no. 16, pp. 947–948, 2006.

[132] T. Li and C. Ding, Non-negative matrix factorizations for clustering: A survey, in *Data Clustering*: *Algorithms and Applications*, C. C. Aggarwal and C. K. Reddy, eds. Boca Raton, FL, USA: Chapman & Hall/CRC, 2014.

[133] D. Greene and P. Cunningham, A matrix factorization approach for integrating multiple data views, in *Proc. European Conf. Machine Learning and Knowledge Discovery in Databases*: *Part I*, Bled, Slovenia, 2009, pp. 423–438.

[134] J. L. Liu, C. Wang, J. Gao, and J. W. Han, Multi-view clustering via joint nonnegative matrix factorization, in *Proc. 2013 SIAM Int. Conf. Data Mining*, Austin, TX, USA, 2013, pp. 252–260.

[135] Y. Jiang, J. Liu, Z. C. Li, and H. Q. Lu, Semi-supervised unified latent factor learning with multi-view data, *Mach. Vision Appl.*, vol. 25, no. 7, pp. 1635–1645, 2014.

[136] M. J. He, Y. Yang, and H. J. Wang, Learning latent features for multi-view clustering based on NMF, in *Proc. Int. Joint Conf. Rough Sets*, Santiago de Chile, Chile, 2016, pp. 459–469.

[137] W. Y. Chang, C. P. Wei, and Y. C. F. Wang, Multi-view nonnegative matrix factorization for clothing image characterization, in *Proc. 22nd Int. Conf. Pattern Recognition*, Stockholm, Sweden, 2014, pp. 1272–1277.

[138] W. H. Ou, S. J. Yu, G. Li, J. Lu, K. S. Zhang, and G. Xie, Multi-view non-negative matrix factorization by patch alignment framework with view consistency, *Neurocomputing*, vol. 204, pp. 116–124, 2016.

[139] C. Xu, D. C. Tao, and C. Xu, Multi-view self-paced learning for clustering, in *Proc. 24th Int. Conf. Artificial Intelligence*, Buenos Aires, Argentina, 2015, pp. 3974–3980.

[140] X. Cai, F. P. Nie, and H. Huang, Multi-view k-means clustering on big data, in *Proc. 23rd Int. joint Conf. Artificial Intelligence*, Beijing, China, 2013, pp. 2598–2604.

[141] M. J. Qian and C. X. Zhai, Unsupervised feature selection for multi-view clustering on text-image web news data, in *Proc. 23rd ACM Int. Conf. Conf. Information and Knowledge Management*, Shanghai, China, 2014, pp. 1963–1966.

[142] H. D. Zhao, Z. M. Ding, and Y. Fu, Multi-view clustering via deep matrix factorization, in *Proc. 31st AAAI Conf. Artificial Intelligence*, San Francisco, CA, USA, 2017, pp. 2921–2927.

[143] H. Wang, F. P. Nie, and H. Huang, Multi-view clustering and feature learning via structured sparsity, in *Proc. 30th Int. Conf. Machine Learning*, Atlanta, GA, USA, 2013, pp. 352–360.

[144] H. F. Liu, H. Y. Mao, and Y. Fu, Robust multi-view feature selection, in *Proc. 16th Int. Conf. Data Mining*, Barcelona, Spain, 2016, pp. 281–290.

[145] X. L. Gong, F. W. Wang, and L. P. Huang, Weighted NMF-based multiple sparse views clustering for web items, in *Proc. 21st Pacific-Asia Conf. Advances in Knowledge Discovery and Data Mining*, Jeju, South Korea, 2017, pp. 416–428.

[146] D. Hidru and A. Goldenberg, EquiNMF: Graph regularized multiview nonnegative matrix factorization, arXiv preprint arXiv: 1409.4018, 2014.

[147] Z. F. Wang, X. W. Kong, H. Y. Fu, M. Li, and Y. J. Zhang, Feature extraction via multi-view non-negative matrix factorization with local graph regularization, in *Proc. 2015 IEEE Int. Conf. Image Proc.*, Quebec City, Canada, 2015, pp. 3500–3504.

[148] Z. Y. Guan, L. J. Zhang, J. Y. Peng, and J. P. Fan, Multi-view concept learning for data representation, *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 11, pp. 3016–3028, 2015.

[149] X. C. Zhang, L. Zhao, L. L. Zong, X. Y. Liu, and H. Yu, Multi-view clustering via multi-manifold regularized nonnegative matrix factorization, in *Proc. 2014 IEEE Int. Conf. Data Mining*, Shenzhen, China, 2014, pp. 1103–1108.

[150] L. L. Zong, X. C. Zhang, L. Zhao, H. Yu, and Q. L. Zhao, Multi-view clustering via multi-manifold regularized nonnegative matrix factorization, *Neural Netw.*, vol. 88, pp. 74–89, 2017.

