

Received December 16, 2019, accepted December 20, 2019, date of publication December 31, 2019, date of current version January 28, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2963306

Adaptive Regularized Semi-Supervised Clustering Ensemble

RUI LUO¹, ZHIWEN YU¹, (Senior Member, IEEE), WENMING CAO², CHENG LIU³, HAU-SAN WONG², AND C. L. PHILIP CHEN¹, (Fellow, IEEE)

¹School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China

²Department of Computer Science, City University of Hong Kong, Hong Kong

³Department of Computer Science, Shantou University, Shantou 515063, China

Corresponding author: Zhiwen Yu (zhwyu@scut.edu.cn)

This work was supported by grants in part by the NSFC under Grant 61722205, Grant 61751205, Grant 61572199, Grant 61572540, and Grant U1611461, in part by the National Key Research and Development Program of China under Grant 2018YFB1700300, in part by the Key Research and Development Program of Guangdong Province under Grant 2018B010107002, and in part by the grant from the Guangdong Natural Science Funds under Grant 2017A030312008.

ABSTRACT Although semi-supervised clustering ensemble methods have achieved satisfactory performance, they fail to effectively utilize the constrained knowledge such as cannot-link and must-link when generating diverse ensemble members. In addition, they ignore negative effects brought about by redundancies and noisy data. To address the above shortcomings, in this paper we propose an approach to combine multiple semi-supervised clustering solutions via adaptively regularizing the weights of clustering ensemble members, which is referred to as ARSCE. First, we generate a series of feature subspaces by randomly selecting feature without replacement to avoid the scenario where there are two identical feature subspaces. Second, we conduct feature transformation on the above obtained feature subspaces while considering the pairwise constraints to find new clustering-friendly spaces, where clustering methods are exploited to generate various clustering solutions. Finally, we design a novel fusion strategy to integrate multiple clustering solutions into a unified clustering partition, where weights are designated for each clustering ensemble member. Extensive experiments are conducted on multiple real-world benchmarks, and experimental results demonstrate the effectiveness and superiority of our proposed method ARSCE over other counterparts.

INDEX TERMS Clustering ensemble, semi-supervised clustering, constraint selection, clustering fusion.

I. INTRODUCTION

Clustering, as one of unsupervised learning methods, aims to split data into several disjoint groups, so that data in the same group are more similar than those from different groups. Despite the success of clustering methods in exploring the underlying structure of data, they suffer from the sensitivity of parameter setting, *i. e.*, the clustering performance is not robust and stable. Additionally, with the available of large-scale data collection devices, we are facing a huge amount of data, most of which are high-dimensional. The high dimensionality of these data has posed a challenge to us, since it requires prohibitively expensive hardware to process them. In this setting, random subspace method

provides a possible alternative by randomly selecting features to explore the underlying structure of data. Thus, we bypass the hardware requirement. Besides, inspired by ensemble supervised learning methods, recent years have witnessed the development of clustering ensemble, which is divided into two steps: the generations of clustering solutions and the fusion of clustering solutions. In this first step, it is expected to provide diverse information about the structures of data by adopting various methods, like random feature subspaces, random sample subspaces, random initiation or random feature transformations. Many researchers have focused on generating diverse base clustering solutions. For example, Ye *et al.* [1] handle the data with a limited number of labeled data by designing an improved semi-supervised K-means clustering method. Yu *et al.* [2] combine constraint weights and ensemble member weights to distinguish the

The associate editor coordinating the review of this manuscript and approving it for publication was Chao Tong.

contribution differences between pairwise constraints and ensemble members. Rathore *et al.* [3] fuse random projection and fuzzy c-means into a semi-supervised clustering ensemble framework designed for handling high-dimensional data. Li *et al.* [4] considers labeled data and pairwise constraints in a hybrid constrained semi-supervised clustering scheme in generating base clustering solutions. In the second step, clustering ensemble methods adopt a consensus method for integrating multiple base solutions in order to obtain a better clustering partition. For example, Strehl and Ghosh [5] formulate the clustering ensemble problem as an optimization problem, which is solved based on a hyper-graph model for the fusion of clustering solutions. Yu *et al.* [6] design a Gaussian distribution-based cluster structure, followed by that the most representative unified cluster structures are found to facilitate the process of clustering by using distribution-based distances. Bai *et al.* [7] propose a weighted consensus measure based on information entropy to evaluate the clustering quality.

Although these clustering ensemble methods have achieved satisfactory performance, they seldom consider the issues below: 1) how to fully exploit prior information provided by experts, denoted as must-link and cannot-link constraints, and 2) how to design a better fusion strategy to integrate all the clustering solutions into a more robust and stable solution, compared with each base clustering solution component. To achieve the above two goals, we propose an adaptive regularized semi-supervised clustering ensemble framework, which is referred to as ARSCE. Specifically, we first generate multiple random feature subspaces, followed by performing transformations on these feature subspaces while considering pairwise constraints. Secondly, we adopt clustering algorithm on the transformed features to generate diverse clustering solutions. Thirdly, we design a new fusion strategy to integrate these clustering partitions into a unified clustering solution by assigning suitable weights for clustering ensemble members. To evaluate the effectiveness of our method, we conducted extensive experiments on multiple real-world benchmark data sets. Experimental results indicate that our method ARSCE achieves better or at least comparable performance, compared with other related counterparts, which verifies its effectiveness and superiority.

