Computing and Informatics, Vol. 31, 2012, 597–612

AN EFFECTIVE SEMI-SUPERVISED CLUSTERING FRAMEWORK INTEGRATING PAIRWISE CONSTRAINTS AND ATTRIBUTE PREFERENCES

Jinlong Wang

School of Computer Engineering
Qingdao Technological University
266033 Qingdao, China
&
Medical College of Qingdao University
266021 Qingdao, China
e-mail: wangjinlong@gmail.com

Shunyao Wu, Can Wen

School of Computer Engineering
Qingdao Technological University
266033 Qingdao, China
e-mail: {shunyaowu, wencan314}@gmail.com

Gang Li

School of Information Technology
Deakin University
3125 Victoria, Australia
e-mail: gang.li@gmail.com

Communicated by Jozef Kelemen

Abstract. Both the instance level knowledge and the attribute level knowledge can improve clustering quality, but how to effectively utilize both of them is an essential problem to solve. This paper proposes a wrapper framework for semi-supervised clustering, which aims to gracely integrate both kinds of priori knowledge in the clustering process, the instance level knowledge in the form of pairwise constraints and the attribute level knowledge in the form of attribute order preferences. The wrapped algorithm is then designed as a semi-supervised clustering process which transforms this clustering problem into an optimization problem. The experimental results demonstrate the effectiveness and potential of proposed method.

Keywords: Semi-supervised clustering, pairwise, attribute preference

1 INTRODUCTION

When some priori knowledge including preferences about the application domain is available, it becomes relatively easier to find a reasonable clustering for the task at hand. The problem of how to effectively utilize available priori knowledge into a clustering system is referred to as *semi-supervised clustering*, which has attracted considerable research attention in recent years [1, 2, 3, 4, 5, 6, 7, 8].

In general, the semi-supervised clustering aims to guide clustering with available priori knowledge so that "more accurate", "easier to understand" result can be achieved.

As an important kind of instance-level priori knowledge, the pairwise constraints are usually available or can be extracted with minimal effort in many applications. There have been a number of methods for incorporating this kind of information into the clustering process [1, 3, 5]. Wagstaff et al. [1] proposed a semi-supervised clustering algorithm incorporating pairwise constraints as hard constraints in 2001, and experimental results proved constraint-based semi-supervised clustering algorithm could improve clustering quality. Xing et al. [3] constructed an optimization problem by pairwise constraints, which aims to make distances among must-link constraints as small as possible while cannot-link constraints large enough, finally applied new metric to corresponding clustering algorithm in 2002. Halkidi et al. [5] proposed a clustering framework based on subjective criteria in the form of pairwise constraints and objective criteria in the form of clustering objective validity criteria, and provided user interaction.

More recently, another kind of priori knowledge, in the form of attribute-level preferences constraints [9, 10], such as "an attribute a_i is more important than another attribute a_j", has also been incorporated into prototype-based clustering, and the semi-supervised clustering is then transformed into convex optimization problem of finding the most suitable attribute weights [11].

However, when both the instance-level and the attribute-level constraints are available, we need a new method to simultaneously utilize both kinds of priori knowledge into the clustering process, especially in some applications. For example, in document clustering, beside labels or pairwise constraints of documents, users may point out some words are important and they are keywords. Thus, effectively incorporating two main priori knowledges is meaningful and essential. Wang et al. [12]

proposed a clustering method with instance and attribute level side information based on metric learning strategy. In order to meet the requirements of users' interactive in some applications, in this paper, we present a semi-supervised clustering framework for this problem with distance metric and soft constraints. Through the introduction of Bregman Divergences [13], this framework is more flexible. First, the instance-level pairwise constraints are used to obtain an initial attribute weight. Then the attribute-level preferences knowledge are utilized to identify a set of attributes so that most of prior knowledge can be respected by the corresponding clustering result.

This paper is organized as follows. Section 2 provides the basic terminology, and describes the notations used in this paper. Section 3 describes the algorithms related with us. Section 4 presents the novel semi-supervised clustering method. Section 5 validates the effectiveness of our method through experiments. In the end, we conclude the paper in Section 6.

2 SOME NOTATIONS

This section briefly describes the concepts used in this paper. Given a set of n data instances $\mathcal{X} = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ in a d dimensional space, where $\mathbf{x}_i = [x_{i1}, \cdots, x_{id}]^t$, and the desired number of clusters is k. The objective of clustering is then to obtain a partition of $\mathcal{X}$ so that some kinds of clustering metric can be optimized.

Definition 1 (Pairwise Constraints [3])**.** For the instance pair $(\mathbf{x}_i, \mathbf{x}_j)$, the "must-link" set $\mathcal{S}$ and "cannot-link" set $\mathcal{D}$ constraints supplied by users can be defined as follows:

- IF $(\mathbf{x}_i, \mathbf{x}_j)$ belongs to the "must-link" set $\mathcal{S}$, then instances $\mathbf{x}_i$ and $\mathbf{x}_j$ belong to the same cluster;
- IF $(\mathbf{x}_i, \mathbf{x}_j)$ belongs to the "cannot-link" set $\mathcal{D}$, then instances $\mathbf{x}_i$ and $\mathbf{x}_j$ belong to different clusters.

