# Semi-Supervised Ensemble Clustering Based on Selected Constraint Projection

Zhiwen Yu, *Senior Member, IEEE*, Peinan Luo, Jiming Liu, *Fellow, IEEE*, Hau-San Wong, Jane You, Guoqiang Han, and Jun Zhang, *Fellow, IEEE*

**Abstract**—Traditional cluster ensemble approaches have several limitations. (1) Few make use of prior knowledge provided by experts. (2) It is difficult to achieve good performance in high-dimensional datasets. (3) All of the weight values of the ensemble members are equal, which ignores different contributions from different ensemble members. (4) Not all pairwise constraints contribute to the final result. In the face of this situation, we propose double weighting semi-supervised ensemble clustering based on selected constraint projection(DCECP) which applies constraint weighting and ensemble member weighting to address these limitations. Specifically, DCECP first adopts the random subspace technique in combination with the constraint projection procedure to handle high-dimensional datasets. Second, it treats prior knowledge of experts as pairwise constraints, and assigns different subsets of pairwise constraints to different ensemble members. An adaptive ensemble member weighting process is designed to associate different weight values with different ensemble members. Third, the weighted normalized cut algorithm is adopted to summarize clustering solutions and generate the final result. Finally, nonparametric statistical tests are used to compare multiple algorithms on real-world datasets. Our experiments on 15 high-dimensional datasets show that DCECP performs better than most clustering algorithms.

**Index Terms**—Cluster ensemble, semi-supervised clustering, projection, pairwise constraint

## 1 Introduction

Cluster ensemble study [1], [2], [3], [4], [5] constitutes an important branch of ensemble learning. When compared with traditional single clustering approaches, they are able to integrate multiple clustering results into a unified result which is more accurate and stable. The cluster ensemble approach can be divided into two stages: ensemble generation and consensus function. The objective of the ensemble generation stage is to obtain diverse ensemble members, while the objective of the consensus function is to generate a unified clustering solution as accurately as possible. Recently, cluster ensembles have gained more attention, due to their useful applications in different research areas [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], such as data mining [6], [7], [8], image processing [9], [10], bioinformatics [11], [12], [13], internet security [14], [15], and so on. For example, Yu et al. [6], [7] proposed an adaptive cluster ensemble approach for noisy data mining. Yang et al. [8] designed the weighted clustering ensemble framework for temporal data mining. Tademir et al. [9] applied the spectral clustering ensemble to handle high spatial resolution remote sensing images. Tsai et al. [10] proposed a self-organizing map-based cluster ensemble framework for multimodal image clustering. Yu et al. [11], [12] studied how to apply cluster ensemble approaches to handle biomolecular data. Hu et al. [13] explored how to use a cluster ensemble approach for microarray gene cluster identification. Zhuang et al. [14] designed the principled cluster ensemble framework for automatic malware and phishing website detection. Li et al. [15] investigated how to apply the cluster ensemble approach for culprit actor detection in social media networks. In general, cluster ensemble approaches have been successfully applied to various areas, and have generated good results in different kinds of datasets.

However, several limitations exist for conventional cluster ensemble approaches. (1) Few consider knowledge provided by experts in specific domains. (2) Few take into consideration how to handle high-dimensional data. (3) Most treat each ensemble member equally even though different ensemble members make different contributions to the fina result. (4) While most pairwise constraints contribute to the final result, some may negatively affect the outcome. To address these limitations, we first propose the semi-supervised clustering ensemble based on selected constraint projection (CESCP) framework, with the following properties. (1) CESCP makes full use of pairwise constraints, which are prior knowledge provided by experts. Different weight values are assigned to different ensemble members according to different subsets of pairwise constraints, which


- *Z. Yu, P. Luo, G. Han, and J. Zhang are with the Guangdong Provincial Key Lab of Computational Intelligence and Cyberspace Information, School of Computer Science and Engineering, South China University of Technology, Guangzhou Shi, Guangdong Sheng 510630, China. E-mail: {zhwyu, csgqhan, csjun}@scut.edu.cn, 411139073@qq.com.*
- *J. Liu is with the Department of Computer Science, Hong Kong Baptist University, Kowloon Tong, Hong Kong. E-mail: jiming@comp.hkbu.edu.hk.*
- *H.-S. Wong is with the Department of Computer Science, City University of Hong Kong, Kowloon Tong, Hong Kong. E-mail: cshswong@cityu.edu.hk.*
- *J. You is with the Department of Computing, Hong Kong Polytechnic University, Hung Hom, Hong Kong. E-mail: csyjia@comp.polyu.edu.hk.*


*Manuscript received 18 Oct. 2017; revised 29 Jan. 2018; accepted 19 Mar. 2018. Date of publication 23 Mar. 2018; date of current version 5 Nov. 2018. (Corresponding author: Zhiwen Yu.)*
*Recommended for acceptance by J. M. Phillips.*
*For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.*
*Digital Object Identifier no. 10.1109/TKDE.2018.2818729*

increase the diversity of the ensemble. (2) CESCP not only adopts the random subspace technique to handle high-dimensional data, but also uses the constraint projection technique to map high-dimensional data into a low-dimensional space. On the basis of CESCP, we further propose the double weighting semi-supervised ensemble clustering approach (DCECP) which includes two weighting process: the constraint weighting process and the adaptive ensemble member weighting process. DCECP has the following characteristics: (1) The adaptive ensemble member weighting process is used to create weights for ensemble members using competition. (2) The weighted normalized cut algorithm is adopted to serve as the consensus function, which is used to partition the consensus matrix generated by different clustering results and obtain a better partition. In addition, nonparametric statistical tests are used to compare the proposed DCECP approach and traditional cluster ensemble approaches on multiple datasets. We evaluate DCECP on 15 high-dimensional datasets. The results illustrate that DCECP works well for these datasets, and outperforms most state-of-the-art cluster ensemble approaches.

The contributions of this paper are as follows: (1) a new semi-supervised clustering ensemble approach based on selected constraint projection is proposed, which not only use the constraint subset selection process to make full use of prior knowledge provided by experts, but also adopts the constraint projection technique to map high-dimensional data into a low-dimensional space. (2) The double weighting semi-supervised ensemble clustering based on selected constraint projection (DCECP) framework is designed, which adopts the adaptive ensemble member weighting process to create weights for ensemble members using competition, and increases the diversity of the ensemble. (3) The random subspace method is combined with constraint projection technique to handle high-dimensional data.

The remainder of this paper is organized as follows. Section 2 introduces related works on semi-supervised clustering approaches and semi-supervised cluster ensemble approaches. Section 3 presents CESCP and DCECP. Section 4 describes the adaptive ensemble member weighting process. Section 5 gives a theoretical analysis of the proposed approach. Section 6 evaluates the performance of the proposed DCECP approach using real high-dimensional datasets. Section 7 gives conclusions and possibilities for future works.

## 2 Related Works

Semi-supervised clustering is a hot topic in machine learning research [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30]. Unlike traditional methods, semi-supervised methods convert experts knowledge into prior information such as label or constraint, and use these to enhance the clustering performance. Recently, a number of semi-supervised methods have been proposed, such as semi-supervised maximum margin clustering [18], semi-supervised kernel mean shift clustering [19], semi-supervised subspace clustering [20], semi-supervised linear discriminant clustering [21], semi-supervised co-clustering [24], semi-supervised fuzzy clustering [26], semi-supervised nonnegative matrix factorization [27], and so on. Some researchers incorporate other techniques to assist clustering. For example, Wang et al. [22] studied the constraint neighborhood projection technique for semi-supervised clustering. Xiong et al. [23] investigated the active learning of constraints in the semi-supervised clustering process. Chang et al. integrated discriminative random fields into a semi-supervised clustering algorithm. Semi-supervised clustering algorithms have been successfully applied to different areas, such as network community detection [28], gene expression data clustering [29], face and hair region segmentation [30], and so on. As a whole, semi-supervised methods take advantage of pairwise constraints generated by the knowledge of experts to enhance quality of the clustering process. However, performance of semi-supervised methods are sensitive to the values of the parameters and the order of pairwise constraints considered, which make them less stable and robust.

To address the limitations of single clustering approaches, cluster ensemble approaches are proposed [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46], which are not sensitive to parameter values. Cluster ensemble approaches can incorporate multiple clustering solutions into an ensemble framework and generate more stable and robust results. Most of them can be divided into four categories. The approaches [31], [32], [33], [34] in the first category focus on how to generate clustering solutions that are as diverse as possible, including the random combination of data transformation operators [31], the adoption of different clustering algorithms [32], the hybrid sampling method [33], the perturbation technique [34], and so on. The approaches [35], [36], [37], [38] in the second category study how to design a consensus function that can generate the final results as accurately as possible; these approaches include the linked based consensus function [35], the locally weighted consensus function [36], the graph based consensus function [37], the model based consensus function [38], and so on. The third type of approaches [39], [40], [41], [42] investigate properties of ensemble members, for instance, the efficiency of the cluster ensemble [39], the stability of the cluster ensemble [40], [41], the diversity of the cluster ensemble [42], and so on. The approaches [43], [44], [45], [46] in the fourth category apply cluster ensemble methods to different areas, such as image segmentation [43], region segmentation [44], geophysical process detection [45], time series clustering [46], and so on. In general, cluster ensemble approaches have been successfully used in different application areas. However, most of them do not take prior information into consideration to further increase the accuracy.

For this reason, our work focuses on semi-supervised ensemble approaches, which consider the advantages of semi-supervised algorithms and cluster ensemble approaches. Semi-supervised clustering ensemble approaches not only consider prior knowledge provided by experts, but also integrate multiple clustering solutions into a unified clustering solution to improve the quality of the final result. Recently, a number of researchers studied how to develop the ensemble framework with semi-supervised methods. For example, Yu et al. [47] designed the adaptive clustering ensemble framework, which uses an adaptive procedure to search for optimal combination of subspaces and generate more accurate results. They also proposed an incremental

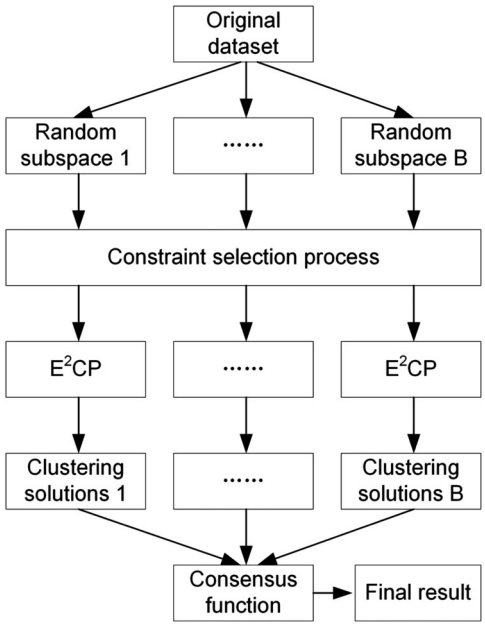



Fig. 1. Overview of the semi-supervised clustering ensemble based on selected constraint projection.

framework which makes use of global fitness and local cost functions to incrementally remove redundant clustering solutions [48]. Wang et al. designed a semi-supervised clustering ensemble approach based on the binary similarity matrix. Yang et al. [59] proposed a bi-weighted ensemble clustering approach. It gives each partition a two-layer weight based on clustering quality and dominance that produces the corresponding clusters. Some works explored properties of ensemble approaches like convergence [50], while some studies have investigated applications for ensemble clustering approaches. For example, Yu et al. [51] applied the semi-supervised ensemble clustering to gene expression data. In general, conventional semi-supervised clustering approaches take the complete set of pairwise constraints into account.

