Ensemble-Driven Support Vector Clustering:
From Ensemble Learning to Automatic Parameter Estimation

Dong Huang[1], Chang-Dong Wang[2,3,4], Jian-Huang Lai[2,3], Yun Liang[1], Shan Bian[1], Yu Chen[1]

[1]*College of Mathematics and Informatics, South China Agricultural University, Guangzhou, China*
[2]*School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China*
[3]*Guangdong Key Laboratory of Information Security Technology, Guangzhou, China*
[4]*Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education, China*
huangdonghere@gmail.com, changdongwang@hotmail.com, stsljh@mail.sysu.edu.cn,
sdliangyun@163.com, bianshan@scau.edu.cn, chenyu204@scau.edu.cn

Abstract

Support vector clustering (SVC) is a versatile clustering technique that is able to identify clusters of arbitrary shapes by exploiting the kernel trick. However, one hurdle that restricts the application of SVC lies in its sensitivity to the kernel parameter and the trade-off parameter. Although many extensions of SVC have been developed, to the best of our knowledge, there is still no algorithm that is able to effectively estimate the two crucial parameters in SVC without supervision. In this paper, we propose a novel support vector clustering approach termed ensemble-driven support vector clustering (EDSVC), which for the first time tackles the automatic parameter estimation problem for SVC based on ensemble learning, and is capable of producing robust clustering results in a purely unsupervised manner. Experimental results on multiple real-world datasets demonstrate the effectiveness of our approach.

1   Introduction

Support vector clustering (SVC) is a flexible clustering technique which is inspired by support vector machines (SVM) [2]. In SVC, the input data points are first mapped from the original space to a high-dimensional kernel space, where a sphere that encloses most of the data points is constructed. When mapped back to the original space, the sphere is split into several components, each of which encloses a set of data points and forms a cluster in the clustering result. With the nonlinear mapping, SVC has two major advantages over the other clustering techniques. First, it is capable of identifying clusters of arbitrary shapes. Second, the number of clusters can be obtained automatically in SVC and needn't to be specified in advance.

Ben-Hur et al. [2] for the first time introduced the SVC technique, which consists of two phases, i.e., sphere construction and cluster labeling. As both of the two phases are time-consuming, some efforts have been made to improve the efficiency of SVC in recent years. To speedup the sphere construction process, Wang et al. [19] proposed the incremental support vector clustering (ISVC) approach which is able to compute the support vectors (SVs) incrementally and construct the sphere efficiently. Huang et al. [6] extended ISVC by introducing an outlier detection mechanism based on dynamical management of bounded support vectors (BSVs). To improve the efficiency of cluster labeling, Lee and Lee [10] exploited the topological property of the trained kernel radius function to assign cluster labels. Ping et al. [15] introduced an adaptive labeling strategy that decomposes clusters into convex hulls, and proposed a fast and scalable SVC (FSSVC) approach. Besides the efficiency, another focus of the SVC research is on its applications. Wang et al. [20] utilized a SVC-based approach to solve the data stream clustering problem. Boecking et al. [3] combined SVC with a triangular alignment kernel and used it for time series clustering. Sun et al. [17] developed a SVC-based model to describe clusters of images with manually tagged words, which is further exploited to deal with automatic or semi-automatic image annotation problems.

Although many SVC based approaches have been developed in recent years [3, 6, 10, 15, 17, 19, 20], most of the existing approaches, if not all, still suffer from a common drawback, i.e., the difficulty in selecting *proper* parameters. Typically, there are two parameters in SVC, namely, the kernel parameter and the

trade-off parameter. The kernel parameter decides how data points are mapped to high-dimensional space and thereby controls the shapes of cluster contours in the data space. The trade-off parameter adjusts the generation of bounded support vectors (BSVs) which significantly affects the ability for SVC to deal with noisy data. These two parameters together impose a great influence on the clustering performance of SVC. However, the existing SVC approaches [3, 6, 10, 15, 17, 19, 20] generally lack the ability to automatically estimate these two parameters and need to tune the parameters in a supervised manner, i.e., implicitly or explicitly need access to the ground-truth knowledge. To partially address this problem, Wang and Lai [18] proposed a position regularized support vector clustering (PSVC) approach, which eliminates the selection of the trade-off parameter, yet is still unable to select the kernel parameter effectively and unsupervisedly. It remains a very challenging and unsolved problem how to automatically and properly select the two parameters for SVC in an unsupervised manner.