Definition 2 (Attribute Order Preferences [11])**.** The set $\mathcal{P}$ contains a set of attribute order preferences, $p_i = (s_i, t_i, \delta)$ with the constant $\delta > 0$. Here the (s, t, δ) means that the attribute s is more important than the attribute t. Meanwhile, $(s, t, 0)$ denotes that the attribute s has a similar importance as the attribute t.

3 RELATED WORKS

3.1 Xing's Metric Learning Method

Literature [3] applied metric learning to clustering and constructed an optimization problem with respect to the Mahalanobis distance by pairwise constraints, which aimed to make sure distances among must-link constraints as small as possible while cannot-link constraints large enough.

$$\min_{A} \sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{S}} \|\mathbf{x}_i - \mathbf{x}_j\|_A^2$$

subject to:

$$\sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{D}} \|\mathbf{x}_i - \mathbf{x}_j\|_A \geq 1$$

$$A \succeq 0 \tag{1}$$

Through solving this convex optimization, Xing et al. [3] made use of the learned matrix A to obtain the rescaling data instances for $\mathbf{x}_i$ with $A^{\frac{1}{2}}\mathbf{x}_i$.

$$d(\mathbf{x},\mathbf{y}) = d_A(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\| = \sqrt{(\mathbf{x}-\mathbf{y})^t A(\mathbf{x}-\mathbf{y})} \tag{2}$$

When A is diagonal, the Mahalanobis distance (Equation (2)) can be transformed into the Euclidean distance. Xing et al. [3] computed the corresponding Equation (3) of this optimization problem with Newton-Raphson technique.

$$g(A) = g(A_{11},\ldots,A_{dd}) = \sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{S}} \|\mathbf{x}_i-\mathbf{x}_j\|_A^2 - \log\Big(\sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{D}} \|\mathbf{x}_i-\mathbf{x}_j\|_A\Big) \tag{3}$$

3.2 Sun's Semi-supervised Learning Framework

Sun et al. [11] made use of attribute order preferences to construct the following optimization problem. Through an iterative updating procedure similar to the EM algorithm, a satisfactory set of weights for attributes can be obtained.

$$\min_{\{\mathbf{w},\xi\},\{\pi_c\}_{c=1}^k,\{\mu_c\}_{c=1}^k} \frac{1}{n}\sum_{c=1}^{k}\sum_{\mathbf{x}_i\in\pi_c} D_{\mathbf{w}}(\mathbf{x}_i,\mu_c) + \lambda_1\sum_{p\in\mathcal{P}}\xi_p - \lambda_2\widehat{H}(\mathbf{w})$$

s.t.

$$\mathbf{w} \in \Delta d$$

$$w_s - w_t \geq \delta - \xi_p \quad \text{for all} \quad p = (s,t,\delta)\in\mathcal{P}$$

$$\xi_p \geq 0 \quad \text{for all} \quad p\in\mathcal{P} \tag{4}$$

where

- $\Delta d = \{\mathbf{w}\in\Re_+^d \mid \mathbf{w}^t\cdot\mathbf{1}_d = 1\}$,
- $\Re_+^d$ denotes the set of nonnegative real numbers, and
- $\mathbf{1}_d = [\underbrace{1,\cdots,1}_{d}]^t$,
- $\xi = [\xi_p]$ where $p\in\mathcal{P}$.

In the function, the first term is intra-cluster distortion of the clusters$\{\pi_c\}_{c=1}^k$, which is an objective clustering validation index, μ_c is a cluster representative for each cluster π_c; the second term reflects the penalty on the constraints of attribute order preferences, which represents the attribute-level subjective criteria; the third term is a regularization term, which guarantees the consistence of attribute weight.

4 A SEMI-SUPERVISED CLUSTERING FRAMEWORK

For supervised learning, attribute selection is an important process of identifying the most effective attribute set for the learning task. Traditionally, two major categories of methods exist in supervised learning: the wrapper framework, which wraps the learning algorithm inside the attribute selection process and uses the learning performance (i.e. accuracy) to estimate the benefits of adding or removing a particular attribute; the filter framework, which selects an attribute set based on some criterion for any learning algorithm to use before the actual learning is carried out.

For semi-supervised clustering, an attribute is a good one if it is either a good clustering criterion by itself, or good cluster criterion when taken together with some other attributes. In order to arrive at a suitable clustering result, the attributes set which can satisfy user's requirement needs to be identified. However, because of a lacking of consistent performance evaluation criterion, the wrapper framework has yet been fully utilized in semi-supervised clustering.