The contributions of this work are summarized as follows:

- 1) We propose a transformation working in random feature subspaces while considering pairwise constraints for finding a clustering-friendly space, where clustering solutions are generated via using traditional clustering methods.
- 2) We design a strategy to fuse all the clustering solutions into a unified clustering solution by adaptively assigning weights for each clustering ensemble solution member.
- 3) We perform experiments over multiple real-world data sets; experimental results verify the effectiveness of our proposed ARSCE in selecting informative constraints and reducing the constraint redundancies.

The remainder of this work is organized as follows. Section II reviews related works on random subspace methods, semi-supervised clustering and clustering ensemble. Section III illustrates our proposed method in detail. In Section IV we present the experimental results and perform analysis. Section V draws the conclusion of this work and describes the possible future directions.

II. RELATED WORK

In this section, we review the related literatures about random subspace, semi-supervised clustering and clustering ensemble.

Random subspace approaches show remarkable advantage when reducing the relativity between different base classifiers. Practically, there are ensemble classification methods based on random subspace. For example, Ho [8] adopts the pseudo-randomly selected subsets of feature vectors to construct multiple trees in randomly chosen subspaces. To deal with high-dimensional data, Yu *et al.* [9] propose a graph-based semi-supervised dimension reduction scheme in random subspaces and perform semi-supervised linear classification in the random feature subspaces. Semi-supervised clustering approaches have been proposed to improve the clustering performance by using prior of label constraints like “cannot-link” and “must-link”. There are many semi-supervised clustering methods proposed in the past years. Specifically, semi-supervised maximum margin clustering [10] extends the margin maximum framework in supervised learning to clustering and penalizes the violation of the given pairwise constraint conditions, which shows a promising performance. In [11], Anand *et al.* utilize pairwise constraints as supervised information for mean shift clustering, where data are projected into a high-dimensional kernel space, followed by imposing pairwise constraints via a linear transformation on them. In [12], Fang *et al.* integrate low-rank representation together with Gaussian fields into a unified framework, allowing pairwise constraints to guide the construction of the affinity matrix. Liu *et al.* [13] combine K-means clustering and linear discriminant analysis, in which the latter aims to find a space via dimensionality reduction where clustering task can achieve satisfactory performance. Wang *et al.* [14] utilize constraint neighborhood projections to mitigate the issue of constraint conflicts while reducing the required number of labeled data. Xiong *et al.* [15] propose to select both pairwise must-link and cannot-link constraints using active learning strategy for semi-supervised clustering iteratively. Huang *et al.* [16] combine pairwise constraints and constraint projections to achieve both sample and feature constraint projections. Chang and Chen [17] adopt discriminative random fields for evaluating the consistency between the results obtained by certain clustering approach and supervised information derived from pairwise constraints. Wang *et al.* [18] obtain all the labels for unlabeled data via propagating pairwise constraints, where supervision information helps to adjust a weight matrix as a regularization term imposed on the objective function of nonnegative

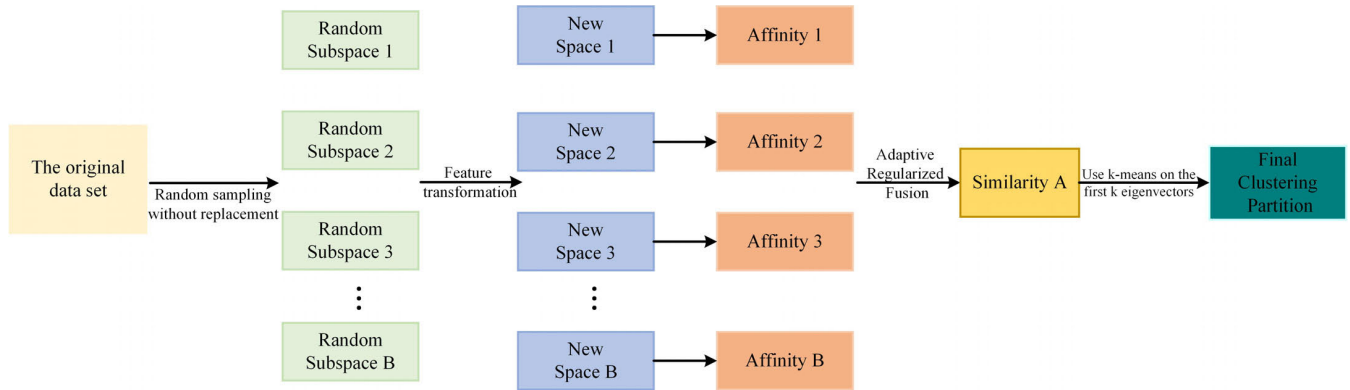


FIGURE 1. The overall framework of our proposed adaptive regularized semi-supervised clustering ensemble.

matrix factorization. Yang *et al.* [19] use the prior about pairwise constraints to improve the community detection performances in a semi-supervised learning framework, where the network topology is integrated with supervised information.