However, not all pairwise constraints contribute to the final result. Some pairwise constraints may not be useful, and some will even affect the result negatively. For this situation, we consider a selection of pairwise constraints, and propose the ensemble clustering framework based on selected constraint projection.

## 3 Semi-Supervised Clustering Ensemble Based on Selected Constraint Projection

Given a high-dimensional dataset $P = \{p_1, p_2, \ldots, p_n\}$, with each feature vector $p_i$ ($i \in \{1, \ldots, n\}$) containing $m$ attributes, semi-supervised clustering ensemble approaches divide the dataset $P$ into $k$ groups so as to minimize the within-cluster sum of squares, and satisfy as many of the pairwise constraints as possible at the same time. Fig. 1 illustrates the framework for semi-supervised clustering ensemble based on selected constraint projection. Specifically, CESCP first adopts the random subspace technique to generate $B$ random subspaces $A_1, A_2, \ldots, A_B$. Then, weighted constraint selection and constraint projection are applied to each random subspace with the clustering method $E^2CP$ [12] to generate a set of clustering results $I^1$, $I^2$, ..., $I^B$. Next, a consensus matrix is constructed using these clustering results. Finally, the normalized cut (Ncut, [52]) is used as a consensus function to obtain the final result.

CESCP uses the random subspace technique to generate a set of random subspaces $A = \{A_1, \ldots A_B\}$ in the first step. Specifically, we have a sampling rate $\tau$ which makes the size of the random subspace $\tau * m$. Then, the subspace attributes are selected by $CESCP$ one by one, whose index is determined as follows:

$$j = \lfloor 1 + \varsigma(m-1) \rfloor, \tag{1}$$

where $j$ is the index of the selected attribute, and $\varsigma$ is a uniform random variable between 0 and 1. Each selected index will be marked to avoid repetition, and the duplicate attributes will be ignored. The above operation will continue until $\tau m$ attributes are selected. The new subspace is constructed by these selected attributes, and it will also be marked to avoid the duplicate subspace problem in the set $A$. Finally, CESCP will generate a set of random subspaces $A_1$, $A_2$, ..., $A_B$ by repeating the above process $B$ times. The advantage of the random subspace technique is that it provides multiple ways to explore the underlying structure of the data in a low-dimensional space.

In the second step, weighted constraint selection and constraint projection are performed in every random subspace $A_i$ ($i \in \{1, \ldots, B\}$) generated in the first step. Since different subspaces of the original data will have different inner structures, we believe that these subspaces will have their own different important constraint sets, which helps the clustering process. From this point, we have three assumptions:

(1) A cannot-link constraint that has a small distance will be more important.
(2) A must-link constraint that has a large distance will be more important.
(3) A pairwise constraint will affect the importance of other constraints.

For a specific subspace, we apply a progressive strategy for constraint selection and weighting. Suppose that the original constraint set is $C$, we will give every constraint $c_i = (x_{i1}, x_{i2})$, where $c_i \in C$, an initial weight $w_i \in (0, 1)$ based on the euclidean distance $dis(c_i)$ of the two data samples $(x_{i1}, x_{i2})$ linked by $c_i$. Inspired by the sigmoid function, the following equation shows how the weight is initialized

$$w_i = \begin{cases} \frac{2-2e^{-dis(c_i)}}{2+2e^{-dis(c_i)}}, & \text{if } c_i \text{ is a must link} \\ \frac{4}{2+2e^{dis(c_i)}}, & \text{if } c_i \text{ is a cannot link.} \end{cases} \tag{2}$$

Initially, our selected constraint set $S$ is empty and the candidate constraint set is $C$. The weighted constraint selection process has the following steps:

(1) The constraint $c_{choose}$ which has the highest weight is selected, added to $S$, and removed from $C$.
(2) The constraints in $S$ are adopted to perform semi-supervised clustering algorithm $E^2CP$ and obtain result $R$.
(3) The within-cluster distance and the number of satisfied constraints are used to evaluate the performance

of the clustering result R as follows:

$$\Delta(R) = \frac{1}{2}\sum_{p_i \in P}\sum_{h=1}^{k}\theta(y_i = h)d(p_i, \mu_h) + \sum_{p_i,p_j \in M} w_{ij}^{M}\theta(y_i \neq y_j) + \sum_{p_i,p_j \in N} w_{ij}^{N}\theta(y_i = y_j) \tag{3}$$

$$\mu_h = \frac{\sum_{i=1}^{h}\theta(y_i = h)p_i}{\sum_{i=1}^{h}\theta(y_i = h),} \tag{4}$$

where $d(p_i, \mu_h)$ denotes the euclidean distance between the feature vectors $p_i$ and $\mu_h$. $\theta$ denotes the indicator function, $\theta(true) = 1$ and $\theta(false) = 0$, $M$ denotes the must-link constraint set and $N$ denotes the cannot-link constraint set. The objective of the cost function is to optimize the squared distance from the feature vectors to their centers and satisfy most of the constraints. If $\Delta(R)$ decreases, according to assumption (3), the added constraints will affect the weight of the other constraints. For each constraint $c_i$ in the candidate constraint set $C$ which is satisfied in the clustering result $R$, its weight will be updated using the following rule

$$w_i = w_i * exp(-1). \tag{5}$$

Next, we sort the constraints in descending order. If the clustering performance decreases, constraint $c_{choose}$ will be removed from $S$.

(4) If $C$ is not empty, go to step (1) or end the constraint weighting selection process.

Constraint projection is a typical data transformation technique to make use of constraint information. As different constraints have different contributions to the clustering procedure, we used a method called weighted constraint projection (WCP) based on constraint projection. Our selected constraint set $S$ has must-link (ML) and cannot-link (CL) constraints, and WCP seeks a set of projection vectors $P = [p_1, p_2, \ldots, p_d]$ which make the pairwise constraints in ML and CL preserve their states in the transformed space $z_i = P^T x_i$. Ideally, data objects involved in ML will be close together, while data objects involved in CL will keep away from each other in the new space. For our problem, we would assign different importance to each constraint. We perform weighted projection through the maximization of the following function $J(P)$ with respect to $P^T P = I$

$$J(P) = \left(\frac{1}{2n_c}\sum_{(x_i,x_j)\in C} w_i \|P^T(x_i - x_j)\|^2\right) \tag{6}$$

$$-\left(\frac{1}{2n_m}\sum_{(x_i,x_j)\in M} w_i \|P^T(x_i - x_j)\|^2\right), \tag{7}$$

where $n_c$ is the size of the cannot-link set and $n_m$ is the size of the must-link set.

Obviously, the problem above can be transformed into a typical eigensystem so that it can be solved efficiently by conventional optimization methods. For convenience, we define $S_C$ and $S_M$ as follows:

$$S_C = \left(\frac{1}{2n_c}\sum_{(x_i,x_j)\in C} w_i \|P^T(x_i - x_j)\|^2\right) \tag{8}$$

$$S_M = \left(\frac{1}{2n_m}\sum_{(x_i,x_j)\in M} w_i \|P^T(x_i - x_j)\|^2\right). \tag{9}$$

From these equations, to maximize $J(P)$, $S_C$ should be maximized while $S_M$ should be minimized. There exists an analytical solution to the problem of finding the projection matrix P

$$S_C = \left(\frac{1}{2n_c}\sum_{(x_i,x_j)\in C} w_i \|P^T(x_i - x_j)\|^2\right) \tag{10}$$

$$= \frac{1}{2n_c}\sum_{(x_i,x_j)\in C} w_i \sum_l P_l^T(x_i - x_j)(x_i - x_j)^T P_l \tag{11}$$

$$= \frac{1}{2n_c}\sum_l P_l^T\left(\sum_{(x_i,x_j)\in C} w_i(x_i - x_j)(x_i - x_j)^T\right)P_l \tag{12}$$

$$= \frac{1}{2n_c}\sum_l P_l^T\left(CC^T\right)P_l \tag{13}$$

$$= \frac{1}{2n_c}\sum_l P_l^T \Lambda_c P_l. \tag{14}$$

The same operation on $S_M$ is

$$S_M = \frac{1}{2n_m}\sum_l P_l^T \Lambda_m P_l. \tag{15}$$

According to (6) and (7), the objective function now becomes

$$J(P) = \frac{1}{2n_c}\sum_l P_l^T \Lambda_c P_l - \frac{1}{2n_m}\sum_l P_l^T \Lambda_m P_l \tag{16}$$

$$= \vartheta\sum_l P_l^T(\Lambda_c - \Lambda_m)P_l \propto \sum_l P_l^T(\Lambda)P_l. \tag{17}$$

By using the Lagrange multiplier technique, the Lagrangian can be written as

$$L_{P_1,\ldots,Pk} = J(P_1, \ldots, P_k) - \sum_{l=1}^{k}\delta_l(P_l^T P_l - 1). \tag{18}$$

Taking the partial derivative of $L_{P_1,\ldots,Pk}$ with respect to each $P_l$ and make it zero, we obtain

$$\frac{\partial L}{\partial P_l} = 2\Lambda P_l - 2\delta_l P_l = 0, \forall l = 1, \ldots, k \tag{19}$$

$$\Longrightarrow \Lambda P_l = \delta_l P_l, \forall l = 1, \ldots, k. \tag{20}$$

From Equation (20), we can see that $P_l$ is an eigenvector of $\Lambda$, and $\delta_l$ is the corresponding eigenvalue. Suppose $P = [p_1, p_2, \ldots, p_d]$ is the solution to Equation (20), and the corresponding eigenvalues are $\gamma_1 \geq \gamma_2 \geq \ldots \gamma_d$. Denote

$\Lambda = diag(\gamma_1 \geq \gamma_2 \geq \ldots \gamma_d)$ and we want to maximize this value. Obviously, if we choose non-negative eigenvalues $\gamma_i$, $\Lambda$ will obtain its maximum value. In our algorithm, $d$ is equal to the number of non-negative eigenvalues which makes $J(P)$ the sum of the $d$ largest eigenvalues of $\Lambda$.

In the third step, CESCP generates a consensus matrix $O$ by summarizing the clustering solutions $\{I^1, I^2, \ldots, I^B\}$ generated by the semi-supervised clustering model $E^2CP$. $E^2CP$ is a high-performance semi-supervised clustering algorithm, which solves the constraint propagation problem in quadratic time based on the $k$-nearest neighbor graph, and uses the propagated pairwise constraints to adjust the similarity matrix for constrained spectral clustering. As a result, we use $E^2CP$ as the base clustering algorithm to improve the performance of CESCP.