Based on the observation that when priori knowledge about the application domains is available, the ratio of satisfied priori constraints by the clustering results can be used as the indicator to guide the search for the best attribute set. In this section, we present an effective wrapper framework for semi-supervised clustering. As shown in Figure 1, this framework starts with a metric learning process which utilizes the instance-level pairwise constraints to get a weighted set of attributes. Then two attributes with the largest weights are used as the seeding set, and the wrapper framework attempts to add each unselected attribute until terminated: for each candidate attribute set, a novel process is called to learn a clustering criterion from both the instance-level and the attribute-level constraints, and to produce a clustering result. The attribute which leads to the biggest improvement in performance is selected. This hill-climbing forward selection process iterates until the performance of adding any attribute is less than the performance of the attribute set already selected. Finally the resulted weighted attribute set is used in clustering algorithm to generate the clustering result.

As our wrapper framework attempts to satisfy as much priori knowledge as possible, we define the salient degree as the sum of the proportion of satisfied constraints:

$$salient = \frac{|sat(\mathcal{S})| + |sat(\mathcal{D})|)}{|\mathcal{S}| + |\mathcal{D}|} + \frac{|sat(\mathcal{P})|}{|\mathcal{P}|}$$

where $satat(*)$ denotes the set of satisfied constraints in the set $*$.

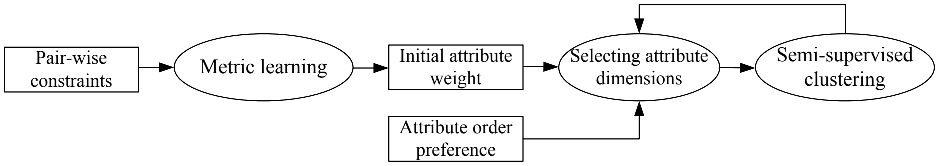


Fig. 1. The framework for semi-supervised clustering

4.1 Initialization of Attribute Weights According to Pairwise Constraints

When instance-level priori knowledge is available as the pairwise constraints $\mathcal{S}$ and $\mathcal{D}$, the metric learning [3] can be used to obtain a distance metric. In order to respect the pairwise constraints as well as maintain the unification of attribute order preferences, a regularization term can be added into distance metric. Thus, this metric learning task can be transformed into an optimization problem as follows:

$$\min_{\omega} \sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{S}} D_\omega(\mathbf{x}_i,\mathbf{x}_j)$$

s.t.

$$\sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{D}} \sqrt{D_\omega(\mathbf{x}_i,\mathbf{x}_j)} \geq 1, \omega \geq 0$$

where

$$D_\omega(\mathbf{x}_i,\mathbf{x}_j) = \sum_{k=1}^{d} \frac{\omega_k}{\upsilon_k} d_\phi(x_{ik}, x_{jk}), \qquad \upsilon_k = \frac{1}{n}\sum_{i=1}^{n} d_\phi(x_{ik}, \overline{\mu}_k)$$

and $d_\phi(.,.)$ is Bregman Divergences [13]. Here, $\mathbf{v}$ is used to regularize distances [11], and $\overline{\mu}$ is the global mean of data points. The MOSEK optimization tool[1] is utlized to solve this problem and produce a weighted attribute set.

4.2 The Wrapper for Semi-Supervised Clustering

The weighted attribute set from the previous step indicates that: the larger the attribute weight, the more important it is for clustering. Starting with two most important attributes according to the weights, the wrapper framework attemptively adds an unselected attribute, then the original attribute-level constraints are mapped into the selected attribute sub-space, and finally each iterative step is ended by another semi-supervised learning process which can utilize both kinds of priori knowledge to generate a clustering result. If the result indicates an improvement in the salient degree, the attribute is then added. The iterative process terminated when no attributes can be added to improve the salient degree of the clustering result.

[1] `http://www.mosek.com/`

4.2.1 Subspace Mapping

For selected attributes in the wrapper process, normalization is applied so that their weights sum to 1. When the selected attribute space does not contain all attributes involved in the attribute preferences constraints $\mathcal{P}$, some attribute-level constraints will become invalid. Therefore, it is necessary to re-establish the attribute-level priori knowledge: for an attribute order preference $p = (s, t, \delta)$; if the attribute s or t is not in the subspace, the corresponding preference is invalid, and we re-establish the attribute order preference information p into $(1, 1, 0)$. If both attributes s and t are selected, then we alter the attribute number in attribute order preferences p to make it accordant with the current attribute space ($s \rightarrow s'$ and $t \rightarrow t'$, denoting that attribute s is corresponding to s', attribute t is corresponding to t' in subspace), and δ is replaced with $\frac{\delta}{sum(\omega_{selected})}$, here $\omega_{selected}$ is the weight of the currently selected attributes.

