Pattern Recognition 106 (2020) 107411

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

New label propagation algorithm with pairwise constraints

Liang Bai[a,b,*], Junbin Wang[b], Jiye Liang[a], Hangyuan Du[b]

[a] Institute of Intelligent Information Processing, Shanxi University, Taiyuan, 030006, Shanxi, China
[b] School of Computer and Information Technology, Shanxi University, Taiyuan, 030006, Shanxi, China

ARTICLE INFO

Article history:
Received 10 May 2019
Revised 17 March 2020
Accepted 29 April 2020
Available online 11 May 2020

Keywords:
Cluster analysis
Semi-supervised clustering
Label propagation algorithm
Pairwise constraint
Spectral clustering

ABSTRACT

The label propagation algorithm is a well-known semi-supervised clustering method, which uses pre-given partial labels as constraints to predict the labels of unlabeled data. However, the algorithm has the following limitations: (1) it does not fully consider the misalignment between the pre-given labels and clustering labels, and (2) it only uses label information as clustering constraints. Real applications not only contain partial label information but pairwise constraints on a dataset. To overcome these deficiencies, a new version of the label propagation algorithm is proposed, which makes use of pairwise relations of labels as constraints to construct an optimization model for spreading labels. Experimental analysis was used to compare the proposed algorithm with 8 other semi-supervised clustering algorithms on 11 benchmark datasets. The experimental results demonstrated that the proposed algorithm is more effective than other algorithms.

1. Introduction

Cluster analysis is an important technique in machine learning and artificial intelligence research [1–3]. The aim of clustering is to partition a set of objects into different clusters such that the objects in the same clusters are highly similar but markedly dissimilar with objects in other clusters. In this regard, various types of clustering algorithms have been proposed and developed, such as partitional clustering [4], hierarchical clustering [5], and density-based clustering [6].

As the clustering process is unsupervised, the cluster structure obtained by a clustering algorithm may not meet user requirements, and may even be irrelevant [7]. Semi-supervised clustering is an important technique to reduce the gap between the clustering result and users' expectations, and makes use of pre-given knowledge about the cluster structure of a dataset to guide the clustering process [8–10]. To date, a number of semi-supervised clustering algorithms have been proposed and reported in the literature (a more detailed review can be found in Section 2). A few of these algorithms have been successfully applied to different fields, such as image segmentation, natural language processing, and social network analysis. Among them, the label propagation (LP) method [9], which is a widely used semi-supervised clustering method, is an extension of graph clustering [11,12]. The method uses partial labels as constraints to ensure that the clustering result has high within-cluster connectivity and that the extent to which these labels differ from pre-given labels is minimized.

However, the performance of the LP algorithm is often affected by several factors, such as the similarity matrix of objects, strategies to solve the optimization problem, and the type of pre-given knowledge. In an attempt to address these factors, several improved LP algorithms have been proposed to enhance the performance in different respects. For example, Wang and Zhang [13], Tang et al. [14] defined different similarity matrices to improve the effectiveness of the LP algorithm. Belkin et al. [15], Zhu et al. [16], 17] added certain regularization terms to the objective function of the LP algorithm with the aim of solving the overfitting problem. Because the LP algorithm can only use positive labels as constraints, Zoidi et al. [18] proposed a new LP algorithm that can accept both positive and negative labels. Lu et al. [19] presented a new LP algorithm with pairwise constraints to solve the inability of the LP algorithm to process pairwise constraints. Even though their algorithm can propagate the pairwise constraints in both the row and column directions, the performance of the algorithm is easily affected by the sparseness of these constraints. In fact, the unavailability of pairwise information for particular data objects could cause propagation errors in the row or column direction, which could also affect other directions. Although they enhanced the performance of the classical LP algorithm to a certain extent, two key issues remained: (1) The LP algorithm is used to process label information, but it does not fully consider the misalignment between pre-given labels and cluster labels; (2) In real applica-

* Corresponding author.
E-mail addresses: bailiang@sxu.edu.cn (L. Bai), wswjbsss@outlook.com (J. Wang), ljy@sxu.edu.cn (J. Liang), duhangyuan@sxu.edu.cn (H. Du).

https://doi.org/10.1016/j.patcog.2020.107411
0031-3203/

tions, different types of prior information are often obtained, such as label information (positive and negative labels) and pairwise information (must-link and cannot-link constraints) [20]. A general LP algorithm capable of working with different types of constraints has not yet been reported.