To tackle this problem, in this paper, we propose an ensemble-driven support vector clustering (EDSVC) approach, which, to the best of our knowledge, is the first algorithm that is capable of automatically estimating the aforementioned two crucial parameters for SVC in a purely unsupervised manner. Our approach is based on the ensemble clustering technique [4, 5, 7, 8, 9, 11, 12, 13, 14, 21, 22], which has proved to be a powerful tool to accumulate information from multiple (weak) clusterings. The ground-truth can be viewed as an expert, while the ensemble of multiple clusterings can be viewed as a crowd a individuals. Without the expert, we appeal to the wisdom of the crowd and learn the parameters under the guidance of a crowd of individual clusterings. By exploiting an ensemble learning-based strategy, our approach can automatically estimate the kernel parameter and the trade-off parameter and thereby obtain robust clustering results without supervision. Experimental results on multiple real-world datasets show the effectiveness of the proposed EDSVC approach.

The rest of this paper is organized as follows. The proposed EDSVC algorithm is introduced in Section 2. The experimental results are reported in Section 3. Finally, we conclude this paper and discuss the future work in Section 4.

2 Proposed Framework

In this section, we introduce the proposed clustering framework. First, we introduce the SVC process in Section 2.1. Then, in Section 2.2, we describe the generation of an ensemble of multiple clusterings, which is

further exploited to guide the parameter estimation process. Finally, we summarize the overall algorithm of EDSVC in Section 2.3.

2.1 Support Vector Clustering

Let $\mathcal{X} = \{x_1, \cdots, x_N\}$ be a data set, where $x_i \in \mathbb{R}^d$ is the i-th data point and N is the number of data points in $\mathcal{X}$. By a nonlinear transformation Φ, the data points in $\mathcal{X}$ are mapped to a high-dimensional space and the smallest enclosing sphere with radius R is found. That is

$$\|\Phi(x_i) - \mu\|^2 = R + \xi_i, \forall i = 1, \ldots, N, \quad (1)$$

where μ is the center of the sphere and $\xi_i \geq 0$, for $i = 1, \ldots, N$, are the slack variables which incorporates soft constraints.

To solve the problem in Eq.(1), the Lagrangian is introduced as follows:

$$L = \quad R^2 - \sum_i (R^2 + \xi_i - \|\Phi(x_i) - \mu\|^2)\beta_i$$
$$- \sum \xi_i \alpha_i + C \sum \xi_i, \quad (2)$$

where $\beta_i \geq 0$ and $\alpha_i \geq 0$ are Lagrange multipliers, C is a constant and $C \sum \xi_i$ is the penalty term. By setting the derivative of L w.r.t. R, μ, and ξ_i, respectively, to zero, we have

$$\sum_i \beta_i = 1, \quad \mu = \sum_i \beta_i \Phi(x_i), \quad \beta_i = C - \alpha_i. \quad (3)$$

The KKT complementary conditions result in

$$\xi_i \alpha_i = 0, \quad (R^2 + \xi_i - \|\Phi(x_i) - \mu\|^2)\beta_i = 0. \quad (4)$$

Then, we eliminate the variables R, μ, ξ_i and α_i, and turn the Lagrangian into the Wolfe dual form. That is

$$\max_{\beta_i} W = \sum_i K(x_i, x_i)\beta_i - \sum_{i,j} \beta_i \beta_j K(x_i, x_j) \quad (5)$$

$$s.t. \quad \sum_i \beta_i = 1, \, 0 \leq \beta_i \leq C, \forall i = 1, \ldots, N,$$

where the dot product $\Phi(x_i) \cdot \Phi(x_j)$ is represented by the Gaussian kernel $K(x_i, x_j) = \exp(-q\|x_i - x_j\|^2)$ with width parameter q.