4.2.2 Wrapped Algorithm: Semi-Supervised Clustering

Wrapped algorithm is a key component in our framework, as shown in Figure 1. In this work, we use a semi-supervised clustering algorithm as the wrapped core algorithm. Similar to the method by Halkidi [5], an optimization problem is constructed to maximize the following objective:

$$\min_{\{\omega,\xi\},\{\pi_c\},\{\mu_c\}} \frac{1}{n} \sum_{c=1}^{k} \sum_{\mathbf{x}_i \in \pi_c} D_w(\mathbf{x}_i, \mu_c) + \lambda_1 \sum_{(s,t,\delta)} max(\delta - (w_s - w_t), 0) - \lambda_2 H(\mathbf{w}) + \lambda_3 S(C)$$

where C is the clustering result, with π_c as the instances set for cluster c and μ_c as the centroid of cluster c.

The first term is the objective clustering validity on the inter-cluster density; the second term is a penalty term for the violation of the attribute order preferences $\mathcal{P}$; the higher the degree of satisfaction with attribute order preferences in clustering, the smaller the penalty value; the third term is a regularization term, which ensures that the weights are as uniform as possible while respecting the preferences; the last term is another penalty term for the violation of the pairwise constraints, $S(C) = 1 - \frac{|sat(\mathcal{S})| + |sat(\mathcal{D})|}{|\mathcal{S}| + |\mathcal{D}|}$.

5 EXPERIMENTAL EVALUATION

Four methods are selected for comparison: the k-means algorithm, Xing's metric learning method which utilizes the pairwise constraints, Sun's semi-supervised learning method which utilizes the attribute constraints and Wang's method which integrates instance and attribute level side information.

Six UCI datasets which have been previously used in semi-supervised clustering, including *iris*, *vowel*, *wdbc*, *pageblocks*, *optdigits* and *pendigits*, are used as our experiment data sets as Table 1.

Dataset	# Examples (n)	# Attribute (d)	# Clusters (k)
iris	1 520	4	3
vowel	990	10	11
wdbc	569	30	2
pageblocks	5 473	10	5
optdigits	5 620	64	10
pendigits	10 992	16	10

Table 1. Dataset characteristics

The *class* information of each data set is used as the ground truth for clustering, then the attribute order preferences are generated as follows [11]: firstly, the within-class distortion Θ_j is calculated for each attribute j $(1 \leq j \leq d)$, $\Theta_j = \frac{1}{v_j}\sum_{c=1}^{k}\sum_{\mathbf{x}_{i\in\pi_c}}(x_{ij}-\mu_{cj})^2$ where $v_j = \frac{1}{n}\sum_{i=1}^{n}(x_{ij}-\frac{1}{n}\sum_{i=1}^{n}x_{ij})^2$. Then, the inverse within-class distortion Γ_j is calculated as $\Gamma_j = \frac{\sum_{l\neq j}\Theta_l}{\Theta_j}$. After that, a rough estimate of the optimal attribute weights is calculated as $\widetilde{w_j} = \frac{\Gamma_j}{\sum_{l=1}^{d}\Gamma_l}$, based on which we can generate a number of attribute preferences constraints. For every dataset in the experiment, we generate 10 different sets of instance-level pairwise constraints by sampling instances from the same cluster or different clusters: each set with $5\,\% \times n$ must-links, and $6\,\% \times n$ cannot-links. Then we generate 10 sets of attribute-level preference constraints, each set contains $\frac{d}{2}$ preferences, where d is the dimensionality.

In the experiments, we set $\lambda_1 = \frac{d}{m}$, $\lambda_2 = d$, $\lambda_3 = 1$ to regulate the value into the range $[0, 1]$. The weighted Euclidean distance $\phi(x) = x^2$ is used in metric learning.

5.1 Cluster Validity Index

We use Clustering Accuracy (ACC) and Normalized Mutual Information (NMI) as the cluster validity indices, same as [11].

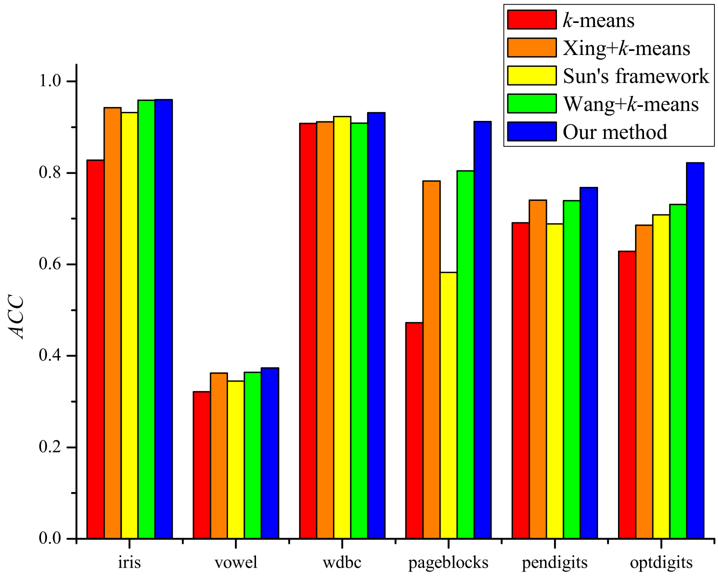
Let $\mathcal{C}$ present the clustering results after applying our approach to X, and $\mathcal{B}$ denote the pre-specified structure. The number of items in $\mathcal{C}$ and $\mathcal{B}$ are both k.