Each clustering solution $I^b$ ($b \in \{1, \ldots, B\}$) can first be transformed to an adjacency matrix $O^b$ with entries $o^b_{ij}$ as follows:

$$o^b_{ij} = \begin{cases} 1 & \text{if } y^b_i = y^b_j, \\ 0 & \text{if } y^b_i \neq y^b_j, \end{cases} \tag{21}$$

where $y^b_i$ and $y^b_j$ denote the predicted labels of the data samples $p_i$ and $p_j$ in the $b$th clustering solution, respectively. An $n \times n$ consensus matrix $O$ is then constructed by combining all of the adjacency matrices $(O^1, O^2, \ldots, O^B)$ as follows:

$$O = \frac{1}{B}\sum_{b=1}^{B} O^b. \tag{22}$$

Finally, CESCP takes Ncut as the consensus function to partition the feature vector set $P$ based on the consensus matrix $O$. We construct a graph ($G = (P, O)$) whose vertices denote the feature vectors, and whose edges correspond to the values of $o_{ij}$ in $O$, which represent the probability that the feature vectors belong to the same cluster. Ncut partitions the graph $G$ into two subgraphs recursively until $k$ subgraphs are obtained. The cost function $\Omega(P_1, P_2)$ of Ncut is defined as follows:

$$\Omega(P_1, P_2) = \frac{\Phi(P_1, P_2)}{\Psi(P_1, P)} + \frac{\Phi(P_1, P_2)}{\Psi(P_2, P)} \tag{23}$$

$$\Phi(P_1, P_2) = \sum_{p_i \in P_1, p_j \in P_2} o_{ij} \tag{24}$$

$$\Psi(P_1, P) = \sum_{p_i \in P_1, p_h \in P} o_{ih}, \tag{25}$$

where $\Omega(P_1, P_2)$ is a dissimilarity measure between $P_1$ and $P_2$, and $o_{ij}$ is the $ij$th entry of $O$. We can convert the above cost function as below:

$$\Omega(P_1, P_2) = \frac{\sum_{(v_i > 0, v_j < 0)} -o_{ij} v_i v_j}{\sum_{v_i > 0} \vartheta_i} + \frac{\sum_{(v_i < 0, v_j > 0)} -o_{ij} v_i v_j}{\sum_{v_i < 0} \vartheta_i}, \tag{26}$$

where $v = [v_1, \ldots, v_n]^T$ is an $n$-dimensional indicator vector ($n$ is the number of feature vectors in $P$), $v_i$ ($i \in \{1, \ldots, n\}$) takes on values in $\{-1, 1\}$, $v_i = 1$ if the $i$th vertex belongs to $P_1$. Otherwise, $v_i = -1$, and $\vartheta_i = \sum_j o_{ij}$.

The corresponding optimization problem is formulated as follows: [13]

$$\min_v \Omega(v) = \min_\alpha \frac{\alpha^T (U - O)\alpha}{\alpha^T U \alpha} \tag{27}$$

$$\alpha = (1 + v) - \iota(1 - v) \tag{28}$$

$$\iota = \frac{\sum_{v_i > 0} \vartheta_i}{\sum_{v_i < 0} \vartheta_i}, \tag{29}$$

with the constraints

$$\alpha_i \in \{-\iota, 1\}, \alpha^T U I = 0, \tag{30}$$

where $U$ is an $n \times n$ diagonal matrix with $\vartheta_i$ ($i \in \{1, \ldots, n\}$) on its diagonal, $I$ denotes the identity matrix, and $\eta_i$ is the $i$th component of $\alpha$.

However, the optimization problem above in its exact form is NP-complete. One possible way to solve the problem is to search for an approximate solution in the real value domain. As the above equation is in the form of a Rayleigh quotient, the above optimization problem can be solved through the following generalized eigenvalue system when $\alpha$ is relaxed to take on real values, as follows:

$$(U - O)\alpha = \Lambda U \alpha, \tag{31}$$

where $\Lambda$ denotes the eigenvalues. The second smallest eigenvector of the generalized eigenvalue system is the solution to the normalized cut problem [52].

## 4 Double Weighting Semi-Supervised Ensemble Clustering Based on Selected Constraint Projection

CESCP improves pairwise constraint selection and constraint projection but pays little attention to the clustering ensemble procedure. We have conducted some research on clustering member integration and made improvement on CESCP. Fig. 2 illustrates the framework for double weighting semi-supervised ensemble clustering based on selected constraint projection. When compared with CESCP, DCECP adopts an adaptive ensemble member weighting process(AEMW) based on a local environment uniform competition process and a non-local environment uniform competition process, to generate the weight of the ensemble member. There are two reasons for our proposed adaptive ensemble member weighting process: (1) Different ensemble members have different clustering qualities; and (2) Different ensemble members are obtained from different subspaces in the data. Algorithm 1 provides a summary of the DCECP approach.

Before introducing AEMW, we will first introduce the modified weighted Ncut since it is the basis of the weighted ensemble clustering in AEMW. In CESCP, we generate an adjacency matrix for every ensemble member, and combine all of the graphs to perform the Ncut partition, so that every ensemble member is treated equally. However, different ensemble members will make different contributions to the final result. We change the construction process of the consensus function of the final clustering solution in order to

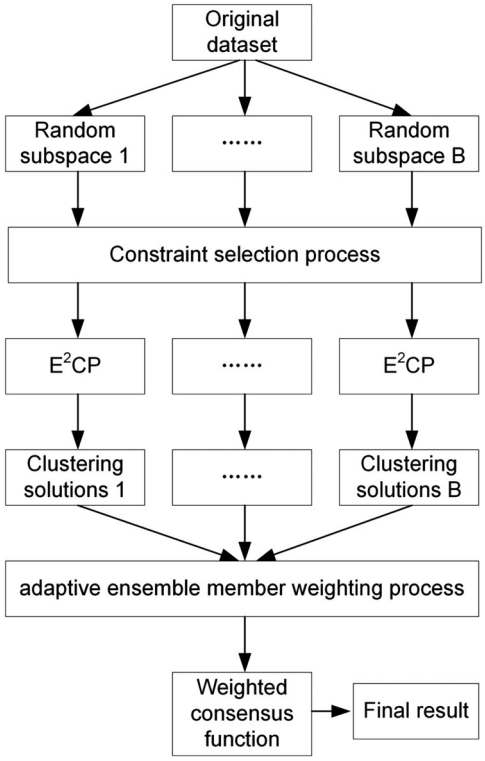



Fig. 2. Overview of the double weighting semi-supervised ensemble clustering based on selected constraint projection.

take the weight vector $W = (w_1, w_2, \ldots, w_B)$ into account, and the final consensus matrix $O$ is then constructed by combining all of the adjacency matrices $(O^1, O^2, \ldots, O^B)$ as follows:

$$O = \sum_{b=1}^{B} w_b O^b \tag{32}$$

$$st. w_1 + w_2 + \cdots + w_B = 1. \tag{33}$$

Finally, we apply the Ncut algorithm on this consensus matrix to obtain the result.

**Algorithm 1.** Double Weighting Semi-Supervised Ensemble Clustering Based on Selected Constraint Projection

**Require:**
**Input**: a high-dimensional dataset $P$;
**Ensure:**
1: Generate $B$ random subspaces $\{A_1, A_2, \ldots, A_B\}$;;
2: For each random subspace
3: Apply constraint selection and weighted projection clustering in CESCP and generate the semi-supervised clustering models $\chi_1, \chi_2, \ldots, \chi_B$;
4: Apply adaptive ensemble member weighting in Algorithm 2 and generate the weight vectors $(w_1, w_2, \ldots, w_B)$ for the B clustering solutions;
5: Obtain consensus matrix $O$ by summarizing the weighted clustering solutions $\{I^1, I^2, \ldots, I^{B'}\}$;
6: Apply consensus function to obtain the final result using the normalized cut approach;
**Output**: the labels of the samples in $P$.

Algorithm 2 provides an overview of the adaptive ensemble member weighting (AEMW) process. The input is the clustering solution set $\{I^1, I^2, \ldots, I^B\}$ and the maximum iteration number, while the output is the corresponding ensemble weight. Specifically, AEMW is an iterative process including three components: candidate weight set initialization, local environment construction, and uniform competition in the local and non-local environments.

**Algorithm 2.** Adaptive Ensemble Member Weighting Process

**Require:**
**Input**:
the clustering solution set $\{I^1, I^2, \ldots, I^B\}$;
maximum iteration $Max_{iter}$;
number of individuals to perform competition $\rho$; maximum unchange times $\omega$.
**Ensure:**
1: Generate initial set of weight vectors $\Psi$ which consists of three sets: equal set, leave one set, random set.
2: Set $t = 0$;
3: Set $unchange = 0$;
4: Set $best_{perform} = 0$;
5: Use the modified weighted Ncut and Equation (3) to calculate the ensemble performance of each vector in $\Psi$, sort the vectors in ascending order according to their clustering performance, and set $best_{perform}$ to the first one.
6: **Repeat**
7: $t = t + 1$;
8: Use the fuzzy membership function to construct the local environment of each vector in $\Psi$.
9: Choose the first $\rho$ weight vectors to perform uniform competition in the local environment and non-local environment
10: Sort the weight vectors in $\Psi$ and find the best performance $best'_{perform}$.
11: **if** $best'_{perform} < best_{perform}$
12: Set $best_{perform} = best'_{perform}$ and $unchange = 0$;
13: **else**
14: $unchange = unchange + 1$;
15: **Until** $unchange = \omega$ or $t = Max_{iter}$.
**Output**: the first weight vector in $\Psi$.

In the first step, AEMW generates $2B$ initial candidate weight vectors where we have $B$ clustering solutions per ensemble. These $2B$ initial candidate weight vectors $\Psi$ are made up of three sets. The first set, named the equal set, contains only one vector $W = (1/B, 1/B, \ldots, 1/B)$ which assigns equal weight to each clustering solution. The second set, the leave one set, contains $B$ vectors $W^i = w_1, w_2, \ldots, w_B$ in which $w_i = 1$ and $w_{j(j \neq i)} = 0$.The third set, the random set, contains $B - 1$ random weight vectors. The weight component in each weight vector in the random set is generated randomly but their sum will be equal to 1.Using the weight vector in $\Psi$, we can obtain a $2B$ ensemble solution with the modified weighted Ncut, and AEMW sorts the weight vectors in $\Psi$ in ascending order according to their ensemble solution performance using Equation (3). In this step, the best performance and its corresponding weight vector are recorded.