With regard to the values of the Lagrangian multipliers β_i, there are two types of representative points. The points with $0 < \beta_i < C$ are termed support vectors (SVs) which lie on the sphere surface. The points with $\beta_i = C$ are termed bounded support vectors (BSVs) which lie outside the sphere and are usually treated as outliers. Obviously, there will be no BSV when setting $C \geq 1$.

In SVC, the cluster boundaries are delineated by exploiting the SVs, and the data points can be assigned with cluster labels by means of various cluster labeling techniques [2, 10]. After cluster labeling, the final clustering result of SVC can be obtained. Formally, we denote the function of the SVC process as follows:

$$\pi^*_{(q,C)} = SVC(\mathcal{X}, q, C), \tag{6}$$

where $\mathcal{X}$ is the input data, q is the kernel parameter, C is the trade-off parameter, and $\pi^*_{(q,C)}$ is the clustering result by SVC with parameters q and C.

2.2 Ensemble Generation

In the existing SVC-based approaches [3, 6, 10, 15, 17, 19, 20], the kernel parameter and the trade-off parameter are generally selected by some *trial and error* strategy, which explicitly or implicitly needs access to the ground-truth knowledge. Different from these methods, in this paper, we propose to automatically estimate the parameters for SVC in a purely unsupervised manner. Instead of using ground-truth knowledge, we take advantage of the ensemble clustering technique and estimate parameters with the help of a diverse set of clusterings.

Ensemble clustering is a clustering technique that aims to combine multiple clusterings into a probably better and more robust clustering [4, 5, 7, 8, 11, 12, 13, 14, 21, 22]. Each of the input clusterings is referred to as a base clustering, which in fact can be generated by any clustering algorithms. In this paper, we generate the ensemble of multiple base clusterings by k-means. For the generation of each base clustering, the number of clusters for k-means is randomly selected in the interval of $[2, \sqrt[3]{N}]$, where N is the number of data points in the dataset $\mathcal{X}$. Formally, the ensemble of base clusterings is denoted as

$$\Pi = \{\pi^1, \cdots, \pi^M\}, \tag{7}$$

where π^i denotes the i-the base clustering in the ensemble Π, and M denotes the number of base clusterings in Π. Without loss of generality, in this paper, we use $M = 10$ base clusterings. Each base clustering is a partition of the original dataset. With the ensemble generated by random initializations and random number of clusters, a single base clustering (i.e., an ensemble member) may be unstable (or low-quality), but the crowd of diverse members can provide different views of the data and may together provide reliable information about the inherent cluster structures. We argue that the ensemble of multiple clusterings can form a robust

guidance for the process of unsupervised parameter estimation, which will be described in detail and demonstrated by extensive experiments in the following sections of this paper.

2.3 Overall Algorithm

In this paper, we aim to estimate the kernel parameter and the trade-off parameter with the help of an ensemble of multiple clusterings and thereby to achieve robust clustering results in an unsupervised manner. Our work is partially inspired by the idea of "the wisdom of crowds", which is the process of taking into consideration the opinion of a crowd of individuals rather than a single expert in the field of economic and social science. In our approach, without ground-truth labels, we appeal to the crowd of multiple base clusterings to assess the clustering performance of SVC with different parameter settings. Specifically, we evaluate the quality of the output clusterings generated by SVC using different parameters q and C, and aim to find the best parameters automatically. The first task here is to define an unsupervised criterion to evaluate the output clusterings.