Our solution to these problems, presented in this paper, is to propose a new label propagation algorithm with pairwise constraints. The proposed algorithm can use the eigenvalue decomposition solution to obtain an optimal result for the label propagation. The proposed algorithm overcomes the label misalignment of the classical LP algorithm and can process pairwise constraints.

The remainder of this paper is organized as follows. Section 2 presents a brief survey of semi-supervised clustering algorithms. Section 3 introduces the classical label propagation algorithm and describes the improved LP algorithm with pairwise constraints. The experimental results are presented in Section 4. The conclusions of this paper are provided in Section 5.

2. Related work

Semi-supervised clustering, which has received considerable attention over the past few years, can incorporate limited prior information, such as a small amount of label information (positive and negative labels) or pairwise constraints (must-link and cannot-link constraints), into clustering algorithms to meet user requirements and to increase the accuracy of data partition.

Depending on the type of prior information, existing semi-supervised clustering algorithms can generally be classified into two categories [21]. The methods in the first category aim to guide the clustering process with label information. Zhou et al. [9] proposed the label propagation algorithm, which is a representative method of graph-based clustering to propagate label information on the graph. Furthermore, several improved LP algorithms [13,14,18] have been proposed to enhance different aspects of clustering effectiveness. However, these improved algorithms cannot consider the misalignment problem between pre-given labels and cluster labels in the process of label propagation. In addition, a number of researchers explored distance metric learning in semi-supervised clustering methods with label information, such as the adaptive kernel method with metric learning [22], information theoretic metric learning [23], and Bregman distance function learning [24]. Basu et al. [25] introduced two semi-supervised variants of k-means clustering that use initial labeled data for seeding. Yu et al. [26] proposed a progressive subspace ensemble learning approach which takes both the feature space and the sample space into account to obtain a more accurate result in terms of semi-supervised clustering. Liu et al. [27] proposed a k-means clustering algorithm with label information. Liu and Wu [28] proposed a matrix factorization method with the label information as additional constraints.

The methods in the second category focus on making use of pairwise constraints to guide the clustering process. Wagstaff et al. [29] devised a semi-supervised variant of k-means named COP-$k$means, which presents a new objective function with penalty terms for violating the pairwise constraints. Yang et al. [30] proposed a constrained self-organizing map ensemble framework to improve COP-k-means. Wei et al. [31] presented a novel semi-supervised clustering ensemble framework based on pairwise constraints and metric learning. Li et al. [32] developed a nonnegative matrix factorization algorithm with pairwise constraints. Kamvar et al. [33] modified the entries of the similarity matrix according to the pairwise constraints and employed spectral clustering to obtain the final result. Similarly, Xu et al. [34] modified the similarity matrix and applied random walk for clustering. Ji and Xu [35], Wang et al. [36] used the constraint matrix as a regularizer to modify the similarity matrix. Kulis et al. [37] used the kernel approach to improve semi-supervised graph clustering. These algorithms are designed to ensure that the clustering result satisfies the constraints as much as possible. However, their performance is easily affected by the size of pairwise constraints, where the pairwise constraints are not propagated in the clustering process. Lu et al. [19] presented a label propagation algorithm with pairwise constraints, which can propagate the pairwise constraints in both the row and column directions. However, the performance of the algorithm is easily affected by the sparseness of pairwise constraints.

To date, semi-supervised clustering methods have been used in many types of practical applications. For example, in the area of image segmentation, Yu et al. [38] proposed a semi-supervised clustering algorithm based on the semantic preservation of distance metric learning for image segmentation. Portela et al. [39] applied semi-supervised clustering for the segmentation of Magnetic Resonance Imaging (MRI) results of the brain. In the area of natural language processing, Huang et al. [40] proposed a semi-supervised document clustering algorithm with language modeling. In the area of social network analysis, Yang et al. [41] proposed a unified semi-supervised framework to integrate the network topology with prior information for community detection. Although these existing methods have already made commendable theoretical and practical contributions, they still have some deficiencies that need to be addressed. This paper presents our proposed extension of the LP algorithm to a new semi-supervised clustering algorithm with pairwise constraints. Compared to existing methods, the proposed algorithm not only solves the misalignment between pre-given labels and cluster labels, but is also able to process different types of prior information.