Clustering Accuracy seems like accuracy in classification, which builds a one to one correspondence between the clusters and the classes.

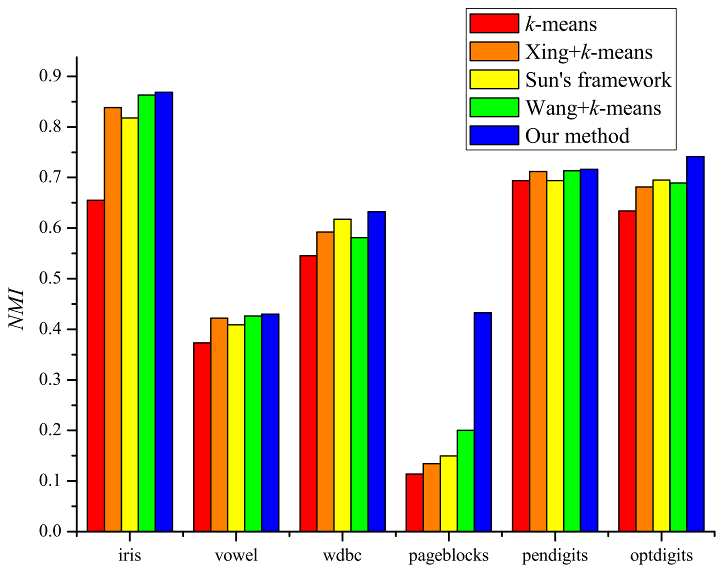
$$Acc(\mathcal{C},\mathcal{B}) = \frac{\max\left(\sum_{i=1}^{k} n_{i,Map(i)}\right)}{n}$$

Here, n stands for number of instances in the dataset, while i stands for the cluster index. $Map(i)$ is class index corresponding to cluster index i, and $n_{i,Map(i)}$ is the number of data points not only belonging to cluster i but to class $Map(i)$.

Normalized Mutual Information is one kind of measure based on information entropy.



a)



b)

Fig. 2. The result comparison of ACC and NMI on six UCI datasets, a) ACC, b) NMI

$$NMI(\mathcal{C}, \mathcal{B}) = \frac{I(\mathcal{C}; \mathcal{B})}{\sqrt{H(\mathcal{C})H(\mathcal{B})}} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{k} n_{ij} \log \frac{n \cdot n_{ij}}{n_i \cdot n'_j}}{\sqrt{\sum_{i=1}^{k} n_i \log \frac{n_i}{n} \sum_{j=1}^{k} n'_j \log \frac{n'_j}{n}}}.$$

Here, H presents the entropy, and I computes the mutual information. We use n_i to express the object number in the i^{th} cluster, n'_j denotes the one in the j^{th} cluster. n_{ij} denotes the item number included in i^{th} and j^{th} cluster.

5.2 Comparison on Cluster Validity Index

The comparison starts with a randomly selected pairwise constraints set, together with a randomly selected attribute preference constraints set; each compared algorithm was running on the same sets of priori knowledge. We run 100 trials, and the average *ACC* and *NMI* were calculated.

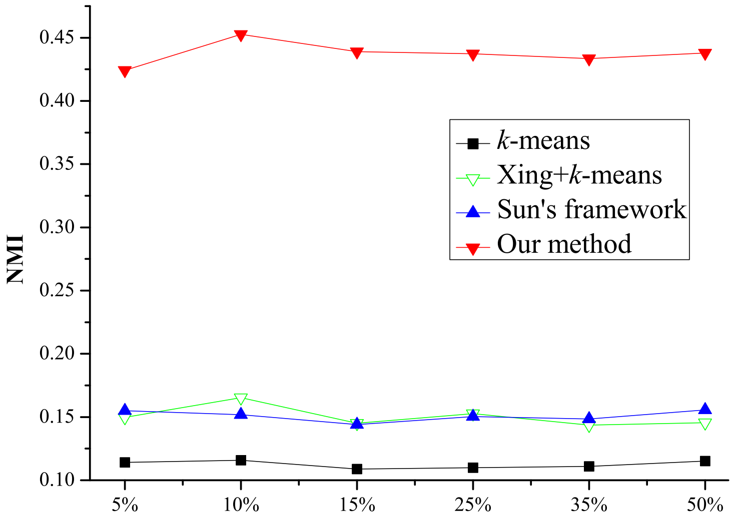
Dataset	k-means	Xing's method	Sun's method	Wang's method
iris	5.6119e−021	0.0683	7.2607e−009	0.5960
vowel	1.4284e−006	0.1739	0.0013	0.2270
wdbc	2.8497e−006	5.2538e−004	0.0364	9.5153e−004
pageblocks	2.5014e−021	1.1882e−008	2.0162e−018	2.2514e−011
pendigits	8.6469e−015	0.0367	4.4041e−012	0.0142
optdigits	2.6210e−005	0.0168	2.9640e−004	0.0010
Statistics	6.0480e−004	0.2581	0.0244	0.3655