To mitigate these issues, clustering ensemble methods have been proposed to fuse multiple clustering solutions into a unified solution. Specifically, Yu *et al.* [20] adopt random transformations in both the sample and feature spaces with a hybrid strategy. Mimaroglu and Aksehirli [21] design a divisive clustering ensemble scheme by combining high-quality clustering solutions into a final result, which does not require input arguments. Yang and Jiang [22] proposed a sampling-based clustering ensemble approach by adopting boosting and bagging strategies for both the global and the local constitutions respectively. Yu and Wong [23] use the perturbation on input to generate the perturbed data, on which partitions are generated by using Neural Gas as the base clustering method. Iam-On *et al.* [24] design a new link-based method that improves the clustering quality via discovering unknown entries with clusters in an ensemble. Huang *et al.* [25] design an ensemble-driven clustering scheme by estimating the clustering uncertainty and locally weighting co-association matrix. Huang *et al.* [26] perform the ensemble clustering by computing trajectories of random walkers over sparse graph representations, based on which consensus functions help to fuse clustering solutions. Topchy *et al.* [27] convert clustering ensemble as a maximum-likelihood problem and rewrite the consensus function using certain mutual information criterion. Liu *et al.* [28] prove the theoretical equivalence between spectral clustering ensemble and weighted k-means. Yousefnezhad *et al.* [29] adopt the wisdom of crowds like diversity, independencies, decentralization and aggregation to enhance the clustering ensemble. Yu *et al.* [30] integrate random subspace, constraint propagation as well as normalized cut algorithm into a semi-supervised clustering. Yang and Jiang [31] propose a bi-weighted ensemble scheme for time series data clustering via HMM-based K-models to reduce dependencies on initializations while automatically performing model selection. Yu *et al.* [32] propose a double selection-based semi-supervised clustering

method working in both features and samples for tumor clustering, which pairwise constraints are as supervised information. Yu *et al.* [33] design an adaptive semi-supervised clustering ensemble method by fully exploiting pairwise constraints via affinity propagation.

Additionally, researchers have made effects to clustering ensemble, in which they need to trade off between quality and diversity. Shi *et al.* [34] transfer the learnt relationship between quality and diversity in a source domain into a target domain based on certain optimization objective functions. For better performance of spectral clustering over large-scale data with very limited resources, Huang *et al.* [35] convert sparse sub-matrix as a bipartite graph and use transferred cut to obtain the clustering result.

III. PROPOSED METHODOLOGY

In this section, we describe the adaptive Regularized semi-supervised clustering ensemble method (ARSCE) in detail. The framework of our method is shown in Figure 1, which is mainly divided into three stages. First, we randomly select features from all the feature candidates to form a series of subspaces. Second, we conduct a weighted constraint selection and constraint mapping in the above subspaces to contribute to improving the clustering quality. Third, we design a scheme to integrate clustering solutions generated in each subspace for a more robust clustering solution.

A. PROBLEM FORMULATION

Suppose there is a data set $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\} \in \mathbb{R}^{m \times n}$, where n and m are the number of samples and the number of features, semi-supervised clustering ensemble aims to cluster \mathbf{X} into k groups, thereby pairwise samples in the same group are more similar to pairwise samples from different groups and satisfy as many constraints as possible simultaneously.

B. THE DETAILED PROCEDURE OF OUR METHOD

1) THE GENERATION OF RANDOM FEATURE SUBSPACES

Given m features, we generate multiple subspaces by randomly selecting $\lfloor m * \rho \rfloor$ features one by one, where τ is the sampling rate, and $\lfloor \cdot \rfloor$ denotes the greatest integer less than

or equal to a number. This selection procedure is repeated B times, which leads to B random subspaces, denoted as $S = \{S_1, S_2, \dots, S_B\}$. Specifically, the index I_{dx} of a feature selected into a subspace is determined by

$$I_{dx} = 1 + \lfloor \rho (m - 1) \rfloor, \quad (1)$$

where ρ denotes a random variable sampled from a uniform distribution whose range is in the interval between 0 and 1. We adopt sampling without replacement to avoid that features are selected repeatedly in each random subspace. Meanwhile, we avoid two identical feature subspaces whose features are perfectly overlapped. These above obtained subspaces provide various perspectives of exploring the structural information of data while ensuring the diversity of information, which plays a key role in ensemble learning.

2) THE WEIGHTED CONSTRAINT SELECTION AND PROJECTION

Since the generated random subspaces contains different features for exploring the underlying structure of the data manifold, they have their own preferable constraint sets. In this setting, we perform the pairwise constraint selection accordingly. When conducting constraint selection and projections, the assumptions shall be held: 1) a must-link constraint indicates a smaller distance between the corresponding data points, and 2) a cannot-link constraint indicates a larger distance between the corresponding data samples. In the constraint selection process, we consider that a must-link constraint with a large distance or a cannot-link constraint with a small distance plays an important role in the clustering, and we should penalize these two cases.