In the second step, AEMW finds the k-nearest neighbors (k here is set to 3) for each of the weight vectors $W$ in the candidate set $\Psi$ it generated in the first step, and treats them as the local environment. We use a fuzzy membership

function $\theta$ to calculate the likelihood of competition among neighbors as follows:

$$\theta(W^e, W^l) = \frac{1}{\sum_{h=1}^{k} \left(\frac{\phi(W^e, W^l)}{\phi(W^e, W^h)}\right)^{\frac{2}{q-1}}}, \tag{34}$$

where $l \in \{1, \ldots, k\}$, $W^l$ is a neighbor of $W^e$, $\frac{2}{q-1}$ denotes the fuzziness exponent, and $\phi(W^e, W^l)$ denotes the euclidean distance between weight vector $W^e$ and $W^l$. The cumulative membership $\eta$ of the neighbors is defined as follows:

$$\eta = \sum_{l=1}^{k} \theta(W^e, W^l). \tag{35}$$

The $l$th membership value interval $\tau_l (l \in (1, \ldots, k))$ is calculated by the $(l-1)$th and $l$th ones $\eta_{l-1}$ and $\eta_l$ as follows:

$$\tau_l = \left[\frac{\eta_{l-1}}{\eta}, \frac{\eta_l}{\eta}\right] \tag{36}$$

$$\eta_l = \eta_{l-1} + \theta(W^e, W^l) \tag{37}$$

$$\eta_{l-1} = \sum_{h=1}^{l-1} \theta(W^e, W^h) \tag{38}$$

$$\eta_0 = 0. \tag{39}$$

The $l$th membership interval $\tau_l$ will become larger when $\theta(W^e, W^l)$ increases. When AEMW generates a random value $r$ between 0 and 1,the likelihood that $r$ falls into interval $\tau_l$ is larger. This means that $W^l$ is likely to be selected to perform competition in the next step of AEMW.

In the third step, for the first $\rho$ weight vectors, AEMW performs a uniform competition in its local and non-local environments. The uniform competition between two weight vectors $W^e$ and $W^p$ is defined as follows:

$$W^e = (w_1^e, w_2^e, \ldots, w_B^e) \tag{40}$$

$$W^p = (w_1^p, w_2^p, \ldots, w_B^p). \tag{41}$$

AEMW makes copies of $W^e$ and $W^p$, as $W^{e'}$ and $W^{p'}$. Then it generates a random value $r$ for $i \in 1, 2, \ldots, B$. If $r$ is greater than 0.5, then $w_i^{e'}$ and $w_i^{p'}$ will exchange their values. After that, AEMW performs normalization on $W^{e'}$ and $W^{p'}$ to ensure that the sum of their $B$ weights is equal to 1 as follows:

$$W^{e'} = \frac{1}{\sum_{i=1}^{B} w_i^{e'}} (w_1^{e'}, w_2^{e'}, \ldots, w_B^{e'}) \tag{42}$$

$$W^{p'} = \frac{1}{\sum_{i=1}^{B} w_i^{p'}} (w_1^{p'}, w_2^{p'}, \ldots, w_B^{p'}). \tag{43}$$

While completing the local competition, AEMW generates a random value $r$ between 0 and 1 for a specific weight vector $W^s$. Suppose that $r$ falls into $\tau_l (l \in (1, \ldots, k))$ in Equation (36), then the $lth$ neighbor of $W^s$ is selected to perform a uniform competition with $W^s$. Non-local competition is similar to local competition, except that $W^s$ conducts the uniform competition with a random weight vector outside of its neighbors. After the competition, the candidate weight vector set is updated and AEMW sorts the weight vectors again to reduce the number of weight vectors in $\Psi$ to $2B$. Finally, AEMW removes the best vector and check if its performance is better than the current best one. If yes, we go to the second step. Otherwise, if the best performance stays unchanged $\omega$ times continuously, or the iteration reaches a maximum value, we end the adaptive weighting process.

## 5 Complexity Analysis

We also perform a time complexity analysis of the double weighting semi-supervised ensemble clustering based on selected constraint projection. The complexity $T_{DCECP}$ of DCECP is estimated as follows:

$$T_{DCECP} = T_{CESCP} + T_{AEMW} \tag{44}$$

where $T_{CESCP}$ and $T_{AEMW}$ represent the time complexity of the constraint-selected projection clustering process and the adaptive ensemble member weighting process, respectively. $T_{CESCP}$ is estimated as follows:

$$T_{CESCP} = T_{CSW} + T_{WCP} + T_{FR}, \tag{45}$$

where $T_{CSW}$, $T_{WCP}$, $T_{FR}$ denote the time complexity for the weighted constraint selection process, the weighted constraint projection and the final result generation step, respectively. $T_{CSW}$ is related to the number of samples $n$, the number of random subspaces $B$, the number of neighbors $K$ in the KNN graph in $E^2CP$, and the number of constraints $c$ as follows:

$$T_{CSW} = O(B \cdot c \cdot K \cdot n^2), \tag{46}$$

$T_{WCP}$ is affected by the number of constraints $c$, and the number of feature attribute $m$ as follows:

$$T_{WCP} = O(c \cdot m^2 + m^3), \tag{47}$$

$T_{FR}$ is related to the number of samples $n$ and the number of clusters $k$ in the clustering solution as follows:

$$T_{FR} = O(k \cdot n^3). \tag{48}$$

$T_{AEMW}$ will be affected by the number of feature vectors $n$, the number of random subspaces $B$, the number of maximum iterations $G$, and the number of clusters $k$ in the clustering solution as follows:

$$T_{AEMW} = O(G \cdot k \cdot B \cdot n^3). \tag{49}$$

Since $B$, $c$, $K$, $G$ and $k$ are significantly smaller than $n^3$ or $m^3$, the time complexity of DCECP is approximately $O(e^3)$, where $e$ is the larger value between $n$ and $m$.

The space complexity of DCECP consists of the space complexity of the constraint-selected projection clustering ensemble members $O(Bnm)$ and the space complexity of the adaptive ensemble member weighting process $O(n^2)$. As a result, the space complexity of DCECP is $O(e^2)$, where $e$ is the large value between $n$ and $m$.

## 6 Experiment

The performances of DCECP and other semi-supervised clustering approaches are evaluated using 15 real-world

TABLE 1
Summary of the Datasets

| Dataset | Source | sample | feature | class |
|---|---|---|---|---|
| Alizadeh-2000-v3 | GENE | 62 | 2093 | 4 |
| Armstrong-2002-v2 | GENE | 72 | 2194 | 3 |
| Bredel-2005 | GENE | 50 | 1739 | 3 |
| Dyrskjot-2003 | GENE | 40 | 1203 | 3 |
| Nutt-2003-v1 | GENE | 50 | 1377 | 4 |
| Pomeroy-2002-v2 | GENE | 42 | 1379 | 5 |
| Su-2001 | GENE | 174 | 1571 | 10 |
| Tomlins-2006-v1 | GENE | 104 | 2315 | 3 |
| mfeat | UCI | 2000 | 649 | 9 |
| semeion | UCI | 1593 | 256 | 10 |
| GLIOMA | ASU | 50 | 4434 | 4 |
| nci9 | ASU | 60 | 9712 | 9 |
| lymphoma | ASU | 96 | 4026 | 9 |
| lung_discrete | ASU | 73 | 325 | 7 |
| yale | ASU | 165 | 1024 | 15 |

datasets as shown in Table 1. Our experiment data includes 8 datasets from GENE repository, 2 datasets from the UCI repository, and 5 datasets from the ASU data repository.

In the validation step, normalized mutual information (NMI, [55]) is used to compare DCECP with other methods. We evaluated each method by calculating NMI mean value and standard deviation after 10 runs.

Given the ground truth result $I$ with $k$ clusters $I = \{C_1, C_2, \ldots, C_k\}$, and the result $I'$ obtained by DCECP with $k'$ clusters $I' = \{C'_1, C'_2, \ldots, C'_{k'}\}$, we use NMI to assess the performance of the clustering result as follows:

$$NMI(I, I') = \frac{2H_1(I; I')}{H_2(I) + H_2(I')} \tag{50}$$

$$H_1(I; I') = \sum_h \sum_l \frac{|C_h \cap C'_l|}{n} \log \frac{n|C_h \cap C'_l|}{|C_h||C'_l|} \tag{51}$$

$$H_2(I) = -\sum_h \frac{|C_h|}{n} \log \frac{|C_h|}{n} \tag{52}$$

$$H_2(I') = -\sum_l \frac{|C'_l|}{n} \log \frac{|C'_l|}{n}, \tag{53}$$

where $n$ denotes the sample size, and $|\cdot|$ denotes the number of elements in the cluster. Higher NMI value represents better clustering performance.

In the following experiments, we first study the effects of the parameters. Then, we explore the effects of weighted constraint projection and the adaptive ensemble member weighting. Next, the proposed approach, DCECP, is compared with single semi-supervised clustering methods and semi-supervised clustering ensemble approaches on the real-world datasets. Finally, nonparametric tests are applied to evaluate the performance of each approach.

### 6.1 Effects of the Parameters

We conduct experiments on six datasets with respect to the average NMI values to investigate the effects of the parameters, which include the sampling rate in the random subspace technique, and the number of pairwise constraints used by the approach.

To explore the effect of the sampling rate $\tau$, we vary its value from 0.1 to 0.5 with an increment of 0.05. The experiment results are shown in Fig. 3. We observed that when $\tau = 0.3$, DCECP achieves good performance on most of the datasets, such as the lung_discrete dataset, the lymphoma dataset, the mfeat dataset and the Alizadeh-2002-v3 dataset. A possible reason could be that when $\tau = 0.3$, the underlying structure of the datasets can be adequately captured. Therefore, $\tau$ is set to 0.3 in the following experiments.

We also vary the number of pairwise constraints $n$ from $0.5n$ to $2n$ with an increment of $0.5n$ to study the effect of the number of pairwise constraints. As we can see in Fig. 4, when the number of pairwise constraints increases, the average NMI values increase gradually on most of the datasets, such as the Alizadeh-2000-v3 dataset, the lymphoma dataset, the mfeat dataset, and so on. The possible reason could be that more pairwise constraints will provide additional useful

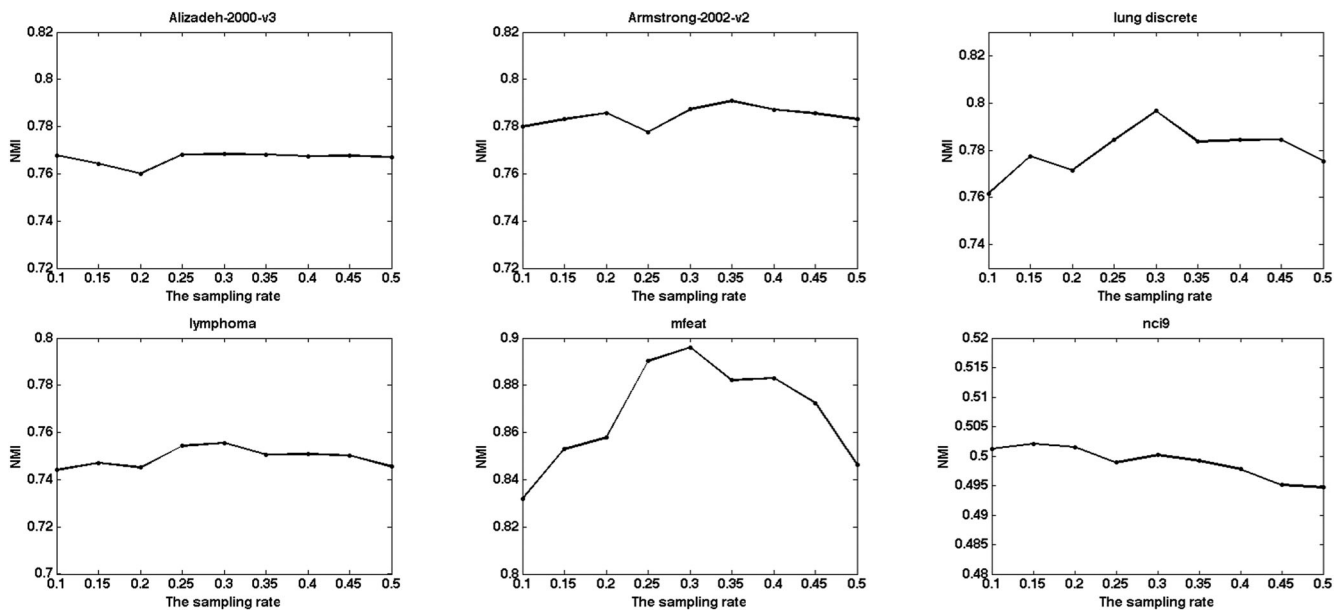



Fig. 3. The effect of the sampling rate $\tau$.