Table 2. The t-Test on ACC

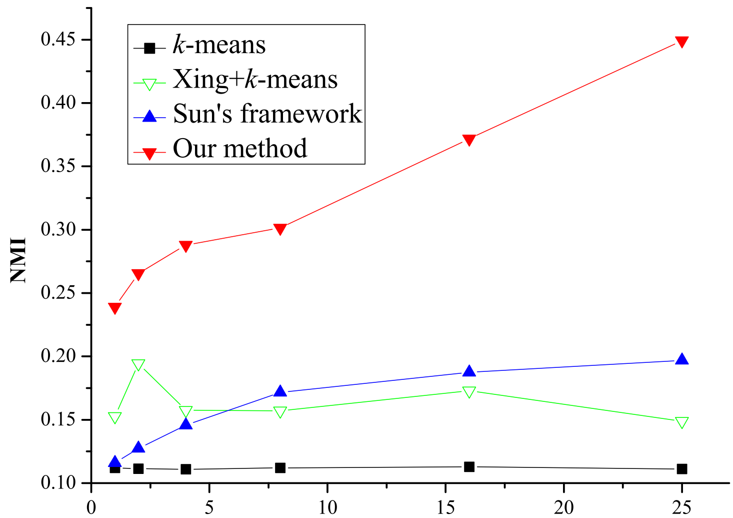
Dataset	k-means	Xing's method	Sun's method	Wang's method
iris	2.7228e−020	0.0494	1.0041e−008	0.3355
vowel	0.0013	0.0032	0.1396	5.1072e−004
wdbc	4.4990e−006	0.0282	0.2885	0.0108
pageblocks	4.9297e−018	2.9924e−013	6.9877e−017	2.2578e−014
pendigits	3.0438e−013	0.2834	4.3073e−011	0.3432
optdigits	7.0695e−005	0.0556	0.0011	4.8727e−004
Statistics	7.2948e−004	0.0972	0.0821	0.1863

Table 3. The t-Test on NMI

Figure 2 gives a comparison in the average *ACC* and *NMI* for the proposed framework, Xing+k-means [3], Sun's framework [11] and Wang+k-means [12]. Obviously we can see that the proposed framework which incorporates two kinds of knowledge effectively outperforms Xing's method, Sun's framework and Wang's method. The accuracy has a small increase over the datasets of *iris*, *vowel*, *wdbc* and *pageblocks*, while there is an significant improvement over the data sets of *optdigits* and *pendigits*. As the attribute order preferences are randomly selected, the quality of attribute preference constraints set is not always fine. Thus, Wang's method based

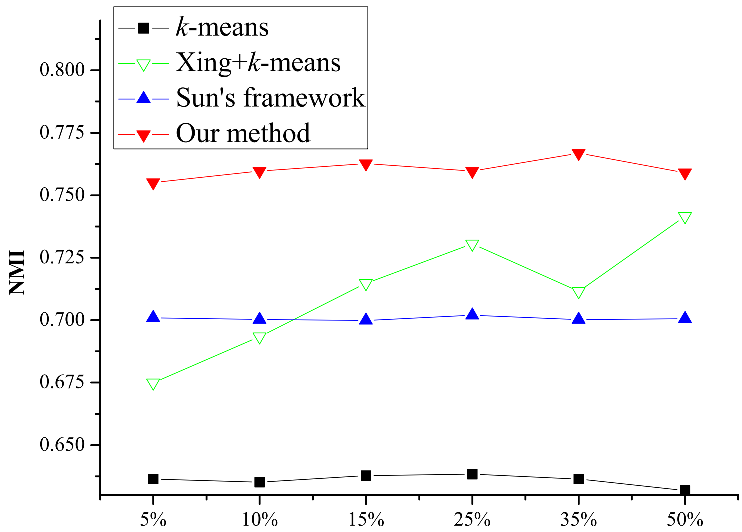


a)

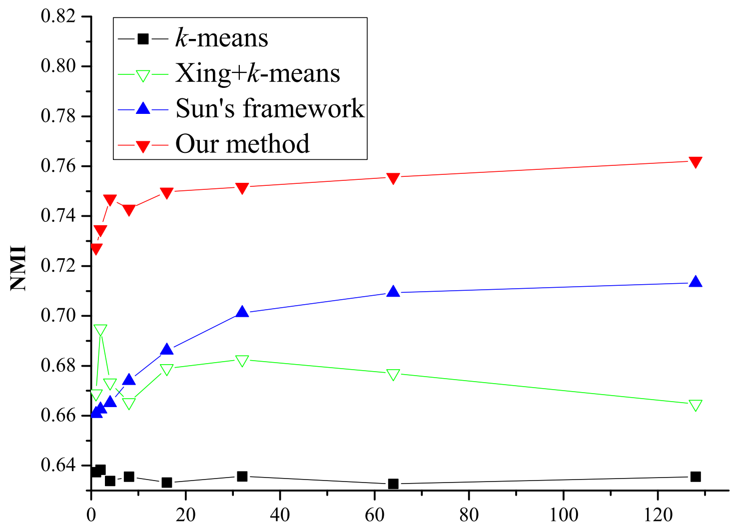


b)