For a given random feature subspace, the objective function of combining the semi-supervised clustering and searching a clustering-friendly space into a unified framework can be defined as follows:

$$\begin{aligned} \min_{F_i, P} & \operatorname{tr} \left(P^T L P \right) + \alpha \| P - X_{S_i} F_i \|_F^2 + \delta \| F_i \|_{2,1} \\ & + \sum_{(x_i, x'_i) \in \mathcal{M}} \gamma_{i,i'} \mathbf{1}(c_i \neq c'_i) + \sum_{(x_i, x'_i) \in \mathcal{C}} \theta_{i,i'} \mathbf{1}(c_i = c'_i), \\ \text{s.t.}, & P^T P = I_i, \end{aligned} \quad (2)$$

where P is the partition matrix. X_{S_i} is the data matrix in the i^{th} feature subspace. F_i denotes a feature transformation from random subspace to a clustering-friendly space. L denotes the Laplacian matrix. \mathcal{C} and \mathcal{M} denote the cannot-link constraint set and must-link constraint set, respectively. $\gamma_{i,i'}$ and $\theta_{i,i'}$ denote the penalty weights for violating must-link constraint and cannot-link constraints, respectively. $\mathbf{1}(x, y)$ is an indicator function, which equals 1 if $x = y$ and 0 otherwise.

Generally, an ideal transformation F_i enjoys the following properties: 1) allowing a better clustering performance, and 2) meeting the pairwise constraints as many as possible.

3) THE INTEGRATION OF AFFINITIES WITH THE FUSION OF DIFFUSION

Suppose that we have obtained multiple clustering-friendly space, then we compute the corresponding affinity graphs and adopt a regularized ensemble diffusion to fuse the similarity information by

$$\begin{aligned} \min_{\beta, A} & \sum_{v=1}^B \beta_v H^v + \mu \sum_{k,i=1}^n (A_{ki} - I_{ki})^2 + \frac{1}{2} \lambda \| \beta \|_2^2, \\ \text{s.t.}, & 0 \leq \beta_v \leq 1, \quad \sum_{v=1}^B \beta_v = 1. \end{aligned} \quad (3)$$

where

$$H^v = \frac{1}{2} \sum_{i,j,k,l=1}^n W_{ij}^v W_{kl}^v \left(\frac{A_{ki}}{\sqrt{D_{ii}^v D_{kk}^v}} - \frac{A_{lj}}{\sqrt{D_{jj}^v D_{ll}^v}} \right)^2, \quad (4)$$

and λ is a trade-off parameter used to adjust the distribution of the learned weights. I is an identity matrix with an appropriate size. $\beta = \{\beta_1, \dots, \beta_M\}$ is a vector of weights whose v^{th} element β^v is the weight of the v^{th} affinity graph. w_{ij}^v denotes the weight of the edge that connects the i^{th} and the j^{th} data points in the v^{th} graph of the v^{th} subspace. A is the final similarity that not only captures the structure of data but also leverages the complementary information in multiple clustering-friendly spaces. The first term in Eq. (3) evaluates the smoothness of the weighted tensor product graph. The motivation is that if data points x_i and x_j are similar to each other and data points x_k and x_l are similar to each other, there should be a small difference between A_{ki} and A_{lj} . The second term is used to preserve the self-similarity in each learnt affinity graph.

After optimizing Eq. (3), we obtain the similarity $A \in \mathbb{R}^{n \times n}$ that integrates all the discriminative and informative features in clustering-friendly spaces. Then we compute the normalized Laplacian matrix L_{sys} together with its first k eigenvectors, denoted as $U = \{\mu_1, \dots, \mu_k\} \in \mathbb{R}^{n \times k}$, followed by adopting k-means clustering, which is defined as follows:

$$\begin{aligned} \min_{C, P} & \| U - PC \|^2, \\ \text{s.t.}, & P \in \{0, 1\}^{n \times k}, \end{aligned} \quad (5)$$

where $C \in \mathbb{R}^{k \times k}$ and P denote the cluster center matrix and the partition matrix, respectively.

The detailed pseudo-code of how to optimize our proposed method ARSCE is illustrated in Algorithm 1.

IV. EXPERIMENTS

In this section, we evaluate the effectiveness of our method on multiple real-world data sets in terms of normalized mutual information (NMI). The statistical descriptions of these data sets are shown in Table 1. Among these data sets there are 8 data sets collected from GENE repository. Besides, we use 2 data sets collected from the UCI machine learning repository and another 5 data sets collected from the ASU data repository.

Algorithm 1 Adaptive Regularize Semi-Supervised Clustering Ensemble (ARSCE)

Require: input data matrix X , the number of subspaces B , the sampling rate τ , trade-off parameters $\alpha, \delta, \gamma, \theta$ and β and λ ;
Ensure: the clustering partition matrix P .

- 1: Generate B random feature subspaces via Eq. (1) for the i^{th} feature subspace;
- 2: **while** $i \leq B$ **do**
- 3: Compute the feature transformation matrix F_i via optimizing Eq. (2);
- 4: Compute the affinities for the transformed data $X_{S_i}F_i$ in a new space;
- 5: Denote the edge weight matrix as W ;
- 6: **end while**
- 7: Compute the final similarity via optimizing Eq. (3);
- 8: Compute the corresponding Laplacian matrix L_{sys} ;
- 9: Compute the first-k eigenvectors $U = \{u_1, \dots, u_k\}$
- 10: Perform clustering on U by using K-means via optimizing Eq. (5);

TABLE 1. The statistical information of the data sets used to evaluate the clustering performance.