The normalized mutual information (NMI) [16] provides a sound indication of the shared information between two clusterings and is widely used for evaluating the quality of clusterings. Generally, the NMI score is computed between a test clustering and the ground-truth clustering so as to evaluate the quality of the test clustering. However, in the unsupervised setting of this work, the ground-truth information is not available. Instead of using the ground-truth, we propose to evaluate the quality of a candidate clustering with the help of an ensemble of multiple clusterings. Let $\pi^*_{(q,C)} = SVC(\mathcal{X}, q, C)$ be a candidate clustering generated by SVC with the kernel parameter q and the trade-off parameter C. The average NMI score is computed between the clustering $\pi^*_{(q,C)}$ and the ensemble Π. That is

$$ANMI(\pi^*_{(q,C)}, \Pi) = \frac{1}{M} \sum_{\pi^i \in \Pi} NMI(\pi^i, \pi^*_{(q,C)}), \tag{8}$$

where $NMI(\pi^i, \pi^*_{(q,C)})$ denotes the NMI score between two clusterings π^i and $\pi^*_{(q,C)}$, and M is the number of base clusterings in Π.

With the evaluation criterion defined in Eq. (8), we can evaluate the clustering results produced by SVC with different parameters. Specifically, we aim to find the optimal parameters $\hat{q}$ and $\hat{C}$ that maximize the average NMI between the ensemble Π and the clustering result produced by SVC. That is

$$(\hat{q}, \hat{C}) = \arg\max_{(q,C)} ANMI(\pi^*_{(q,C)}, \Pi). \tag{9}$$

In this paper, we solve the optimization problem in Eq. (9) by evaluating a set of candidate values. Let $\mathcal{Q} = \{q_1, \cdots, q_{n_q}\}$ be a set of n_q candidate values for the parameter q. Let $\mathcal{C} = \{C_1, \cdots, C_{n_c}\}$ be a set of n_c candidate values for the parameter C. In this paper, we use $n_q = 100$ values for q and $n_c = 100$ values for C. Of the two parameters q and C, one parameter is optimized by fixing the other one, and vice versa. First, by fixing the parameter $C = C_0$, we test the candidate values in $\mathcal{Q}$ for the parameter q and find the best value $\hat{q}$ that optimized Eq. (9). Then, by fixing the parameter $q = \hat{q}$, we test the candidate values in $\mathcal{C}$ for the parameter C and find the optimal value $\hat{C}$. With the optimal parameters obtained by Eq. (9), the final clustering result is generated by SVC using the estimated parameters $\hat{q}$ and $\hat{C}$.

For clarity, the proposed EDSVC algorithm is summarized in Algorithm 1.

Algorithm 1 (Ensemble-Driven Support Vector Clustering)

Input: $\mathcal{X} = \{x_1, \cdots, x_N\}$
1: Generate the ensemble $\Pi = \{\pi^1, \cdots, \pi^M\}$:
    for $i = 1, 2, \cdots, M$
        Select k randomly in the interval of $[2, \sqrt[3]{N}]$.
        Build the i-th base clustering π^i using k-means.
    end for
2: Unsupervised parameter estimation:
    By fixing $C = C_0$, find the optimal $\hat{q}$ in $\mathcal{Q}$ w.r.t. Eq. (9).
    By fixing $q = \hat{q}$, find the optimal $\hat{C}$ in $\mathcal{C}$ w.r.t. Eq. (9).
3: Obtain the clustering result by SVC using the estimated parameters $\hat{q}$ and $\hat{C}$:
    $\tilde{\pi} = SVC(\mathcal{X}, \hat{q}, \hat{C})$
Output: the final clustering $\tilde{\pi}$.

3 Experiments

In this section, we conduct experiments on multiple real-world datasets to evaluate the performance of the proposed EDSVC approach. We first introduce the datasets and the evaluation metric in Section 3.1. Then we compare EDSVC against the base clusterings in the ensemble Π, against four ensemble clustering methods, and against position regularized support vector clustering (PSVC) [18], in Section 3.2, Section 3.3, and Section 3.4, respectively.