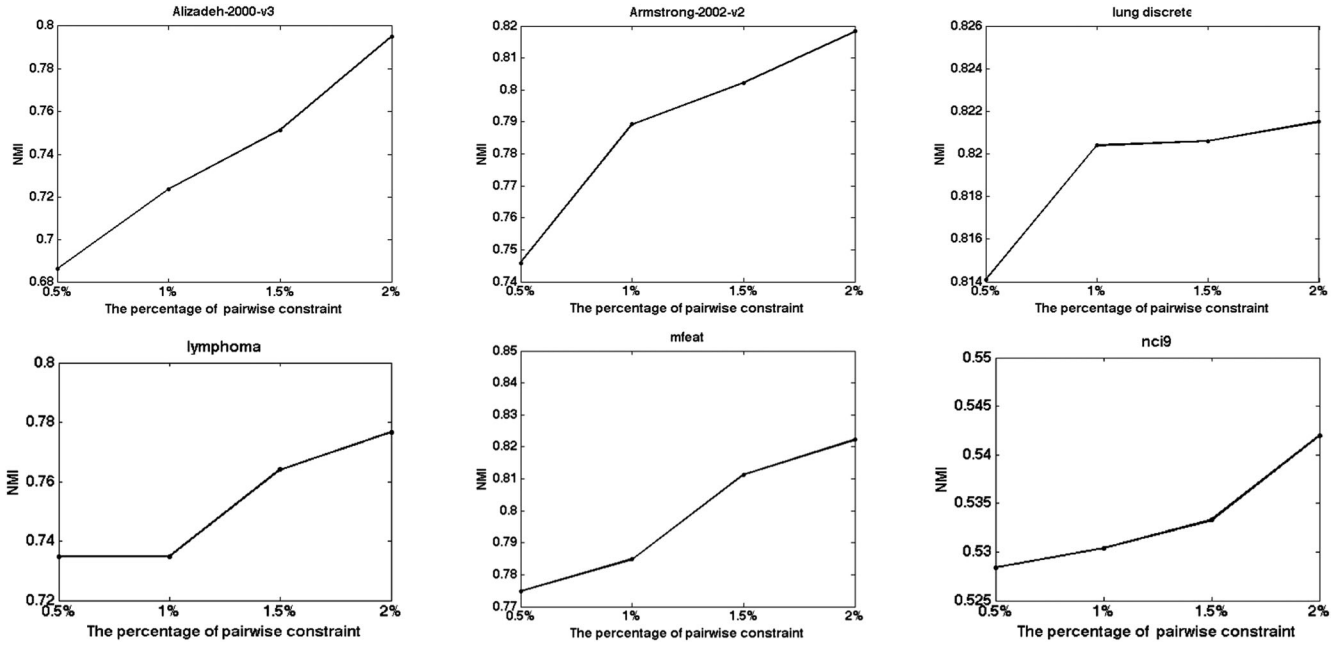



Fig. 4. The effect of the number of pairwise constraints.

information, which will improve the final result. However, more pairwise constraints require more labeling efforts from experts. To keep a balance between the labeling effort and the performance of the algorithm, we set $n$ as the default value for the number of pairwise constraints.

### 6.2 Effects of the Constraint Projection Process

To investigate the effects of the constraint projection process, we remove the constraint projection process from DCECP and name it DCECP (without WCP). Then we compare DCECP with DCECP (without WCP) on eight datasets in Table 1. DCECP (without WCP) denotes DCECP without the constraint weighting and weighted constraint projection processes.

It can be seen from Table 2 that DCECP outperforms DCECP (without WCP) on most of the datasets. The NMI 0.8738, 0.7653, 0.6442, 0.6868, and 0.7221 obtained by DCECP on the Armstrong-2002-v2, Dyrskjot-2003, Nutt-2003-v1, Pomeroy-2002-v2,and Tomlins-2006-v1 datasets, respectively, are 0.1025, 0.0.0745, 0.0.1968, 0.061, and 0.1092 larger than those obtained by DCECP (without WCP). The possible reasons could be that DCECP assigns different importance to prior knowledge (pairwise constraints) in different subspaces of the data, and uses this weight information in its constraint projection process. This process makes the original data space closer to the distribution which the constraints describe. In this new distribution, ML objects are close together and CL objects are far from each other so that the clustering approach can perform better.

### 6.3 Effects of the Adaptive Ensemble Member Weighting Process

To investigate the effects of the adaptive ensemble member weighting process, we remove the adaptive ensemble member weighting process from DCECP and name it DCECP (without AEMW). Then we compare DCECP with DCECP (without AEMW) on the same datasets used above.

Table 3 shows the results obtained by DCECP and DCECP (without AEMW). We can see that DCECP outperforms DCECP (without AEMW) on most of the datasets. For example, the NMI values of 0.8738, 0.7653, 0.6442, 0.6868, and 0.7221 obtained by DCECP on the Armstrong-2002-v2, Dyrskjot-2003, Nutt-2003-v1, Pomeroy-2002-v2, and Tomlins-2006-v1 datasets, respectively, are 0.0355, 0.0.1203, 0.0.2927, 0.0897, and 0.0616 larger than those obtained by DCECP (without AEMW). This confirms the effectiveness of the AEMW process in our framework. AEMW uses an evolutionary process to calculate the weight

TABLE 2
The Comparison of DCECP and DCECP (Without WCP)

| Datasets | DCECP(without WCP) | DCECP |
|---|---|---|
| Alizadeh-2000-v3 | $0.6622 \pm 0$ | **0.7737** $\pm$ 0.0048 |
| Armstrong-2002-v2 | 0.7713$\pm$0.0100 | **0.8738** $\pm$ 0.0376 |
| Bredel-2005 | $0.4425 \pm 0$ | **0.5681** $\pm$ 0.0225 |
| Dyrskjot-2003 | $0.6620 \pm 0$ | **0.7653** $\pm$ 0.0232 |
| Nutt-2003-v1 | $0.3983 \pm 0.0173$ | **0.6442** $\pm$ 0.0099 |
| Pomeroy-2002-v2 | $0.5688 \pm 0.0224$ | **0.6868** $\pm$ 0.0701 |
| Su-2001 | $0.7273 \pm 0$ | **0.7579** $\pm$ 0.0266 |
| Tomlins-2006-v1 | $0.5770 \pm 0.0100$ | **0.7221** $\pm$ 0.0238 |

TABLE 3
The Comparison of DCECP with DCECP (Without AEMW)

| Datasets | DCECP(without AEMW) | DCECP |
|---|---|---|
| Alizadeh-2000-v3 | **0.7779** $\pm$ 0.0003 | 0.7737 $\pm$ 0.0048 |
| Armstrong-2002-v2 | $0.8383 \pm 0.0016$ | **0.8738** $\pm$ 0.0376 |
| Bredel-2005 | $0.524 \pm 0.0002$ | **0.5681** $\pm$ 0.0225 |
| Dyrskjot-2003 | 0.645 $\pm$0.0112 | **0.7653** $\pm$ 0.0232 |
| Nutt-2003-v1 | $0.3515 \pm 0.0006$ | **0.6442** $\pm$ 0.0099 |
| Pomeroy-2002-v2 | $0.5971 \pm 0.0033$ | **0.6868** $\pm$ 0.0701 |
| Su-2001 | $0.7339 \pm 0.0002$ | **0.7579** $\pm$ 0.0266 |
| Tomlins-2006-v1 | $0.6605 \pm 0.0008$ | **0.7221** $\pm$ 0.0238 |

TABLE 4
Comparison with Single Semi-Supervised Clustering Approaches (the Table does not Show the Standard Deviation Values Which Are Less than 0.001)

| Datasets | DCECP | PC-kmeans | $E^2CP$ |
|---|---|---|---|
| Alizadeh-2000-v3 | **0.7737** ± 0.0048 | 0.5210 ± 0.1565 | 0.7084 |
| Armstrong-2002-v2 | **0.8738** ± 0.0376 | 0.6244 ± 0.1217 | 0.8424 |
| Bredel-2005 | **0.5681** ± 0.0225 | 0.5426 ± 0.0693 | 0.3763 |
| Dyrskjot-2003 | **0.7653** ± 0.0232 | 0.6402 ± 0.1212 | 0.6891 |
| Nutt-2003-v1 | **0.6442** ± 0.0099 | 0.5490 ± 0.1105 | 0.4426 ± 0.0245 |
| Pomeroy-2002-v2 | **0.6868** ± 0.0701 | 0.434 ± 0.1364 | 0.6198 ± 0.0469 |
| Su-2001 | **0.7579** ± 0.0266 | 0.6203 ± 0.0860 | 0.7050 ± 0.0100 |
| Tomlins-2006-v1 | **0.7221** ± 0.0238 | 0.4269 ± 0.0854 | 0.6203 ± 0.0100 |
| mfeat | **0.8961** ± 0.0399 | 0.5796 ± 0.0155 | 0.7572 ± 0.0075 |
| semeion | **0.7038** ± 0.0577 | 0.5541 ± 0.0287 | 0.6315 |
| GLIOMA | 0.5191 ± 0.0425 | 0.4314 | **0.5748** |
| nci9 | **0.8192** ± 0.0841 | 0.1203 ± 0.0538 | 0.4937 ± 0.0107 |
| lymphoma | **0.7556** ± 0.0481 | 0.277 ± 0.0891 | 0.7515 ± 0.0033 |
| lung_discrete | **0.7966** ± 0.0382 | 0.3292 | 0.7626 ± 0.0223 |
| yale | **0.5621** ± 0.0552 | 0.2265 ± 0.0412 | 0.5496 ± 0.0087 |

of the ensemble member based on its performance, so that it can assign different importance to different ensemble members to achieve better performance.

### 6.4 Comparison of Single Semi-Supervised Clustering Approaches

We have compared DCECP with the pairwise constraints based K-means algorithm (PC-Kmeans, [56]) and the constraint propagation approach ($E^2CP$, [57]) based on NMI on all the datasets in Table 1. PC-Kmeans [56] is one of the most popular semi-supervised clustering algorithms. It will serve as the baseline for the comparison of other semi-supervised clustering algorithms.

We can see the comparison result of DCECP and single semi-supervised methods on all of the datasets from Table 4. As shown in the table, DCECP outperforms other semi-supervised methods. The possible reasons are as follows: (1) The random subspace technique helps reduce the dimension of the original space, which is useful for reducing the noisy and redundant features, and (2) DCECP assigns different importance to ensemble members and integrates them using an adaptive weighting process, which provides more accurate and robust results.