[151] D. Kuang, C. Ding, and H. Park, Symmetric nonnegative matrix factorization for graph clustering, in *Proc. 2012 SIAM Int. Conf. Data Mining*, Anaheim, CA, USA, 2012, pp. 106–117.

[152] X. C. Zhang, Z. X. Wang, L. L. Zong, and H. Yu, Multi-view clustering via graph regularized symmetric nonnegative matrix factorization, in *Proc. 2016 IEEE Int. Conf. Cloud Computing and Big Data Analysis*, Chengdu, China, 2016, pp. 109–114.

[153] W. Xu and Y. H. Gong, Document clustering by concept factorization, in *Proc. 27th Annu. Int. ACM SIGIR Conf. Research and Development in Information Retrieval*, Sheffield, UK, 2004, pp. 202–209.

[154] H. Wang, Y. Yang, and T. R. Li, Multi-view clustering via concept factorization with local manifold regularization, in *Proc. 16th Int. Conf. Data Mining*, Barcelona, Spain, 2016, pp. 1245–1250.

[155] K. Zhan, J. H. Shi, J. Wang, and F. Tian, Graph-regularized concept factorization for multi-view document clustering, *Journal of Visual Communication and Image Representation*, vol. 48, pp. 411–418, 2017.

[156] W. X. Shao, L. F. He, and P. S. Yu, Multiple incomplete views clustering via weighted nonnegative matrix factorization with $L2,1$ regularization, in *European Conf. Machine Learning and Knowledge Discovery in Databases*, Porto, Portugal, 2015, pp. 318–334.

[157] W. X. Shao, L. F. He, C. T. Lu, and P. S. Yu, Online multi-view clustering with incomplete views, arXiv preprint arXiv: 1611.00481, 2016.

[158] W. X. Shao, L. F. He, and P. S. Yu, Clustering on multi-source incomplete data via tensor modeling and factorization, in *Proc. 19th Pacific-Asia Conf. Advances in Knowledge Discovery and Data Mining*, Ho Chi Minh City, Vietnam, 2015, pp. 485–497.

[159] S. Y. Li, Y. Jiang, and Z. H. Zhou, Partial multi-view clustering, in *Proc. 28th AAAI Conf. Artificial Intelligence*, Québec City, Canada, 2014, pp. 1968–1974.

[160] B. Qian, X. B. Shen, Y. Y. Gu, Z. M. Tang, and Y. H. Ding, Double constrained NMF for partial multi-view clustering, in *Proc. 2016 Int. Conf. Digital Image Computing: Techniques and Applications*, Gold Coast, Australia, 2016, pp. 1–7.

[161] N. Rai, S. Negi, S. Chaudhury, and O. Deshmukh, Partial multi-view clustering using graph regularized NMF, in *Proc. 23rd Int. Conf. Pattern Recognition*, Cancun, Mexico, 2016, pp. 2192–2197.

[162] X. C. Zhang, L. L. Zong, X. Y. Liu, and H. Yu, Constrained NMF-based multi-view clustering on unmapped data, in *Proc. 29th AAAI Conf. Artificial Intelligence*, Austin, TX, USA, 2015, pp. 3174–3180.

[163] R. Caruana, Multitask learning, *Mach. Learn.*, vol. 28, no. 1, pp. 41–75, 1997.

[164] Q. Q. Gu and J. Zhou, Learning the shared subspace for multi-task clustering and transductive transfer classification, in *Proc. 9th IEEE Int. Conf. Data Mining*, Miami, FL, USA, 2009, pp. 159–168.

[165] Z. H. Zhang and J. Zhou, Multi-task clustering via domain adaptation, *Pattern Recogn.*, vol. 45, no. 1, pp. 465–473, 2012.

[166] S. N. Xie, H. T. Lu, and Y. C. He, Multi-task co-clustering via nonnegative matrix factorization, in *Proc. 21st Int. Conf. Pattern Recognition*, Tsukuba, Japan, 2012, pp. 2954–2958.

[167] S. Al-Stouhi and C. K. Reddy, Multi-task clustering using constrained symmetric non-negative matrix factorization, in *Proc. 2014 SIAM Int. Conf. Data Mining*, Philadelphia, PA, USA, 2014, pp. 785–793.

[168] X. Wang, B. Y. Qian, J. P. Ye, and I. Davidson, Multi-objective multi-view spectral clustering via Pareto optimization, in *Proc. 2013 SIAM Int. Conf. Data Mining*, Austin, TX, USA, 2013, pp. 234–242.