3.1 Datasets and Evaluation Metric

In this paper, the experiments are conducted on six real-world datasets, namely, *Wine*, *Ionosphere*, *Breast Cancer* (*BC*), *Banknote Authentication* (*BA*), *Yeast*, and *Semeion*. The six datasets are publicly available at the

Table 1. Description of datasets

Dataset	#Object	#Attribute	#Class
Wine	178	13	3
Ionosphere	351	34	2
BC	683	9	2
BA	1,372	4	2
Yeast	1,484	8	10
Semeion	1,593	256	10

UCI machine learning repository [1]. The details of the benchmark datasets are given in Table 1.

In the experiments, the normalized mutual information (NMI) [16] is used to evaluate the quality of clusterings. The NMI score provides a sound indication of the shared information between clusterings. A higher NMI indicates a better clustering result.

3.2 Comparison Against Base Clusterings

In this paper, we propose the EDSVC approach which bridges the gap between ensemble clustering and support vector clustering. For each benchmark dataset, an ensemble of multiple base clusterings is exploited to guide the SVC process. The purpose is to utilize the information of multiple clusterings to obtain a probably better and more robust clustering result. In this section, we compare the clustering results of EDSVC against the average performance of the base clusterings. As can be seen in Fig. 1, the EDSVC approach is able to produce significantly better clustering results (in terms of NMI) than the base clusterings on all of the six benchmark datasets.
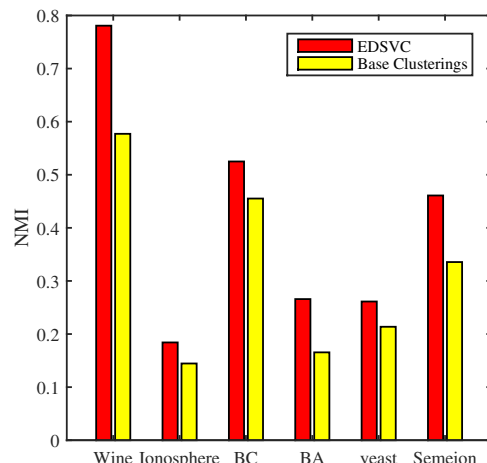

Figure 1. Comparing EDSVC with the base clusterings in terms of NMI.

Table 2. The performance (in terms of NMI) of EDSVC compared against the ensemble clustering approaches. The best score in each column is highlighted in bold.

Method	Wine	Ionosphere	BC	BA	Yeast	Semeion
EDSVC	**0.781**	**0.184**	**0.525**	**0.266**	**0.267**	**0.461**
COMUSA [13]	0.580	0.080	0.219	0.128	0.150	0.122
DICLENS [12]	0.611	0.067	0.369	0.205	0.184	0.326
COMUSACL [14]	0.671	0.164	0.480	0.232	0.204	0.373
COMUSACL-DEW [14]	0.703	0.134	0.494	0.212	0.264	0.344