### 6.5 Comparison of Semi-Supervised Clustering Ensemble Approaches

In the following experiments, DCECP (A7) is compared with a number of ensemble methods including neural gas clustering ensemble (NGCE, A1), random K-means clustering ensemble (RSKE, A2, [11]), bagging K-means clustering ensemble(BAGKE, A3, [58]), hierarchical clustering ensemble(HCCE, A4, [15]), exhaustive and efficient constraint propagation clustering ensemble($E^2CPE$, A5, [57]), and incremental semi-supervised clustering ensemble(ISSCE, A6, [48]).

The experiment result of seven ensemble approaches is shown in Fig. 1 in the supplementary file, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TKDE.2018.2818729, and Table 5. We make three interesting observations in Fig. 1 in the supplementary file, available online. First, DCECP (A7) achieves the best performance on 14 out of 15 datasets. The main reasons are as follows: (1) The use of random subspace is suitable for DCECP to alleviate the effect of high dimensionality. (2) Constraint weighting and constraint projection with weighting help transform the original feature space, and improve clustering performance. (3) The adaptive ensemble member weighting process is effective for combining different ensemble members and obtaining better results. Second, the state-of-the-art clustering method ISSCE(A6), which was proposed by us, is less effective than DCECP (A7) on most of the datasets. A possible reason for this is that besides clustering ensembles, the DCECP framework optimizes the feature space which makes good use of prior information encapsulated in the pairwise constraints, such that the overall quality of the final result is improved. Third, the results obtained by the cluster ensemble approaches such as NGCE(A1),RSKE (A2), BAGKE (A3) and HCCE (A4) are less satisfactory than those obtained by semi-supervised methods. This indicates that pairwise constraints play an important role in improving performance. In summary, DCECP is good at dealing

TABLE 5
Comparison of Semi-Supervised Clustering Ensemble Approaches on the Datasets in Table 1 with Respect to NMI (the Table does not Show the Standard Deviation Values Which Are Less than 0.001)

| Datasets | NGCE | RSKE | BAGKE | HCCE | $E^2CPE$ | ISSCE | DCECP |
|---|---|---|---|---|---|---|---|
| Alizadeh-2000-v3 | 0.5017 ± 0.072 | 0.6085 ± 0.2587 | 0.5016 ± 0.0566 | 0.3131 ± 0.1145 | 0.6985 | 0.721 ± 0.0141 | **0.7737** ± 0.0048 |
| Armstrong-2002-v2 | 0.6463 ± 0.0587 | 0.6571 ± 0.3063 | 0.6889 ± 0.0938 | 0.0939 ± 0.0548 | 0.8389 ± 0.01 | 0.8536 ± 0.0141 | **0.8738** ± 0.0376 |
| Bredel-2005 | 0.3768 ± 0.0697 | 0.3679 ± 0.1711 | 0.3754 ± 0.0316 | 0.0664 ± 0.02 | 0.4425 | 0.5666 ± 0.03 | **0.5681** ± 0.0225 |
| Dyrskjot-2003 | 0.4114 ± 0.0505 | 0.5369 ± 0.3089 | 0.4761 ± 0.0849 | 0.393 ± 0.1253 | 0.662 | 0.7345 ± 0.0837 | **0.7653** ± 0.0232 |
| Nutt-2003-v1 | 0.5149 ± 0.0344 | 0.505 ± 0.1892 | 0.4978 ± 0.0316 | 0.3405 ± 0.1396 | 0.3983 ± 0.0173 | 0.4749 ± 0.04 | **0.6442** ± 0.0099 |
| Pomeroy-2002-v2 | 0.5628 ± 0.0429 | 0.556 ± 0.2648 | 0.4702 ± 0.0616 | 0.2951 ± 0.1082 | 0.5688 ± 0.0224 | 0.6408 ± 0.0671 | **0.6868** ± 0.0701 |
| Su-2001 | 0.523±0.0288 | 0.7±0.1556 | 0.6699 ± 0.0283 | 0.3984 ± 0.0888 | 0.7273 | 0.7408 ±0.02 | 0.7579± 0.0266 |
| Tomlins-2006-v1 | 0.4522 ± 0.0206 | 0.4574 ± 0.1640 | 0.4181 ± 0.0469 | 0.2389 ± 0.0490 | 0.577 ± 0.01 | 0.6557 ± 0.0245 | **0.7221** ± 0.0238 |
| mfeat | 0.5287 ± 0.032 | 0.5908 ± 0.0468 | 0.5152 ± 0.0452 | 0.3457 ± 0.2093 | 0.7371 | 0.8565 ± 0.0498 | **0.8961** ± 0.0398 |
| semeion | 0.5054 ± 0.0189 | 0.4996 ± 0.0119 | 0.5275 ± 0.0389 | 0.1295 ± 0.1238 | 0.6452 | 0.6561 ± 0.028 | **0.7038** ± 0.0577 |
| GLIOMA | 0.4867±0.0672 | 0.4859±0.0324 | 0.4765±0.0706 | 0.2804±0.1069 | 0.5713±0.0112 | **0.6133** ± 0.0282 | 0.5191 ± 0.0425 |
| nci9 | 0.449 ± 0.0319 | 0.4564 ± 0.015 | 0.4669 ± 0.0355 | 0.3295 ± 0.0437 | 0.5491 ± 0.0109 | 0.5327 ± 0.0154 | **0.8192** ± 0.0841 |
| lymphoma | 0.6634 ± 0.0255 | 0.6806 ± 0.0237 | 0.6653 ± 0.0501 | 0.3924 ± 0.0584 | 0.7574 ± 0.014 | 0.7291 ± 0.0183 | **0.7556** ± 0.0481 |
| lung_discrete | 0.4139 ± 0.017 | 0.6789 ± 0.0586 | 0.7115 ± 0.0416 | 0.4761 ± 0.0954 | 0.7333 ± 0.0084 | 0.8244 ± 0.0117 | **0.7966** ± 0.0382 |
| yale | 0.4257 ± 0.0176 | 0.4897 ± 0.0242 | 0.4966 ± 0.026 | 0.3053 ± 0.034 | 0.5571 ± 0.0073 | 0.546 ± 0.019 | **0.5621** ± 0.0552 |

TABLE 6
Average Rankings of the Algorithms (Friedman)

| Algorithm | Ranking |
|---|---|
| DDCECP | 1.13 |
| BAGKE | 4.87 |
| $E^2$CPE | 2.93 |
| HCCE | 6.93 |
| ISSCE | 2.33 |
| NGCE | 5.13 |
| RSKE | 4.67 |

TABLE 7
Results of Holm's Procedure Based on the Performances of Compared Algorithms at Significance Level $\alpha = 0.05$ (Control Test: DCECP is the Control Algorithm)

| i | Algortihm | $z$ | $p$-Value | $\alpha/i$ | Hypothesis |
|---|---|---|---|---|---|
| 6 | **HCCE** | 7.35E+00 | 1.94E-13 | 8.33E-03 | rejected |
| 5 | **NGCE** | 5.07E+00 | 3.96E-07 | 1.00E-02 | rejected |
| 4 | **BAGKE** | 4.73E+00 | 2.21E-06 | 1.25E-02 | rejected |
| 3 | **RSKE** | 4.48E+00 | 7.49E-06 | 1.67E-02 | rejected |
| 2 | **$E^2$CPE** | 2.28E+00 | 2.25E-02 | 2.50E-02 | rejected |
| 1 | ISSCE | 1.52E+00 | 1.28E-01 | 5.00E-02 | not rejected |

TABLE 8
Results of Holm's Procedure Based on the Performances of Compared Algorithms at Significance Level $\alpha = 0.05$ (Multiple Test)

| i | Algortihm | $z$ | $p$-Value | $\alpha/i$ | Hypothesis |
|---|---|---|---|---|---|
| 21 | **DCECP versus HCCE** | 7.35E+00 | 1.94E-13 | 2.38E-03 | rejected |
| 20 | **HCCE versus ISSCE** | 5.83E+00 | 5.49E-09 | 2.50E-03 | rejected |
| 19 | **DCECP versus NGCE** | 5.07E+00 | 3.96E-07 | 2.63E-03 | rejected |
| 18 | **$E^2$CPE versus HCCE** | 5.07E+00 | 3.96E-07 | 2.78E-03 | rejected |
| 17 | **DCECP versus BAGKE** | 4.73E+00 | 2.21E-06 | 2.94E-03 | rejected |
| 16 | **DCECP versus RSKE** | 4.48E+00 | 7.49E-06 | 3.13E-03 | rejected |
| 15 | **ISSCE versus NGCE** | 3.55E+00 | 3.86E-04 | 3.33E-03 | rejected |
| 14 | **BAGKE versus ISSCE** | 3.21E+00 | 1.32E-03 | 3.57E-03 | rejected |
| 13 | **ISSCE versus RSKE** | 2.96E+00 | 3.10E-03 | 3.85E-03 | rejected |
| 12 | **HCCE versus RSKE** | 2.87E+00 | 4.06E-03 | 4.17E-03 | rejected |
| 11 | $E^2$CPE versus NGCE | 2.79E+00 | 5.29E-03 | 4.55E-03 | not rejected |
| 10 | BAGKE versus HCCE | 2.62E+00 | 8.79E-03 | 5.00E-03 | not rejected |
| 9 | BAGKE versus $E^2$CPE | 2.45E+00 | 1.42E-02 | 5.56E-03 | not rejected |
| 8 | DCECP versus $E^2$CPE | 2.28E+00 | 2.25E-02 | 6.25E-03 | not rejected |
| 7 | HCCE versus NGCE | 2.28E+00 | 2.25E-02 | 7.14E-03 | not rejected |
| 6 | $E^2$CPE versus RSKE | 2.20E+00 | 2.80E-02 | 8.33E-03 | not rejected |
| 5 | DCECP versus ISSCE | 1.52E+00 | 1.28E-01 | 1.00E-02 | not rejected |
| 4 | $E^2$CPE versus ISSCE | 7.61E-01 | 4.47E-01 | 1.25E-02 | not rejected |
| 3 | NGCE versus RSKE | 5.92E-01 | 5.54E-01 | 1.67E-02 | not rejected |
| 2 | BAGKE versus NGCE | 3.38E-01 | 7.35E-01 | 2.50E-02 | not rejected |
| 1 | BAGKE versus RSKE | 2.54E-01 | 8.00E-01 | 5.00E-02 | not rejected |

with different kinds of real-world datasets, especially high-dimensional datasets.

### 6.6 Nonparametric Tests

We perform nonparametric tests to evaluate the performance of different ensemble methods including NGCE, RSKE, BAGKE, HCCE, $E^2$CPE, ISSCE and DCECP, over the 15 datasets listed in Table 1, to identify the significant differences among the results shown in Fig. 3. Table 6 shows the rank of the algorithms (a higher rank denotes a better algorithm). Tables 7, 8, 9, and Table 1 in the supplementary file, available online, show the comparison results using three statistical tests: Friedman test, Holm test and multiple test. We can see that DCECP attains the highest average ranking among all of the testing ensemble methods.