3. The label propagation algorithm

3.1. The classical label propagation algorithm

Let $X = \{x_1, \ldots, x_l, x_{l+1}, \ldots, x_n\}$ be a dataset with n objects and $L = \{1, \ldots, k\}$ be a set of class labels. The first l objects $x_1, \ldots, x_l$ in X are assumed to be labeled, where $y_i \in L$ for x_i $(i \leq l)$, and other objects are unlabeled. W is an $n \times n$ similarity matrix, where W_{ij} is the similarity between objects x_i and x_j. X can be viewed as a graph $G = (X, W)$, where a node represents an object on the dataset X and an edge weight represents the similarity between its linked nodes. F is an $n \times k$ membership matrix, where k is the number of clusters and $F_{ij} \in [0, 1]$ is the membership of x_i to the jth cluster.

The optimization problem of the LP algorithm [9] is described as

$$\min_F Q(F) = \mu\ S(F) + (1-\mu)\Gamma(F). \tag{1}$$

The objective function $Q(F)$ includes two terms $S(F)$ and $\Gamma(F)$ and μ is a positive parameter to control the trade-off between terms.

$S(F)$ is the objective function of the spectral clustering and is expressed as follows

$$S(F) = \frac{1}{2}\left(\sum_{i,j=1}^{n} W_{ij}\left|\left|\frac{1}{\sqrt{D_{ii}}}F_i - \frac{1}{\sqrt{D_{jj}}}F_j\right|\right|^2\right) = F^T LF. \tag{2}$$

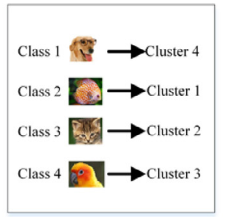
In the definition of $S(F)$, W is computed by a Gaussian kernel function [42]

$$W_{ij} = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right), \tag{3}$$

where σ is set according to the covariance of the given dataset, $D = [D_{ii}]$ is an $n \times n$ diagonal matrix, where $D_{ii} = \sum_{j=1}^{n} W_{ij}$, and

Objects \ Classes	Class 1	Class 2	Class 3	Class 4
x_1	1	0	0	0
x_2	0	1	0	0
x_3	0	1	0	0
x_4	0	0	1	0
x_5	0	0	0	1
x_6	0	0	1	0

Membership matrix Y of objects to classes



Relation between classes and clusters

Objects \ Clusters	Cluster 1	Cluster 2	Cluster 3	Cluster 4
x_1	0	0	0	1
x_2	1	0	0	0
x_3	1	0	0	0
x_4	0	1	0	0
x_5	0	0	1	0
x_6	0	1	0	0

Membership matrix F of objects to clusters

Fig. 1. Misalignment between class labels and cluster labels.

$L = I - D^{-1/2}WD^{-1/2}$ is the normalized Laplacian matrix, where I is an $n \times n$ identity matrix.

$\Gamma(F)$ is a cost function that penalizes the divergence of the cluster labels from the pre-given labels and is defined as

$$\Gamma(F) = ||F - Y||^2 = (F - Y)^T(F - Y). \tag{4}$$

In the definition of $\Gamma(F)$, $Y \in \mathbb{R}^{n\times k}$ is an $n \times k$ membership matrix that is used to save the pre-given label information as follows

$$Y_{ij} = \begin{cases} 1, & \text{if the label of} x_i \text{is} y_j, \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

The LP algorithm aims to minimize $\Gamma(F)$ to reduce the difference between the pre-given labels and the clustering result. The optimal solution of the optimization problem in the LP algorithm can be obtained by the following formula

$$F^* = \beta\ (I - \alpha L)^{-1}Y, \tag{6}$$

where $\beta = \frac{\mu}{1+\mu}$ and $\alpha = \frac{1}{1+\mu}$.