Data set	Source	#Sample	#Features	#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
yale	ASU	165	1024	15

A. EVALUATION CRITERION

We adopt the normalized mutual information (NMI) to evaluate the quality of a clustering result, as defined below:

$$NMI(Y, Y') = \frac{\sum_{l=1}^c \sum_{h=1}^c t_{l,h} \log(\frac{n \times t_{l,h}}{t_l \tilde{t}_h})}{\sqrt{(\sum_{l=1}^c t_l \log \frac{t_l}{n})(\sum_{h=1}^c \tilde{t}_h \log \frac{\tilde{t}_h}{n})}}, \quad (6)$$

where Y and Y' are the ground-true labels and the clustering results, respectively. c denotes the number of clusters. t_l and \tilde{t}_h denote the numbers of samples in the l^{th} ground-truth class and the h^{th} cluster, respectively. $t_{l,h}$ denotes the number of samples in the intersection of the l^{th} ground-truth class and the h^{th} cluster. Generally, a larger NMI value indicates a better clustering performance.

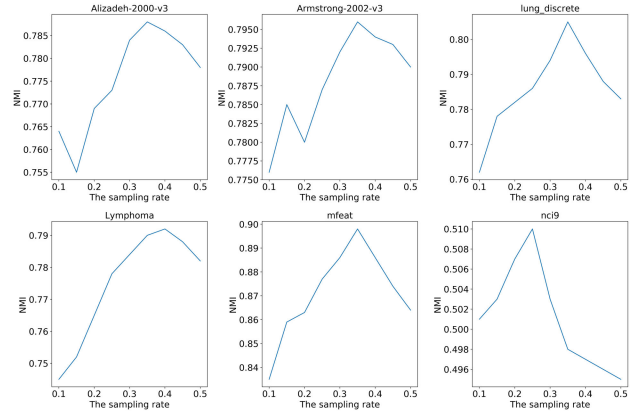


FIGURE 2. The effects of sampling rate on the clustering performance in terms of normalized mutual information (NMI).

B. PARAMETER SETTING

There are eight trade-off parameters in our proposed method ARSCE. For the sampling rate τ , we search the optimal value in the range $\{0.1, 0.2, 0.3, 0.4, 0.5\}$. For α, δ and β , we search the optimal values in the range $\{0.1, 0.2, 0.3, \dots, 0.9\}$, respectively. For γ and θ , we search the optimal values in the range $\{10^{-3}, 10^{-2}, 10^{-1}, 0, 10^1, 10^2, 10^3\}$. Similarly, we search the optimal λ in the interval $\{0.1, 0.2, 0.3, \dots, 0.9\}$. Apart from these trade-off parameters, we need to explore effects of pairwise constraint ratio on performance by searching in the range $\{0.05, 0.1, 0.15, \dots, 0.4\}$. Another factor needed to be considered is the number of ensemble members, and we search the optimal value of ensemble member numbers in the range $\{3, 6, 9, \dots, 24\}$. Concerning parameter settings of comparative counterparts, we choose the optimal values by following the original papers.

C. EXPERIMENTAL ANALYSIS

1) THE EFFECTS OF PARAMETERS

In this part, we first explore effects of sampling rate on clustering performance in terms of normalized mutual information (NMI), where the sampling rate determines the number of features in each subspace. This experiment is conducted over six data sets in Table 1, namely Alizadeh-2000-v3, Armstrong-2002-v2, lung disease, lymphoma, mfeat and nci9. Here the sampling rate varies in the range between 0.1 and 0.5. Figure 2 illustrates effects of sampling rate on clustering performance. From this figure, we observe that in general, performance becomes better with the increase of sampling rate initially. It means that there are increasing informative features selected to facilitate the clustering. However, when the sampling ratio reaches a certain value, we notice that the clustering performance shows an obvious downward trend. A possible reason is that redundant features are selected in this setting, which produces negative effects on clustering. In the majority of cases, the optimal sample

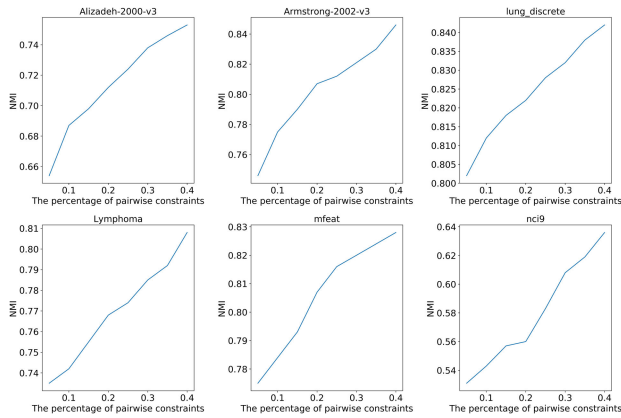


FIGURE 3. The effects of pairwise constraints on the clustering performance in terms of normalized mutual information (NMI).

ratio falls in the range between 0.3 and 0.4, while for the data set nci9, the optimal value of the sampling ratio is in the range between 0.2 and 0.3. In other words, different data have their respective preferable sampling rates. In this setting, we need to choose the optimal sampling ratio dedicatedly. From the perspective of feature selection, we consider that it is necessary to explore a more reasonable strategy when constructing the random feature subspaces by selecting more effective informative features. Thus, it can allow generate multiple diverse clustering partitions with satisfactory performances.