TABLE 9
Results of Holm's Procedure Based on the Performances of Compared Algorithms at Significance Level $\alpha = 0.1$ (Multiple Test)

| i | Algortihm | $z$ | $p$-Value | $\alpha/i$ | Hypothesis |
|---|---|---|---|---|---|
| 21 | **DCECP versus HCCE** | 7.35E+00 | 1.94E-13 | 4.76E-03 | rejected |
| 20 | **HCCE versus ISSCE** | 5.83E+00 | 5.49E-09 | 5.00E-03 | rejected |
| 19 | **DCECP versus NGCE** | 5.07E+00 | 3.96E-07 | 5.26E-03 | rejected |
| 18 | **$E^2$CPE versus HCCE** | 5.07E+00 | 3.96E-07 | 5.56E-03 | rejected |
| 17 | **DCECP versus BAGKE** | 4.73E+00 | 2.21E-06 | 5.88E-03 | rejected |
| 16 | **DCECP versus RSKE** | 4.48E+00 | 7.49E-06 | 6.25E-03 | rejected |
| 15 | **ISSCE versus NGCE** | 3.55E+00 | 3.86E-04 | 6.67E-03 | rejected |
| 14 | **BAGKE versus ISSCE** | 3.21E+00 | 1.32E-03 | 7.14E-03 | rejected |
| 13 | **ISSCE versus RSKE** | 2.96E+00 | 3.10E-03 | 7.69E-03 | rejected |
| 12 | **HCCE versus RSKE** | 2.87E+00 | 4.06E-03 | 8.33E-03 | rejected |
| 11 | **$E^2$CPE versus NGCE** | 2.79E+00 | 5.29E-03 | 9.09E-03 | rejected |
| 10 | **BAGKE versus HCCE** | 2.62E+00 | 8.79E-03 | 1.00E-02 | rejected |
| 9 | BAGKE versus $E^2$CPE | 2.45E+00 | 1.42E-02 | 1.11E-02 | not rejected |
| 8 | DCECP versus $E^2$CPE | 2.28E+00 | 2.25E-02 | 1.25E-02 | not rejected |
| 7 | HCCE versus NGCE | 2.28E+00 | 2.25E-02 | 1.43E-02 | not rejected |
| 6 | $E^2$CPE versus RSKE | 2.20E+00 | 2.80E-02 | 1.67E-02 | not rejected |
| 5 | DCECP versus ISSCE | 1.52E+00 | 1.28E-01 | 2.00E-02 | not rejected |
| 4 | $E^2$CPE versus ISSCE | 7.61E-01 | 4.47E-01 | 2.50E-02 | not rejected |
| 3 | NGCE versus RSKE | 5.92E-01 | 5.54E-01 | 3.33E-02 | not rejected |
| 2 | BAGKE versus NGCE | 3.38E-01 | 7.35E-01 | 5.00E-02 | not rejected |
| 1 | BAGKE versus RSKE | 2.54E-01 | 8.00E-01 | 1.00E-01 | not rejected |

## 7 Conclusion and Future Work

In this paper, we propose the double weighting semi-supervised ensemble clustering based on selected constraint projection to address the limitations of traditional semi-supervised clustering ensemble methods. The proposed approach has three advantages compared with conventional semi-supervised clustering ensemble methods. (1) DCECP adopts the random subspace technique in combination with the constraint project procedure to perform high-dimensional data clustering. (2) It generates different subsets of pairwise constraints to reduce the effect of redundant pairwise constraints. (3) An adaptive ensemble member weighting process is designed to emphasize the importance of different ensemble members, and avoid the effect of deleterious ensemble members. In addition, the nonparametric tests further verify the high performance of DCECP. Experiments are constructed based on 15 real-world high-dimensional datasets and we reach the following conclusions. (1) Suitable parameter values will improve the performance of the proposed approach. (2) The adaptive ensemble member weighting process plays an important role in DCECP. (3) The results obtained by DCECP on most high-dimensional datasets are better than those obtained by other semi-supervised clustering ensemble approaches. In future, we will consider how to apply the semi-supervised clustering ensemble approach based on selected constraint projection to other research areas.

## ACKNOWLEDGMENTS

The work described in this paper was partially funded by grants from the NSFC No. 61722205, No. 61751205, No. 61572199, No. 61572540, No. 61472145, and No. U1611461, the grant from the Guangdong Natural Science Funds ( No. 2017A030312008), the grant from the Science and Technology Planning Project of Guangdong Province, China (No. 2015A050502011, No. 2016B090918042, No. 2016A050503015, No. 2016B010127003), the grant from the Guangzhou Science and Technology Planning Project (No. 201704030051), the grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (No. CityU 11300715), the grant from the Hong Kong General Research Grant(GRF, Ref. No. 152202/14E) and the PolyU Central Research Grant (G-YBJW), and a grant from the City University of Hong Kong [Project No. 7004884].

## REFERENCES

[1] A. Strehl and J. Ghosh, "Cluster ensembles–A knowledge reuse framework for combining multiple partitions," *J.Mach. Learn. Res.*, vol. 3, pp. 583–617, 2002.

[2] I. T. Christou, "Coordination of cluster ensembles via exact methods," *IEEE Tran. Pattern Anal. Mach. Intell.*, vol. 33, no. 2, pp. 279–293, Feb. 2011.

[3] N. Iam-On, T. Boongoen, S. Garrett, and C. Price, "A link-based approach to the cluster ensemble problem," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2396–2409, Dec. 2011.

[4] Z. Yu, L. Li, H.-S. Wong, J. You, G. Han, Y. Gao, and G. Yu, "Probabilistic cluster structure ensemble," *Inf. Sci.*, vol. 267, pp. 16–34, 2014.

[5] Z. Yu, X. Zhu, H. S. Wong, J. You, J. Zhang, and G. Han, "Distribution-based cluster structure selection," *IEEE Trans. Cybern.*, vol. 47, no. 11, pp. 3554–3567, Nov. 2017.

[6] Z. Yu, L. Li, J. Liu, J. Zhang, and G. Han, "Adaptive noise immune cluster ensemble using affinity propagation," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 12, pp. 3176–3189, Dec. 2015.

[7] Z. Yu, J. You, H.-S. Wong, and G. Han, "From cluster ensemble to structure ensemble," *Inf. Sci.*, vol. 168, pp. 81–99, 2012.

[8] Y. Yang and K. Chen, "Temporal data clustering via weighted clustering ensemble with different representations," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 2, pp. 307–320, Feb. 2011.

[9] K. Tademir, Y. Moazzen, and I. Yildirim, "An approximate spectral clustering ensemble for high spatial resolution remote-sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 5, pp. 1996–2004, May 2015.

[10] J. T. Tsai, Y. Y. Lin, and H. Y. M. Liao, "Per-cluster ensemble kernel learning for multi-modal image clustering with group-dependent feature selection," *IEEE Trans. Multimedia*, vol. 16, no. 8, pp. 2229–2241, Dec. 2014.

[11] Z. Yu, H. Chen, J. You, G. Han, and L. Li, "Hybrid fuzzy cluster ensemble framework for tumor clustering from biomolecular data," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 10, no. 3, pp. 657–670, May 2013.

[12] Z. Yu, H.-S. Wong, J. You, Q. Yang, and H. Liao, "Knowledge based cluster ensemble for cancer discovery from biomolecular data," *IEEE Trans. NanoBioSci.*, vol. 10, no. 2, pp. 76–85, Jun. 2011.

[13] X. Hu, E. K. Park, and X. Zhang, "Microarray gene cluster identification and annotation through cluster ensemble and EM-based informative textual summarization," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 5, pp. 832–840, Sep. 2009.

[14] W. Zhuang, Y. Ye, Y. Chen, and T. Li, "Ensemble clustering for internet security applications," *IEEE Trans. Syst. Man Cybern. Part C*, vol. 42, no. 6, pp. 1784–1796, Nov. 2012.

[15] F. Li, K. Wu, J. Lei, M. Wen, Z. Bi, and C. Gu, "Steganalysis over large-scale social networks with high-order joint features and clustering ensembles," *IEEE Trans. Inf. Forensics Secur.*, vol. 11, no. 2, pp. 344–357, Feb. 2016.

[16] X. Zhu, "Semi-supervised learning literature survey," Dept. Comput. Sci., Univ. Wisconsin-Madison, Madison, WI, USA, 2008.

[17] L. Wu, S. C. H. Hoi, R. Jin, J. Zhu, and N. Yu, "Learning Bregman distance functions for semi-supervised clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 3, pp. 478–491, Mar. 2012.

[18] H. Zeng and Y.-M. Cheung, "Semi-supervised maximum margin clustering with pairwise constraints," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 5, pp. 926–939, May 2012.

[19] S. Anand, S. Mittal, O. Tuzel, and P. Meer, "Semi-supervised Kernel mean shift clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 6, pp. 1201–1215, Jun. 2014.

[20] X. Fang, Y. Xu, X. Li, Z. Lai, and W. K. Wong, "Robust semi-supervised subspace clustering via non-negative low-rank representation," *IEEE Trans. Cybern.*, vol. 46, no. 8, pp. 1828–1838, Aug. 2016.

[21] C. L. Liu, W. H. Hsaio, C. H. Lee, and F. S. Gou, "Semi-supervised linear discriminant clustering," *IEEE Trans. Cybern.*, vol. 44, no. 7, pp. 989–1000, Jul. 2014.

[22] H. Wang, T. Li, and Y. Yang, "Constraint neighborhood projections for semi-supervised clustering," *IEEE Trans. Cybern.*, vol. 44, no. 5, pp. 636–643, May 2014.

[23] S. Xiong, J. Azimi, and X. Z. Fern, "Active learning of constraints for semi-supervised clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 1, pp. 43–54, Jan. 2014.

[24] S. Huang, H. Wang, T. Li, Y. Yang, and T. Li, "Constraint co-projections for semi-supervised co-clustering," *IEEE Trans. Cybern.*, vol. 46, no. 12, pp. 3047–3058, Dec. 2016.

[25] C.-C. Chang and H.-Y. Chen, "Semi-supervised clustering with discriminative random fields," *Pattern Recognit.*, vol. 45, no. 12, pp. 4402–4413, 2012.

[26] N. Grira, M. Crucianu, and N. Boujemaa, "Active semi-supervised fuzzy clustering," *Pattern Recognit.*, vol. 41, no. 5, pp. 1834–1844, 2008.

[27] D. Wang, X. Gao, and X. Wang, "Semi-supervised nonnegative matrix factorization via constraint propagation," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 233–244, Jan. 2016.

[28] L. Yang, X. Cao, D. Jin, X. Wang, and D. Meng, "A unified semi-supervised community detection framework using latent space graph regularization," *IEEE Trans. Cybern.*, vol. 45, no. 11, pp. 2585–2598, Nov. 2015.

[29] I. A. Maraziotis, "A semi-supervised fuzzy clustering algorithm applied to gene expression data," *Pattern Recognit.*, vol. 45, no. 1, pp. 637–648, 2012.

[30] I. Ahn and C. Kim, "Face and hair region labeling using semi-supervised spectral clustering-based multiple segmentations," *IEEE Trans. Multimedia*, vol. 18, no. 7, pp. 1414–1421, Jul. 2016.