[169] A. Wahid, X. Y. Gao, and P. Andreae, Multi-view clustering of web documents using multi-objective genetic algorithm, in *Proc. 2016 IEEE Congress on Evolutionary Computation*, Beijing, China, 2014, pp. 2625–2632.

[170] A. Wahid, X. Y. Gao, and P. Andreae, Multi-objective multi-view clustering ensemble based on evolutionary approach, in *Proc. 2015 IEEE Congress on Evolutionary Computation*, Sendai, Japan, 2015, pp. 1696–1703.

[171] X. T. Zhang, X. C. Zhang, H. Liu, and X. Y. Liu, Multi-task clustering through instances transfer, *Neurocomputing*, vol. 251, pp. 145–155, 2017.

[172] H. Shi, Y. Li, Y. H. Han, and Q. H. Hu, Cluster structure preserving unsupervised feature selection for multi-view tasks, *Neurocomputing*, vol. 175, pp. 686–697, 2016.

[173] X. C. Zhang, X. T. Zhang, and H. Liu, Multi-task multi-view clustering for non-negative data, in *Proc. 24th Int. Conf. Artificial Intelligence*, Buenos Aires, Argentina, 2015, pp. 4055–4061.

[174] X. T. Zhang, X. C. Zhang, H. Liu, and X. Y. Liu, Multi-task multi-view clustering, *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 12, pp. 3324–3338, 2016.

[175] R. Bekkerman and J. Jeon, Multi-modal clustering for multimedia collections, in *Proc. 2007 IEEE Conf. Computer Vision and Pattern Recognition*, Minneapolis, MN, USA, 2007, pp. 1–8.

[176] X. J. Xie and S. L. Sun, Multi-view clustering ensembles, in *Proc. 2013 Int. Conf. Machine Learning and Cybernetics*, Tianjin, China, 2013, pp. 51–56.

[177] X. J. Chen, X. F. Xu, J. Z. Huang, and Y. M. Ye, Tw-k-means: Automated two-level variable weighting clustering algorithm for multiview data, *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 4, pp. 932–944, 2013.

[178] Y. M. Xu, C. D. Wang, and J. H. Lai, Weighted multi-view clustering with feature selection, *Pattern Recogn.*, vol. 53, pp. 25–35, 2016.

[179] B. Jiang, F. Y. Qiu, L. P. Wang, and Z. J. Zhang, Bi-level weighted multi-view clustering via hybrid particle swarm optimization, *Inf. Process. Manage.*, vol. 52, no. 3, pp. 387–398, 2016.

[180] B. Jiang, F. Y. Qiu, and L. P. Wang, Multi-view clustering via simultaneous weighting on views and features, *Appl. Soft Comput.*, vol. 47, pp. 304–315, 2016.

[181] G. Cleuziou, M. Exbrayat, L. Martin, and J. H. Sublemontier, CoFKM: A centralized method for multiple-view clustering, in *Proc. 9th IEEE Int. Conf. Data Mining*, Miami, FL, USA, 2009, pp. 752–757.

[182] F. de A. T. de Carvalho, F. M. de Melo, and Y. Lechevallier, A multi-view relational fuzzy c-medoid vectors clustering algorithm, *Neurocomputing*, vol. 163, pp. 115–123, 2015.

[183] Y. Z. Jiang, F. L. Chung, S. T. Wang, Z. H. Deng, J. Wang, and P. J. Qian, Collaborative fuzzy clustering from multiple weighted views. *IEEE Trans. Cybern.*, vol. 45, no. 4, pp. 688–701, 2015.

[184] Y. T. Wang and L. H. Chen, Multi-view fuzzy clustering with minimax optimization for effective clustering of data from multiple sources, *Expert Syst. Appl.*, vol. 72, pp. 457–466, 2017.

[185] S. F. Hussain, G. Bisson, and C. Grimal, An improved co-similarity measure for document clustering, in *Proc. 9th Int. Conf. Machine Learning and Applications*, Washington, DC, USA, 2010, pp. 190–197.

Yan Yang received the BS and MS, degrees from Huazhong University of Science and Technology, Wuhan, China, in 1984 and 1987, respectively. She received the PhD degree from Southwest Jiaotong University, Chengdu, China, in 2007. From 2002 to 2003 and 2004 to 2005, she was a visiting scholar with the University of Waterloo, Waterloo, Canada. She is currently a professor and vice dean with the School of Information Science and Technology, Southwest Jiaotong University, Chengdu, China. Her current research interests include multi-view learning, big data analysis and mining, ensemble learning, semi-supervised learning, and cloud computing.

Hao Wang received the BEng degree from Nanyang Institute of Technology, Nanyang, China, in 2014. He is currently pursuing the PhD degree in Southwest Jiaotong University, Chengdu, China. His current research interests include data mining, multi-view learning, natural language processing, and lifelong machine learning.