Fig. 3. Clustering accuracy versus different knowledge in pageblocks dataset, a) pairwise, b) attribute order



a)



b)

Fig. 4. Clustering accuracy versus different knowledge in optdigits dataset, a) pairwise, b) attribute order

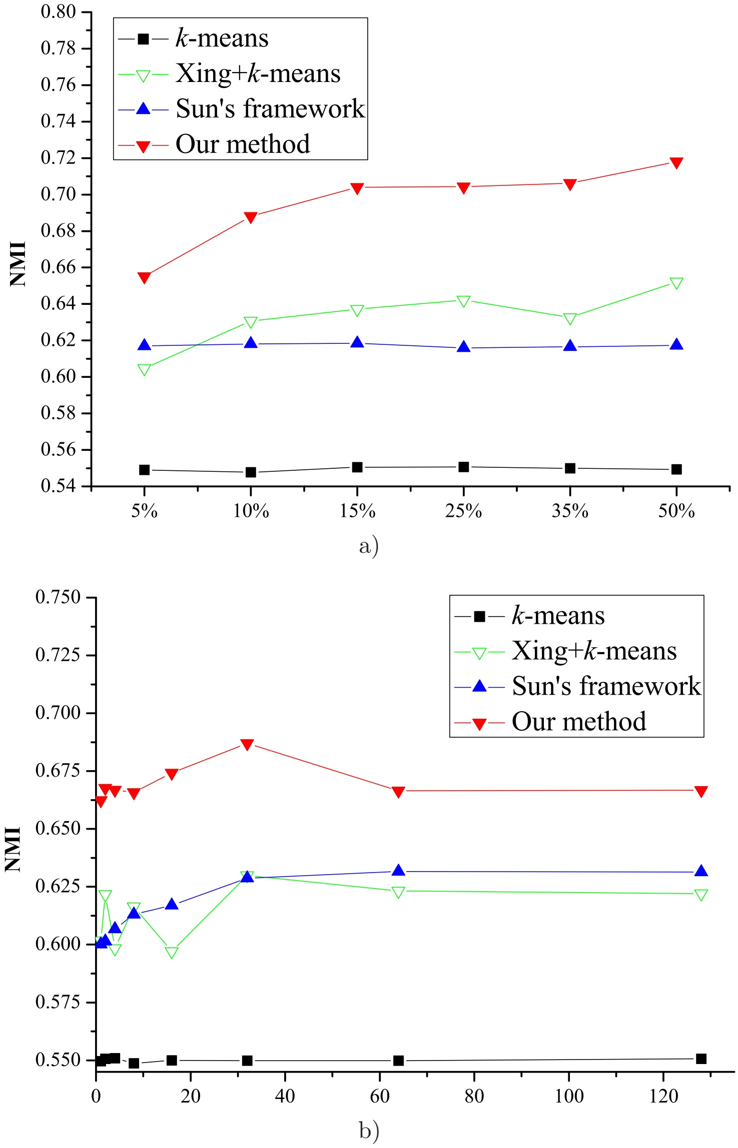


Fig. 5. Clustering accuracy versus different knowledge in wdbc dataset, a) pairwise, b) attribute order

on metric learning is worse than our method which combines distance-based and constraint-based approaches; it demonstrates the semi-supervised clustering combining distance-based and constraint-based approaches is better than those only based on distance-based one.

Tables 2 and 3 are the t-test comparing our approach versus competing methods with ACC and NMI index. The t-test indicates that the improvement by the proposed framework is statistical.

5.3 Clustering Accuracy Versus Knowledge

Here we evaluate the effect of increasing amount of priori knowledge on the clustering accuracy. First, we keep the number of attributes order preferences at $\frac{d}{2}$, while increasing the number of pairwise constraints from $5\,\% \times n$ to $50\,\% \times n$. Second, we keep the number of pairwise constraints at $5\,\% \times n$ pairs of must-links and $6\,\% \times n$ pairs cannot-links, while increasing the number of attribute order preferences from 1 to 128 (25 on pageblocks dataset, because its maximum number of preferences is 25). The experimental results (Figures 3, 4 and 5) show our method with little additional prior knowledge can achieve better clustering quality than Xing's method and Sun's framework.

6 CONCLUSION

This paper presents an effective semi-supervised clustering method for incorporating instance level and attribute level information. This method uses selecting and weighting through incorporating attribute level information into results with pairwise instance level information. The experimental results validate our method.

Acknowledgement

We thank the anonymous reviewers for their detailed and extremely helpful comments.