In the following, we would explore the effects of pairwise constraints on clustering performance by increasing the percentages of pairwise constraints. Generally speaking, a larger percentage of pairwise constraints indicates that we have more supervised information to drive clustering methods to find a better clustering performance. Figure 3 illustrates the effects of pairwise constraints on performance over six data sets in Table 1. From this figure, we find that with an increasing number of pairwise constraints available, the performances show upward trends of different levels. It means that these pairwise constraints provide effective supervised information, which contributes to the clustering process when finding a clustering-friendly space.

In addition, we explore the effects of the ensemble member number on clustering performance with respect to normalized mutual information (NMI), which is shown in Figure 4. From this figure, we find that the performance shows an upward trend with the increase of ensemble member numbers. It means that more ensemble members can provide much more informative and complementary information for a better clustering. However, when the number of ensemble members approaches a value, the performance shows a slight improvement, and the improvement gap is diminishing. In this setting, we shall tradeoff between the performance improvement and the computation cost, since an increasing number of ensemble members indicates a larger computation cost and time.

Besides, we study the effects of λ on clustering performance concerning normalized mutual information (NMI),

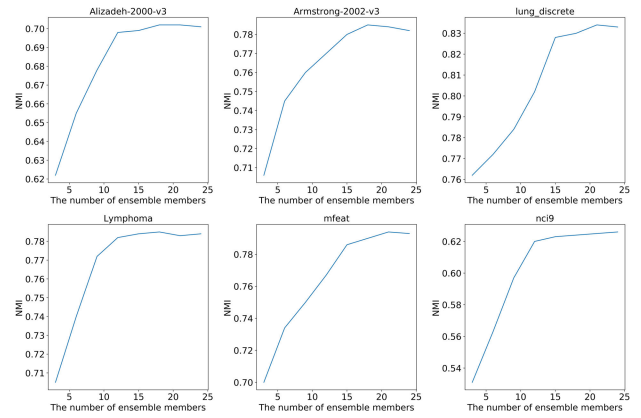


FIGURE 4. The effects of the ensemble member numbers on the clustering performances in terms of normalized mutual information (NMI).

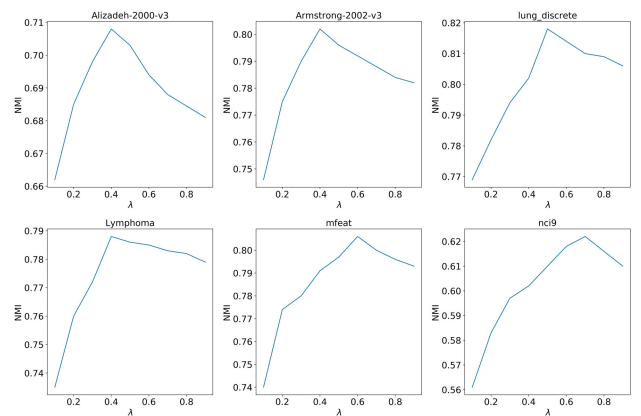


FIGURE 5. The effects of λ on the clustering performances in terms of normalized mutual information (NMI).

which is shown in Figure 5. In this exploration, we change the value of λ in the range between 0.1 and 0.9. From Fig. 5, we notice that when λ increases, the clustering performance shows a quick upward trend before heating the peak, followed by showing downward trends with different extents. Apart from this observation, we find that different data sets have their respective preferable λ . It indicates that our proposed method is sensitive to λ , which is used to control the distributions of the weight affinity graphs for new learnt spaces. Basically speaking, we find that in the majority cases the optimal value of the trade-off parameter λ is in the range between 0.4 and 0.6 except on the data nci9, in which case the optimal λ in the range between 0.6 and 0.8. We consider that there exists some differences between the distribution of data samples over the clusters and those of other data sets. As a result, we need to select its preferable value to adjust the weights of the learnt affinity graph for a better performance.

Next, we analyze the comparative results obtained by recent semi-supervised clustering ensemble approaches and our proposed one. The counterparts include neural gas-based clustering ensemble algorithm (NGCE [36]), random K-means-based clustering ensemble algorithm (RSKE [37]), bagging-based K-means clustering ensemble algorithm

TABLE 2. Comparison of semi-supervised clustering ensemble methods on the data sets in Table 1 in TRMS of normalized mutual information (NMI).