However, the LP algorithm has two key shortcomings:

(1) Misalignment between the pre-given labels and cluster labels may occur during the propagation process. Unlike supervised learning, where the class labels represent specific classes, the cluster labels only express grouping characteristics of the data and are not directly comparable across different clustering results. For example, the example in Fig. 1 is useful to illustrate the misalignment problem. The example contains four pre-given class labels, i.e., Class 1, Class 2, Class 3, and Class 4, and four cluster labels, i.e., Cluster 1, Cluster 2, Cluster 3, and Cluster 4. Y is a membership matrix that represents the relations between objects and class labels. F is a membership matrix that represents the relations between objects and cluster labels. In this example, although the partition of objects described by F is the same as that of Y, the cluster labels do not correspond to the class labels. It is obvious that the labels of different partitions should be aligned. However, the LP algorithm does not consider the alignment problem in $\Gamma(F)$, which may affect label propagation.

(2) In practical applications, the obtained prior information not only includes label information but also pairwise constraints. However, according to the definition of $\Gamma(F)$, it is known that the LP algorithm can only process label information.

These shortcomings motivated us to propose a new LP algorithm, which is described in the next section.

3.2. New label propagation algorithm

To overcome the above key shortcomings of the LP algorithm, the divergence between the pairwise relations of the prior information and clustering result is used as a penalty factor, instead of the difference between partition matrices. Therefore, we first introduce the pairwise relations of prior information (including label information and pairwise constraints).

Label information reflects the membership relation between objects and clusters, which includes both positive and negative labels. The positive label information Y is defined by Eq. (5). The negative labels reflect which class the object does not belong to. Here, $Y^- \in \mathbb{R}^{n\times k}$ is used to represent the negative label information as follows

$$Y^-_{ij} = \begin{cases} -1, & \text{if the label of} x_i \text{is not} y_j, \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

Pairwise constraints reflect the relation between objects, which includes must-link and cannot-link constraints. They can be formalized as follows. Let $M = \{(x_i, x_j) : y_i = y_j, 1 \le i, j \le n\}$ be a set of must-link constraints, and $C = \{(x_i, x_j) : y_i \neq y_j, 1 \le i, j \le n\}$ be a set of cannot-link constraints, where y_i and y_j are the labels of objects x_i and x_j, respectively. We use an $n \times n$ matrix A to represent the pairwise constraints as follows

$$A_{ij} = \begin{cases} 1, & \text{if} (x_i, x_j) \in M, \\ -1, & \text{if} (x_i, x_j) \in C, \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

Based on the definitions of Y, Y^-, and A, the pairwise relation matrix P for different types of prior information can be defined as follows

$$P = \begin{cases} YY^T, & \text{given positive labels,} \\ \frac{1}{k-1}\left(Y^-Y^{-T}\right), & \text{given negative labels,} \\ A, & \text{given pairwise constraints,} \end{cases} \tag{9}$$

where YY^T and $\frac{1}{k-1}\left(Y^-Y^{-T}\right)$ are the pairwise representations of Y and Y^-, respectively, where k is the number of clusters on the dataset. Because negative labels cannot lead to the conclusion that two objects are definitely in the same class, $\frac{1}{k-1}$ is used in the pairwise representation to reflect the probability that two objects belong to the same class.

Based on the definition of P, the penalty function $\Gamma(F)$ of the LP algorithm is redefined as follows

$$\Gamma'(F) = ||FF^T - P||^2, \tag{10}$$

where FF^T is the pairwise representation of F. $\Gamma'(F)$ is the divergence between the pre-given and clustering pairwise representations. The new penalty function can solve the misalignment problem of the pre-given labels and cluster labels. The example in Fig. 1 is again used to show the advantage of the penalty function $\Gamma'(F)$, which is shown in Fig. 2.

According to the figure, YY^T is equal to FF^T. Thus, the use of pairwise matrices can overcome the misalignment problem.