This work was partially supported by the National Natural Science Foundation of P.R. China (No. 60802066), the China Postdoctoral Science Foundation (No. 20100471494), the Excellent Young Scientist Foundation of Shandong Province of China under Grant (No. 2008BS01009), and Deakin CRGS Grant 2011.

REFERENCES

[1] Wagstaff, K.—Cardie, C.—Rogers, S.—Schrödl, S.: Constrained k-Means Clustering With Background Knowledge. In: ICML '01: Proceedings of the Eighteenth International Conference on Machine Learning, Morgan Kaufmann Publishers Inc., Williams College, Williamstown, MA 2001, pp. 577–584.

[2] Klein, D.—Kamvar, S. D.—Manning, C. D.: From Instance-Level Constraints to Space-Level Constraints: Making the Most of Prior Knowledge in Data Clustering. In: ICML '02 – Proceedings of the Nineteenth International Conference on Machine Learning, Morgan Kaufmann Publishers Inc., Sydney, NSW, Australia 2002, pp. 307–314.

[3] Xing, E. P.–NG, A. Y.—Jordan, M. I.—Russell, S. J.: Distance Metric Learning With Application to Clustering with Side-Information. In: NIPS '02 – Advances in Neural Information Processing Systems 15, MIT Press, Vancouver, British Columbia, Canada 2002, pp. 505–512.

[4] Basu, S.—Bilenko, M.–Mooney, J. R.: A Probabilistic Framework for Semisupervised Clustering. In: KDD '04: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 59–68, ACM, Seattle, WA, USA 2004.

[5] Halkidi, M.—Gunopulos, D.—Vazirgiannis, M.—Kumar, N.—Domeniconi, C.: A Clustering Framework Based on Subjective and Objective Validity Criteria. ACM Transactions on Knowledge Discovery from Data, Vol. 1, 2008, No. 4, pp. 1–25.

[6] Xiang, S.—Nie, F.—Zhang, C.: Learning a Mahalanobis Distance Metric for Data Clustering and Classification. Pattern Recognition, Vol. 41, 2008, No. 12, pp. 3600–3612.

[7] Wang, J.—Wu, S.—Vu, H. Q.—Li, G.: Text Document Clustering With Metric Learning. In: SIGIR '10: Proceeding of the Thirty-third International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, Geneva, Switzerland 2010, pp. 783–784.

[8] Yin, X.—Chen, S.—Hu, E.—Zhang, D.: Semi-Supervised Clustering with Metric Learning: An Adaptive Kernel Method. Pattern Recognition, Vol. 43, 2010, No. 4, pp. 1320–1333.

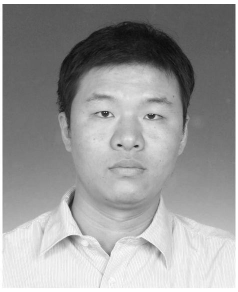
[9] Yu, S.—Yu, K.—Tresp, V.—Kriegel, H. P.: Collaborative Ordinal Regression. In: ICML '06: Proceedings of the Twenty-third International Conference on Machine Learning, ACM 2006, pp. 1089–1096.

[10] Zhu, X.—Goldberg, A.: Kernel Regression With Order Preferences. In: AAAI '07: Proceedings of the Twenty-second AAAI Conference on Artificial Intelligence, AAAI Press 2007, pp. 681–687.

[11] Sun, J.—Zhao, W.—Xue, J.—Shen, Z.—Shen, Y.: Clustering with Feature Order Preferences. In: PRICAI '08: Proceedings of the Tenth Pacific Rim International Conference on Artificial Intelligence, Springer-Verlag, Hanoi, Vietnam 2008, pp. 382–393.

[12] Wang, J.—Wu, S.—Li, G.: Clustering with Instance and Attribute Level Side Information. International Journal of Computational Intelligence Systems, Vol. 3, 2010, No. 6, pp. 770–785.

[13] Banerjee, A.—Merugu, S.—Dhillon, I. S.—Ghosh, J.: Clustering with Bregman Divergences. The Journal of Machine Learning Research, Vol. 6, 2005, pp. 1705–1749.



Jinlong Wang is an Associate Professor at School of Computer Engineering, Qingdao Technological University. He received the Diploma and Ph. D. degree from College of Computer Science and Technology from Zhejiang University, China in 2002 and 2007, respectively. His research interests include data mining, machine learning and artificial intelligence.

Shunyao Wu is currently a master student at School of Computer Engineering, Qingdao Technological University.

Canis Wen currently a Master student at School of Computer Engineering, Qingdao Technological University.

Gangis Li a lecturer in the School of Information Technology, Deakin University. He completed his Ph. D. in data mining in 2005 at Deakin University, received the Bachelor's degree in computer science from Xi'an Petroleum Institute in 1994, and the Master by research degree from Shanghai University of Science and technology in 1997. His research interests include data mining, wireless sensor networks and sentiment mining from multimedia.