(a) *Wine*   (b) *Ionosphere*   (c) *BC*

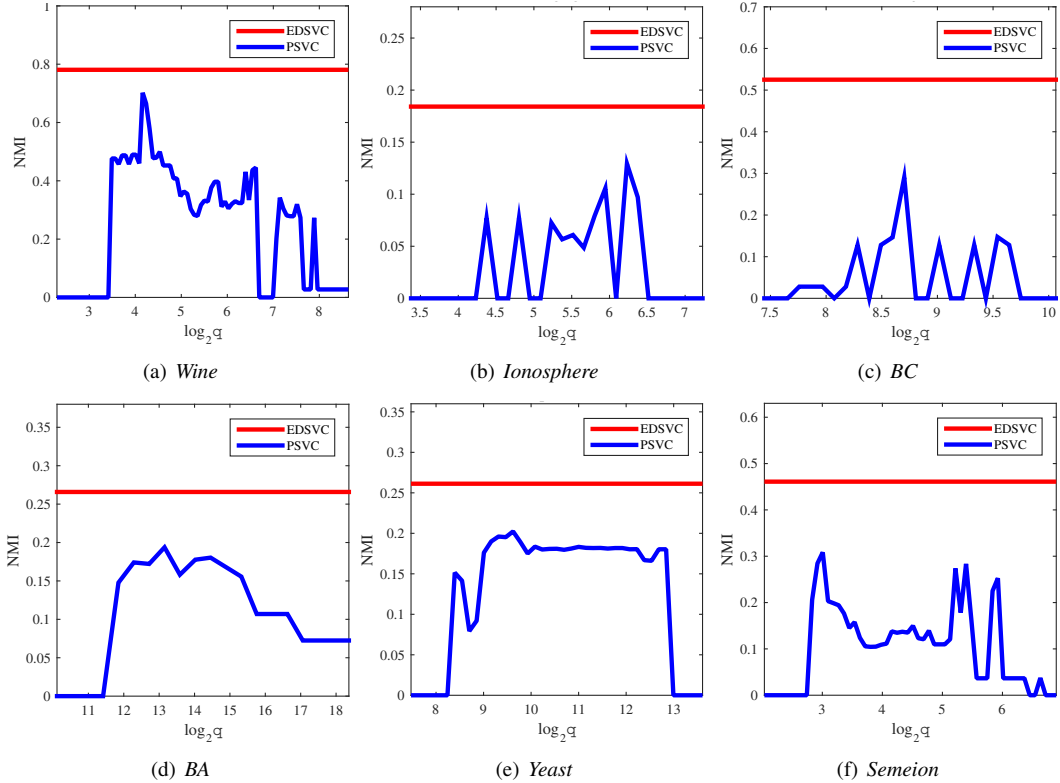(d) *BA*   (e) *Yeast*   (f) *Semeion*

Figure 2. Comparing the performances of EDSVC and PSVC [18] (with varying parameter q) on the benchmark datasets. Note that the X axis corresponds to $\log_2 q$.

3.3 Comparison Against Ensemble Clustering Methods

In this section, we compare EDSVC against four state-of-the-art ensemble clustering approaches, namely, COMUSA [13], DICLENS [12], COMUSACL [14], and COMUSACL-DEW [14]. To provide a fair comparison, on each benchmark dataset, the same ensemble of clusterings are used for both EDSVC and the ensemble clustering approaches. The performances of EDSVC and the four baseline approaches in terms of NMI are reported in Table 2. Our EDSVC approach outperforms the baseline ensemble clustering approaches on all of the six benchmark datasets. Especially, for the *Wine*, *Ionosphere*, *BA*, and *Semeion* datasets, our approach yields significantly better perfor-

mance than the baseline approaches.

3.4 Comparison Against PSVC

In this section, we further compare EDSVC against the position regularized support vector clustering (PSVC) approach [18]. The PSVC approach eliminates the trade-off parameter C, but is still unable to automatically estimate the kernel parameter q. In comparison, our EDSVC approach is capable of automatically estimating q and C and producing robust clustering results. Because PSVC cannot determine the parameter q automatically, we illustrate the NMI scores of our EDSVC and that of PSVC w.r.t. varying q in Fig. 2. As can be seen in Fig. 2, the clustering performance of PSVC is sensitive to the kernel parameter q. Even when the

best q is selected for PSVC on each benchmark dataset, our EDSVC approach can still produce consistently better clustering results than PSVC, which demonstrates the effectiveness of the proposed unsupervised parameter selection strategy and the robustness of the proposed EDSVC approach.

4 Conclusion and Future Work

In this paper, we propose a novel support vector clustering (SVC) approach termed ensemble-driven support vector clustering (EDSVC), which for the first time tackles the problem of unsupervised parameter estimation for SVC based on the ensemble clustering technique. Without needing access to the ground-truth, our approach is able to automatically estimate the kernel parameter q and the trade-off parameter C under the guidance of an ensemble of multiple base clusterings. We conduct experiments on six real-world datasets. The experimental results have demonstrated the superiority of our approach over other ensemble-based and SVC-based approaches. In the future work, we plan to develop a scalable version of EDSVC and enable it to handle the clustering problem of large-scale datasets.