Furthermore, the optimization problem of the label propagation algorithm is modified as follows

$$\begin{aligned} \min_F Q(F) &= \frac{1}{2}\mu\left(\sum_{i,j=1}^{n} W_{ij}\left|\left|\frac{1}{\sqrt{D_{ii}}}F_i - \frac{1}{\sqrt{D_{jj}}}F_j\right|\right|^2\right) + \frac{1}{2}(1-\mu)||FF^T - P||^2 \\ &= \mu\ tr(F^TLF) + \frac{1}{2}(1-\mu)\ tr\left((FF^T - P)^T(FF^T - P)\right). \end{aligned} \tag{11}$$

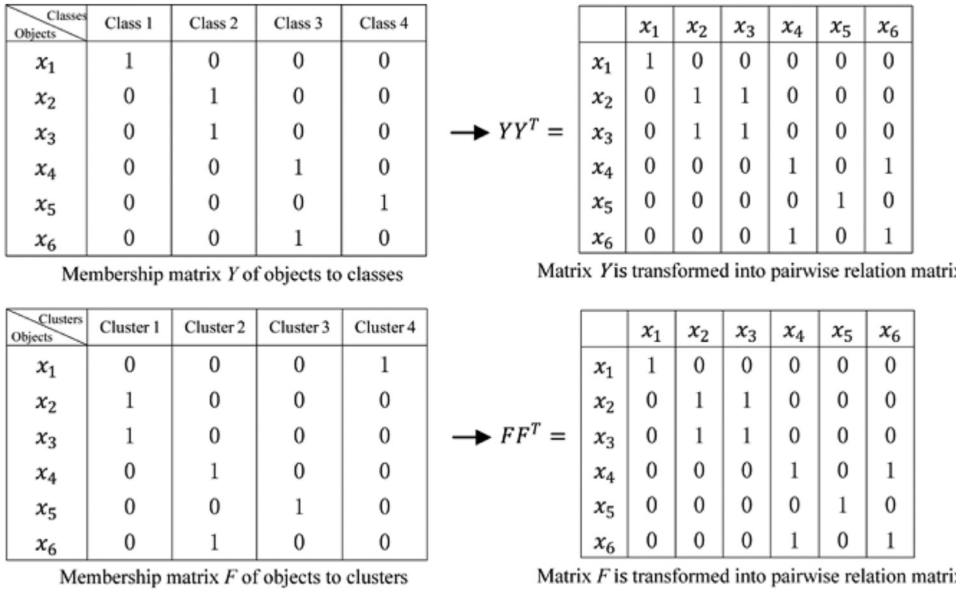


Fig. 2. Procedure of transforming label matrices into pairwise matrices.

In the new objective function, $\Gamma'(F)$ can be converted into the following equation

$$\begin{aligned} & tr\Big((FF^T-P)^T(FF^T-P)\Big) \\ & =2tr(I-F^TPF). \end{aligned} \tag{12}$$

Then,

$$\begin{aligned} Q(F)=&\mu\ tr(F^TLF)+(1-\mu)tr(I-F^TPF) \\ =&tr\Big(F^T(\mu\ L+(1-\mu)(I-P))F\Big). \end{aligned} \tag{13}$$

Because $L'=\mu L+(1-\mu)(I-P)$ can be seen as a new Laplacian matrix, the optimization problem of the new LP algorithm can be converted into a spectral clustering problem. Thus, the eigenvalue decomposition method is used to obtain the optimal solution of $Q(F)$. The proposed algorithm, named NLPPC, is formally summarized in Algorithm 1. Its computational complexity is $O(n^2+n^2k)$,

Algorithm 1 New label propagation with pairwise constraints (NLPPC).

Input: $\mathcal{X}$, k, P, μ.
Output: F^*.

Compute the similarity matrix W of $\mathcal{X}$ by Eq.(3);
Compute the Laplacian matrix $L=I-D^{-1/2}WD^{-1/2}$;
Compute the new Laplacian matrix $L'=\mu L+(1-\mu)(I-P)$;
Compute $F=eigs(L',k)$ to obtain the top k eigenvectors of L';
Get F^* by applying Ward's linkage algorithm to F;
return F^*;

which is as much as that of the original spectral clustering algorithm, where n is the number of objects on a dataset, where $O(n^2)$ is the time cost of constructing the similarity matrix W and $O(n^2k)$ is the time cost of obtaining the top k eigenvectors of L'. In addition, $O(3n^2)$ space is needed in which to save the similarity matrix W, prior information P, and the pairwise matrix of clustering FF^T.

4. Experimental analysis

4.1. Experimental setup

The experiment was designed to test the proposed algorithm on 11 benchmark datasets, which can be downloaded from http://archive.ics.uci.edu/ml. Details of these datasets appear in Table 1.