[31] Z. Yu, H.-S. Wong, J. You, G. Yu, and G. Han, "Hybrid cluster ensemble framework based on the random combination of data transformation operators," *Pattern Recognit.*, vol. 45, no. 5, pp. 1826–1837, 2012.

[32] S. Mimaroglu and E. Aksehirli, "DICLENS: Divisive clustering ensemble with automatic cluster number," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 9, no. 2, pp. 408–420, Mar./Apr. 2012.

[33] Y. Yang and J. Jiang, "Hybrid sampling-based clustering ensemble with global and local constitutions," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 5, pp. 952–965, May 2016.

[34] Z. Yu and H. S. Wong, "Class discovery from gene expression data based on perturbation and cluster ensemble," *IEEE Trans. Nanobiosci.*, vol. 8, no. 2, pp. 147–160, Jun. 2009.

[35] N. Iam-On, T. Boongeon, S. Garrett, and C. Price, "A link-based cluster ensemble approach for categorical data clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 3, pp. 413–425, Mar. 2012.

[36] D. Huang, C. D. Wang, and J. H. Lai, "Locally weighted ensemble clustering," *IEEE Trans. Cybern.*, to be published, doi: 10.1109/TCYB.2017.2702343

[37] D. Huang, J. H. Lai, and C. D. Wang, "Robust ensemble clustering using probability trajectories," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 5, pp. 1312–1326, May 2016.

[38] A. Topchy, A. K. Jain, and W. Punch, "Clustering ensembles: Models of consensus and weak partitions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 12, pp. 1866–1881, Dec. 2005.

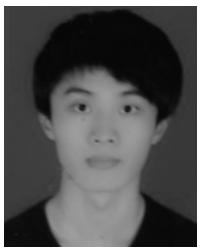
[39] T. Wang, "CA-Tree: A hierarchical structure for efficient and scalable coassociation-based cluster ensembles," *IEEE Trans. Syst. Man Cybern. Part B (Cybern.)*, vol. 41, no. 3, pp. 686–698, Jun. 2011.

[40] L. I. Kuncheva and D. P. Vetrov, "Evaluation of stability of k-means cluster ensembles with respect to random initialization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1798–1808, Nov. 2006.

[41] H. Liu, J. Wu, T. Liu, D. Tao, and Y. Fu, "Spectral ensemble clustering via weighted K-means: Theoretical and practical evidence," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 5, pp. 1129–1143, May 2017.

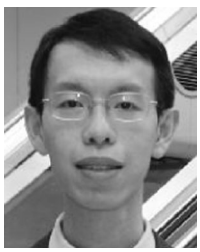
[42] M. Yousefnezhad, S. J. Huang, and D. Zhang, "WoCE: A framework for clustering ensemble by exploiting the wisdom of crowds theory," *IEEE Trans. Cybern.*, vol. 48, no. 2, pp. 486–499, Feb. 2018, doi: 10.1109/TCYB.2016.2642999.
[43] X. Zhang, L. Jiao, F. Liu, L. Bo, and M. Gong, "Spectral clustering ensemble applied to SAR image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2126–2136, Jul. 2008.
[44] H. Li, F. Meng, Q. Wu, and B. Luo, "Unsupervised multiclass region cosegmentation via ensemble clustering and energy minimization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 5, pp. 789–801, May 2014.
[45] P. Kthur, M. Sips, H. Dobslaw, and D. Dransch, "Visual analytics for comparison of ocean model output with reference data: Detecting and analyzing geophysical processes using clustering ensembles," *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 12, pp. 1893–1902, Dec. 2014.
[46] N. Asadi, A. Mirzaei, and E. Haghshenas, "Creating discriminative models for time series classification and clustering by HMM ensembles," *IEEE Trans. Cybern.*, vol. 46, no. 12, pp. 2899–2910, Dec. 2016.
[47] Z. Yu, Z. Kuang, J. Liu, H. Chen, J. Zhang, J. You, H.-S. Wong, and G. Han, "Adaptive ensembling of semi-supervised clustering solutions," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 8, pp. 1577–1590, Aug. 2017.
[48] Z. Yu, P. Luo, J. You, H.-S. Wong, H. Leung, S. Wu, J. Zhang, and G. Han, "Incremental semi-supervised clustering ensemble for high dimensional data clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 3, pp. 701–714, Mar. 2016.
[49] H. Wang, J. Qi, W. Zheng, and M. Wang, "Semi-supervised cluster ensemble based on binary similarity matrix," in *Proc. 2nd IEEE Int. Conf. Inf. Manage. Eng.*, 2010, pp. 251–254.
[50] D. Chen, Y. Yang, H. Wang, and A. Mahmood, "Convergence analysis of semi-supervised clustering ensemble," in *Proc. IEEE 3rd Int. Conf. Inf. Sci. Technol.*, 2013, pp. 783–788.
[51] Z. Yu, H. Chen, J. You, H.-S. Wong, J. Liu, L. Li, and G. Han, "Double selection based semi-supervised clustering ensemble for tumor clustering from gene expression profiles," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 11, no. 4, pp. 727–740, Jul./Aug. 2014.
[52] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.
[53] M. C. de Souto, I. G. Costa, D. S. de Araujo, T. B. Ludermir, and A. Schliep, "Clustering cancer gene expression data: A comparative study," *BMC Bioinf.*, vol. 9, 2008, Art. no. 497, doi: 10.1186/1471–2105-9-497.
[54] M. Lichman, "UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]," Irvine, CA: University of California, School of Information and Computer Science, 2013.
[55] N. X. Vinh, J. Epps, and J. Bailey, "Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance," *J. Mach. Learn. Res.*, vol. 11, pp. 2837–2854, 2010.
[56] S. Basu, A. Banjeree, E. R. Mooney, A. Banerjee, and R. J. Mooney, "Active semi-supervision for pairwise constrained clustering," in *Proc. 4th SIAM Int. Conf. Data Mining*, 2004, pp. 333–344.
[57] Z. Lu and Y. Peng, "Exhaustive and efficient constraint propagation: A graph-based learning approach and its applications," *Int. J. Comput. Vis.*, vol. 103, no. 3, pp. 306–325, 2013.
[58] S. Monti, P. Tamayo, J. Mesirov, and T. Glub, "Consensus clustering: A resamlping based method for class discovery and visualization of gene expression microarray data," *J. Mach. Learn. Res.*, vol. 52, pp. 1–2, 2003.
[59] Y. Yang and J. Jiang, "Bi-weighted ensemble via HMM-based approaches for temporal data clustering," *Pattern Recognit.*, vol. 76, pp. 391–403, 2018.

**Zhiwen Yu** (S'06-M'08-SM'14) received the PhD degree from the City University of Hong Kong, in 2008. He is a professor with the School of Computer Science and Engineering, South China University of Technology, and an adjunct professor with Sun Yat-Sen University. His research interests include focus on data mining, machine learning, bioinformatics, and pattern recognition. Until now, he has published more than 90 referred journal papers and international conference papers, in the *IEEE Transactions on Evolutionary Computation*, the *IEEE Transactions on Cybernetics*, the *IEEE Transactions on Multimedia*, the *IEEE Transactions on Circuits and Systems for Video Technology*, the *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, the *Transactions on NanoBioscience*, the *Journal of Public Relations Research*, the *Information Systems Journal*, the *Bioinformatics*, SIGKDD, and so on. He is a senior member of the IEEE, ACM, CCF (China Computer Federation), and CAAI (Chinese Association for Artificial Intelligence). More information about him can be found at http://www.hgml.cn/yuzhiwen.htm.



**Peinan Luo** received the BSc degree from the South China University of Technology in China. He is working toward the MPhil degree in the School of Computer Science and Engineering, South China University of Technology. His research interests include pattern recognition, machine learning, and data mining.

**Jiming Liu** (F'11) received the MEng and PhD degrees from McGill University. He is currently the chair professor of computer science and the acting dean of science at the Hong Kong Baptist University. His research interests include data analytics, data mining and machine learning, complex network analytics, data-driven complex systems modeling, and health informatics. He has served as the editor-in-chief of the *Web Intelligence Journal* (IOS), and the associate editor of *Big Data and Information Analytics* (AIMS), the *IEEE Transactions on Knowledge and Data Engineering*, the *IEEE Transactions on Cybernetics*, *Neuroscience and Biomedical Engineering* (Bentham), and *Computational Intelligence* (Wiley), among others. He is a fellow of the IEEE.



**Hau-San Wong** received the BSc and MPhil degrees in electronic engineering from the Chinese University of Hong Kong, and the PhD degree in electrical and information engineering from the University of Sydney. He is currently an associate professor with the Department of Computer Science, City University of Hong Kong. He has also held research positions with the University of Sydney and Hong Kong Baptist University. His research interests include multimedia information processing, multimodal human-computer interaction, and machine learning.

**Jane You** is currently a professor with the Department of Computing, Hong Kong Polytechnic University and the chair of the Department Research Committee. She has worked extensively in the fields of image processing, medical imaging, computer-aided diagnosis, and pattern recognition. So far, she has more than 190 research papers published with more than 1,000 non-self citations. She has been a principal investigator for one ITF project, three GRF projects, and many other joint grants since she joined PolyU in late 1998. She is also a team member for two successful patents (one HK patent and one US patent) and three awards including Hong Kong Government Industrial Awards. Her current work on retinal imaging won a Special Prize and Gold Medal with Jury's Commendation at the 39th International Exhibition of Inventions of Geneva (April 2011) and the second place in an international competition (SPIE Medical Imaging'2009 Retinopathy Online Challenge (ROC'2009)). Her research output on retinal imaging has been successfully led to technology transfer with clinical applications. She is also an associate editor of *Pattern Recognition* and other journals.

**Guoqiang Han** received the BSc degree from Zhejiang University, in 1982, and the master's and PhD degrees from Sun Yat-Sen University, in 1985 and 1988, respectively. He is a professor with the School of Computer Science and Engineering, South China University of Technology (SCUT), Guangzhou, China. He is the head of the School of Computer Science and Engineering, SCUT. His research interests include multimedia, computational intelligence, machine learning, and computer graphics. He has published more than 100 research papers.

**Jun Zhang** (M'02-SM'08-F'16) received the PhD degree in electrical engineering from the City University of Hong Kong, Kowloon, Hong Kong, in 2002. Since 2004, he has been with Sun Yat-Sen University, Guangzhou, China, where he is currently a Cheung Kong professor. He has authored seven research books and book chapters, and more than 100 technical papers in his research areas. His current research interests include computational intelligence, cloud computing, big data, and wireless sensor networks. He was a recipient of the China National Funds for Distinguished Young Scientists from the National Natural Science Foundation of China in 2011 and the First-Grade Award in Natural Science Research from the Ministry of Education, China, in 2009. He is currently an associate editor of the *IEEE Transactions on Evolutionary Computation*, the *IEEE Transactions on Industrial Electronics*, and the *IEEE Transactions on Cybernetics*. He is the founding and current chair of the IEEE Guangzhou Subsection and ACM Guangzhou Chapter. He is a fellow of the IEEE.