Data set	NGCE	RSKE	BAGKE	HCCE	E ² CPE	ISSCE	DCECP	ARSCE
Alizadeh-2000-v3	0.502±0.007	0.609±0.259	0.502±0.057	0.313±0.115	0.699	0.721±0.014	0.774±0.005	0.793±0.007
Armstrong-2002-v2	0.646±0.059	0.657±0.306	0.689±0.094	0.094±0.055	0.839±0.010	0.854±0.014	0.874±0.038	0.892±0.026
Bredel-2005	0.377±0.070	0.368±0.171	0.375±0.032	0.066±0.021	0.443	0.567±0.003	0.568±0.023	0.583±0.032
Dyrskjot-2003	0.411±0.051	0.537±0.309	0.476±0.085	0.393±0.125	0.662	0.735±0.084	0.765±0.023	0.789±0.023
Nutt-2003-v1	0.515±0.034	0.505±0.190	0.498±0.032	0.341±0.140	0.398±0.017	0.475±0.042	0.644±0.010	0.656±0.043
Pomeroy-2002-v2	0.563±0.043	0.556±0.265	0.470±0.062	0.295±0.108	0.569±0.022	0.641±0.067	0.689±0.071	0.701±0.031
Su-2001	0.523±0.028	0.700±0.156	0.670±0.028	0.398±0.089	0.727	0.741±0.020	0.758±0.027	0.769±0.051
Tomlins-2006-v1	0.455±0.021	0.457±0.164	0.418±0.047	0.239±0.049	0.577±0.010	0.656±0.025	0.722±0.024	0.713±0.026
mfeat	0.529±0.032	0.591±0.047	0.515±0.045	0.346±0.209	0.737	0.857±0.050	0.891±0.040	0.885±0.026
semeion	0.505±0.019	0.500±0.012	0.528±0.039	0.130±0.124	0.645	0.656±0.028	0.704±0.058	0.714±0.056
GLIOMA	0.487±0.067	0.486±0.032	0.477±0.071	0.280±0.107	0.571±0.011	0.613±0.028	0.519±0.043	0.634±0.038
nci9	0.449±0.032	0.456±0.015	0.467±0.036	0.330±0.044	0.549±0.011	0.533±0.015	0.819±0.084	0.809±0.103
Lymphoma	0.663±0.026	0.681±0.024	0.665±0.050	0.392±0.058	0.757±0.014	0.729±0.019	0.756±0.048	0.775±0.201
lung discrete	0.414±0.017	0.679±0.059	0.712±0.042	0.476±0.095	0.733±0.008	0.824±0.012	0.797±0.038	0.819±0.043
yale	0.426±0.018	0.490±0.024	0.497±0.026	0.305±0.034	0.557±0.007	0.546±0.019	0.562±0.055	0.586±0.063

(BAGKE [38]), hierarchical clustering ensemble algorithm (HCCE [39]), exhaustive and efficient clustering ensemble algorithm through constraint propagation (E²CPE [40]), incremental semi-supervised clustering ensemble algorithm (ISSCE [30]), and double weighting semi-supervised ensemble clustering algorithm (DCECP [2]). The comparison results are shown in Table 2, where we do not provide standard deviation whose values are smaller than 0.001. From this table, we have the following observations:

1) E²CPE can achieve better performances, compared with NGCE, RSKE, BAGKE and HCCE, since it uses constraint propagation tricks to leverage the supervised information, which helps to guide the clustering process. It indicates the efficiency of pairwise constraints in boosting the clustering quality.

2) Both the constraint weighting and constraint projection weighting transform feature subspaces into a clustering-friendly space, where high-quality clustering solutions with enough diversity are obtained. It is reflected by the fact that ISSCE and DCECP have achieved much better performances than E2CPE in the majority of data sets.

3) Our proposed method has achieved the best or at least comparable performances on all the data sets, which indicates that it is necessary to adopt an adaptive clustering ensemble via assigning proper weights for base clustering solutions to combine them for a better clustering partition. In other words, it verifies the effectiveness of fusion via diffusion.

V. CONCLUSION

In this paper, we propose a novel constraint-selection based clustering ensemble. First, we design a scheme to learn effective features that are beneficial clustering and meeting the prior clustering constraint conditions. Second, we propose to fuse all the clustering solutions by using the fusion of diffusion rather than the voting mechanism. We conduct experiments on multiple real-world benchmark data sets by comparing other recent related algorithms. Experimental results demonstrate both the effectiveness and the superiority of our proposed method ARSCE. In future, we shall explore

the feasibility of using deep neural networks to learn informative and discriminative features. Besides, we shall improve new strategies of integrating multiple clustering solutions for a better performance. We will further improve our method ARSCE in terms of its sensitivity to the sample ratio. Meanwhile, we shall extend our method ARSCE to real-world application in computer vision fields.

REFERENCES

- [1] Y. Hanmin, L. Hao, and S. Qianting, "An improved semi-supervised K-means clustering algorithm," in *Proc. IEEE Inf. Technol., Netw., Electron. Autom. Control Conf.*, May. 2016.
- [2] Z. Yu, P. Luo, J. Liu, H.-S. Wong, J. You, G. Han, and J. Zhang, "Semi-supervised ensemble clustering based on selected constraint projection," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 12, pp. 2394–2407, Dec. 2018.
- [3] P. Rathore, J. C. Bezdek, S. M. Erfani, S. Rajasegarar, and M. Palaniswami, "Ensemble fuzzy clustering using cumulative aggregation on random projections," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 3, pp. 1510–1524, Jun. 2018.
- [4] X. Li, L. Wang, Y. Song, and X. Zhao, "A hybrid constrained semi-supervised clustering algorithm," in *Proc. 7th Int. Conf. Fuzzy Syst. Knowl. Discovery*, Aug. 2010, pp. 1597–1601.
- [5] 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.
- [6] 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.
- [7] L. Bai, J. Liang, H. Du, and Y. Guo, "An information-theoretical framework for cluster ensemble," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 8, pp. 1464–1477, Aug. 2018.
- [8] I. Barandiaran, "The random subspace method for constructing decision forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 1–22, Aug. 1998.
- [9] G. Yu, G. Zhang, and C. Domeniconi, "Semi-supervised classification based on random subspace dimensionality reduction," *Pattern Recognit.*, vol. 45, no. 3, pp. 1119–1135, 2012.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] 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.