Two widely used validity indices were employed: the Normalized Mutual Information (*NMI*) [43] and Adjusted Rand Index (*ARI*) [44] to evaluate the performance of the proposed algorithm. These indices attempt to measure the similarity between the ground-truth partition and the clustering result on a dataset. Given a dataset X with N objects and two partitions of these objects, namely $\mathbb{C}=\{c_1,c_2,\cdots,c_k\}$ (the clustering result) and $\mathbb{P}=\{p_1,p_2,\cdots,p_k\}$ (the true partition). Let $n_{ij}=|c_i\cap p_j|$ be the number of common nodes of groups c_i and p_j, $b_i=\sum_{j=1}^N n_{ij}$ and $d_j=\sum_{i=1}^N n_{ij}$. The normalized mutual information (NMI) [43] is defined as

$$NMI=\frac{2\sum_i\sum_j n_{ij}\log\frac{n_{ij}N}{b_id_j}}{-\sum_i b_i\log\frac{b_i}{N}-\sum_j d_j\log\frac{d_j}{N}}.$$

The adjusted rand index [44] is defined as

$$ARI=\frac{\sum_{ij}\binom{n_{ij}}{2}-[\sum_i\binom{b_i}{2}\sum_j\binom{d_j}{2}]/\binom{N}{2}}{\frac{1}{2}[\sum_i\binom{b_i}{2}+\sum_j\binom{d_j}{2}]-[\sum_i\binom{b_i}{2}\sum_j\binom{d_j}{2}]/\binom{N}{2}}.$$

Table 1
Experimental datasets.

Dataset	#Instances	#Features	#Classes
Iris	150	4	3
Wine	178	13	3
Heart	270	13	2
ORL	400	1024	40
Breast	569	31	2
Messidor	1151	20	2
Bank	1372	4	2
Isolet	1560	617	26
Digits	5620	64	10
Statlog	6435	36	6
MNIST	10000	784	10

Table 2
NMI values for semi-supervised clustering methods with positive labels.

Datasets	percent	LP	LNP	CNMF	PLCC	NLPPC	SC
	15%	0.8580	0.7784	0.6250	0.6826	**0.8797**	
Iris	25%	0.8610	0.8440	0.6795	0.7455	**0.8790**	0.7777
	35%	0.8924	0.8629	0.7411	0.8267	**0.9222**	
	15%	0.8349	0.7446	0.7084	0.8156	**0.9045**	
Wine	25%	0.8626	0.8450	0.7798	0.8705	**0.9327**	0.8609
	35%	0.9258	0.8715	0.8159	0.8958	**0.9423**	
	15%	0.3492	0.1429	0.2343	0.3211	**0.3661**	
Heart	25%	0.4196	0.2236	0.3242	0.3740	**0.4351**	0.2765
	35%	0.4808	0.3735	0.4385	0.4767	**0.5172**	
	15%	**0.8242**	0.7653	0.2045	0.7253	0.8230	
ORL	25%	0.8783	0.8441	0.2045	0.7228	**0.8840**	0.8909
	35%	0.9248	0.8939	0.5795	0.7458	**0.9401**	
	15%	0.7227	0.6808	0.5108	0.5912	**0.7579**	
Breast	25%	0.7597	0.7296	0.5697	0.6945	**0.7813**	0.6305
	35%	**0.8294**	0.8053	0.5883	0.7330	0.8146	
	15%	1.0000	0.9156	0.6545	0.7092	**1.0000**	
Messidor	25%	1.0000	0.9482	0.8581	1.0000	**1.0000**	0.3405
	35%	1.0000	0.9850	0.9389	1.0000	**1.0000**	
	15%	0.9600	0.6419	0.1161	0.0908	**0.9784**	
Bank	25%	0.9523	0.7647	0.1434	0.1350	**0.9871**	0.2131
	35%	0.9706	0.8908	0.2527	0.2597	**0.9888**	
	15%	0.8428	0.7874	0.1096	0.7788	**0.8539**	
Isolet	25%	0.8778	0.8351	0.1588	0.8091	**0.8800**	0.8073
	35%	0.9050	0.8564	0.2013	0.8162	**0.9167**	
	15%	0.9643	0.9521	0.0552	0.7359	**0.9665**	
Digits	25%	0.9704	0.9617	0.1533	0.7891	**0.9716**	0.8914
	35%	0.9756	0.9634	0.2949	0.7774	**0.9767**	
	15%	**0.7718**	0.7461	0.5823	0.5879	0.7677	
Statlog	25%	**0.8075**	0.7933	0.5778	0.6311	0.8052	0.6658
	35%	0.8259	0.8204	0.6109	0.6956	**0.8312**	
	15%	0.8922	0.8501	0.1007	0.5586	**0.9015**	
MNIST	25%	0.9117	0.8776	0.2029	0.6214	**0.9163**	0.6655
	35%	0.9306	0.9045	0.2996	0.6446	**0.9321**	
ALL Datasets	Average	0.8479	0.7848	0.4338	0.6625	**0.8622**	0.6382