5 Acknowledgement

References

[1] K. Bache and M. Lichman. UCI machine learning repository. Available online: http://archive.ics.uci.edu/ml, 2013.

[2] A. Ben-Hur, D. Horn, H. T. Siegelmann, and V. Vapnik. Support vector clustering. *JMLR*, 2:125–137, 2001.

[3] B. Boecking, S. K. Chalup, D. Seese, and A. S. Wong. Support vector clustering of time series data with alignment kernels. *PRL*, 45:129–135, 2014.

[4] A. L. N. Fred and A. K. Jain. Combining multiple clusterings using evidence accumulation. *IEEE Trans. PAMI*, 27(6):835–850, 2005.

[5] D. Huang, J. Lai, and C.-D. Wang. Ensemble clustering using factor graph. *Pattern Recognition*, 50:131–142, 2016.

[6] D. Huang, J.-H. Lai, and C.-D. Wang. Incremental support vector clustering with outlier detection. In *ICPR*, pages 2339–2342, 2012.

[7] D. Huang, J.-H. Lai, and C.-D. Wang. Combining multiple clusterings via crowd agreement estimation and multi-granularity link analysis. *Neurocomputing*, 170:240–250, 2015.

[8] D. Huang, J.-H. Lai, and C.-D. Wang. Robust ensemble clustering using probability trajectories. *IEEE Trans. KDE*, 28(5):1312–1326, 2016.

[9] D. Huang, J.-H. Lai, C.-D. Wang, and P. C. Yuen. Ensembling over-segmentations: From weak evidence to strong segmentation. *Neurocomputing*, 207:416–427, 2016.

[10] J. Lee and D. Lee. An improved cluster labeling method for support vector clustering. *IEEE Trans. PAMI*, 27(3):461–464, March 2005.

[11] H. Liu, T. Liu, J. Wu, D. Tao, and Y. Fu. Spectral ensemble clustering. In *KDD*, pages 715–724, 2015.

[12] S. Mimaroglu and E. Aksehirli. DICLENS: Divisive clustering ensemble with automatic cluster number. *IEEE/ACM Trans. CBB*, 9(2):408–420, 2012.

[13] S. Mimaroglu and E. Erdil. Combining multiple clusterings using similarity graph. *Pattern Recognition*, 44(3):694–703, 2011.

[14] S. Mimaroglu and E. Erdil. An efficient and scalable family of algorithms for combining clusterings. *EAAI*, 26(10):2525–2539, 2013.

[15] Y. Ping, Y. F. Chang, Y. Zhou, Y. J. Tian, Y. X. Yang, and Z. Zhang. Fast and scalable support vector clustering for large-scale data analysis. *Knowledge and Information Systems*, 43:281–310, 2015.

[16] A. Strehl and J. Ghosh. Cluster ensembles: A knowledge reuse framework for combining multiple partitions. *JMLR*, 3:583–617, 2002.

[17] L. Sun, H. Ge, S. Yoshida, Y. Liang, and G. Tan. Support vector description of clusters for content-based image annotation. *Pattern Recognition*, 47(3):1361–1374, 2014.

[18] C.-D. Wang and J. Lai. Position regularized support vector domain description. *Pattern Recognition*, 46(3):875–884, 2013.

[19] C.-D. Wang, J.-H. Lai, and D. Huang. Incremental support vector clustering. In *ICDM Workshops*, pages 839–846, 2011.

[20] C.-D. Wang, J.-H. Lai, D. Huang, and W.-S. Zheng. SVStream: A support vector based algorithm for clustering data streams. *IEEE Trans. KDE*, 25 (6):1410–1424, 2013.

[21] J. Wu, H. Liu, H. Xiong, and J. Cao. A theoretic framework of k-means-based consensus clustering. In *IJCAI*, 2013.

[22] J. Yi, T. Yang, R. Jin, and A. K. Jain. Robust ensemble clustering by matrix completion. In *ICDM*, pages 1176–1181, 2012.