- [14] H. Wang, T. Li, T. Li, and Y. Yang, "Constraint neighborhood projections for semi-supervised clustering," *IEEE Trans. Cybern.*, vol. 44, no. 5, pp. 636–643, May. 2014.
- [15] S. Xiong, J. Azimi, and X. Fern, "Active learning of constraints for semi-supervised clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 1, pp. 43–54, Jan. 2014.
- [16] 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.
- [17] C.-C. Chang and H.-Y. Chen, "Semi-supervised clustering with discriminative random fields," *Pattern Recognit.*, vol. 45, no. 12, pp. 4402–4413, Dec. 2012.
- [18] 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.
- [19] 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.
- [20] Z. Yu, H.-S. Wong, J. You, G. Yu, and G. Han, "Hybrid cluster ensemble framework from the random combination of data transformation operators," *Pattern Recognit.*, vol. 45, no. 5, pp. 1826–1837, May. 2012.
- [21] 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. 2012.
- [22] 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.
- [23] 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.
- [24] 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.
- [25] D. Huang, C.-D. Wang, and J.-H. Lai, "Locally weighted ensemble clustering," *IEEE Trans. Cybern.*, vol. 48, no. 5, pp. 1460–1473, May. 2018.
- [26] 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.
- [27] A. Topchy, A. 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.
- [28] 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.
- [29] 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.
- [30] 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.
- [31] Y. Yang and J. Jiang, "Bi-weighted ensemble via HMM-based approaches for temporal data clustering," *Pattern Recognit.*, vol. 76, pp. 391–403, Apr. 2018.
- [32] 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. 2014.
- [33] 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.
- [34] Y. Shi, Z. Yu, C. L. P. Chen, J. You, H.-S. Wong, Y. Wang, and J. Zhang, "Transfer clustering ensemble selection," *IEEE Trans. Cybern.*, to be published.
- [35] D. Huang, C.-D. Wang, J. Wu, J.-H. Lai, and C. K. Kwok, "Ultra-scalable spectral clustering and ensemble clustering," *IEEE Trans. Knowl. Data Eng.*, to be published.
- [36] Z. Yu, J. You, and G. Wen, "Neural gas based cluster ensemble algorithm and its application to cancer data," in *Proc. Int. Conf. Mach. Learn.*, vol. 1, Jul. 2011, pp. 15–20.
- [37] 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.
- [38] S. Monti, P. Tamayo, J. Mesirov, and T. Glub, "Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data," *J. Mach. Learn. Res.*, vol. 52, pp. 91–118, Jul. 2003.
- [39] 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 Security*, vol. 11, no. 2, pp. 344–357, Feb. 2016.
- [40] 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, Jul. 2013.



RUI LUO received the B.Sc. degree from the School of Computer Science and Engineering, South China University of Technology (SCUT), China, where she is currently pursuing the M.S. degree. Her research interests include machine learning and data mining.



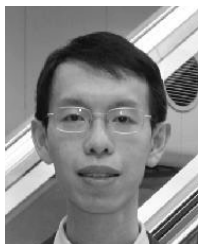
ZHIWEN YU (Senior Member, IEEE) received the Ph.D. degree from the City University of Hong Kong, in 2008. He is currently a Professor with the School of Computer Science and Engineering, South China University of Technology, and also an Adjunct Professor with Sun Yat-sen University. His research areas focus on data mining, machine learning, bioinformatics, and pattern recognition. He has published more than 90 referred journal articles and international conference papers, including TEVC, TCYB, TMM, TCVST, TCBB, TNB, PR, IS, *Bioinformatics*, SIGKDD, and so on. He is a senior member of the ACM, China Computer Federation (CCF), and Chinese Association for Artificial Intelligence (CAAI).



WENMING CAO received M.S. degree from the School of Automation, Huazhong University of Science and Technology (HUST), Wuhan, China, 2015, and the Ph.D. degree in computer science from the City University of Hong Kong, in 2019. His research interests include data mining and machine learning.



CHENG LIU received the Ph.D. degree in computer science from the City University of Hong Kong, Hong Kong, in 2018. He is currently an Assistant Professor with the Department of Computer Science, Shantou University, Shantou, China. His research interests include machine learning and pattern recognition.



processing, multimodal human–computer interaction, and machine learning.

HAU-SAN WONG received the B.Sc. and M.Phil. degrees in electronic engineering from the Chinese University of Hong Kong, and the Ph.D. 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 in the University of Sydney and Hong Kong Baptist University. His research interests include multimedia information



C. L. PHILIP CHEN (Fellow, IEEE) received the M.S. degree from the University of Michigan, Ann Arbor, MI, USA, in 1985, and the Ph.D. degree from Purdue University, West Lafayette, IN, USA, 1988, all in electrical engineering. He is currently the Chair Professor and the Dean of the College of Computer Science and Engineering, South China University of Technology, Guangzhou, China. His current research interests include computational intelligence, systems, and cybernetics.

• • •