Table 3
ARI values for semi-supervised clustering methods with positive labels.

Datasets	percent	LP	LNP	CNMF	PLCC	NLPPC	SC
	15%	0.8652	0.7925	0.5616	0.6464	**0.9222**	
Iris	25%	0.8816	0.8670	0.6033	0.7021	**0.9037**	0.7445
	35%	0.9040	0.8847	0.6890	0.8346	**0.9799**	
	15%	0.8504	0.7327	0.6963	0.8307	**0.9167**	
Wine	25%	0.8817	0.8554	0.7766	0.8907	**0.9651**	0.8708
	35%	**0.9408**	0.8877	0.8401	0.9232	0.9295	
	15%	0.4276	0.1569	0.2893	0.4070	**0.4424**	
Heart	25%	0.5170	0.2776	0.4036	0.4663	**0.5918**	0.3576
	35%	0.5868	0.4676	0.5418	0.5819	**0.6384**	
	15%	0.5356	0.3417	0.0048	0.3357	**0.5828**	
ORL	25%	**0.6613**	0.5358	0.0048	0.3309	0.6448	0.7016
	35%	0.7921	0.7085	0.1619	0.3782	**0.8799**	
	15%	0.8214	0.7781	0.5606	0.6969	**0.8501**	
Breast	25%	0.8457	0.8228	0.6394	0.7941	**0.9034**	0.7114
	35%	**0.9034**	0.8800	0.6765	0.8243	0.9033	
	15%	1.0000	0.9456	0.6909	0.7242	**1.0000**	
Messidor	25%	1.0000	0.9716	0.8677	1.0000	**1.0000**	0.3554
	35%	1.0000	0.9931	0.9608	1.0000	**1.0000**	
	15%	0.9779	0.6872	0.1579	0.1243	**0.9942**	
Bank	25%	0.8336	0.8240	0.1904	0.1840	**0.9942**	0.1321
	35%	0.9855	0.9356	0.3299	0.3352	**0.9942**	
	15%	**0.7019**	0.6000	0.0049	0.5586	0.6028	
Isolet	25%	0.7702	0.6963	0.0191	0.6211	**0.8096**	0.5590
	35%	0.8256	0.7346	0.0423	0.6242	**0.8257**	
	15%	0.9695	0.9569	0.0208	0.6453	**0.9714**	
Digits	25%	0.9753	0.9668	0.0488	0.7206	**0.9779**	0.8321
	35%	0.9795	0.9681	0.1048	0.7050	**0.9819**	
	15%	**0.7970**	0.7759	0.5135	0.5046	0.7892	
Statlog	25%	0.8325	0.8256	0.4950	0.5626	**0.8426**	0.6505
	35%	0.8518	0.8515	0.5489	0.6498	**0.8659**	
	15%	0.8999	0.8556	0.0543	0.4462	**0.9177**	
MNIST	25%	0.9192	0.8866	0.0804	0.5459	**0.9271**	0.5166
	35%	0.9424	0.9151	0.1191	0.5672	**0.9457**	
ALL Datasets	Average	0.8430	0.7691	0.3848	0.6110	**0.8662**	0.5847

Table 4
NMI values for semi-supervised clustering methods with pairwise constraints.

Datasets	percent	COP	CVQE	PCLP	NLPPC
	15%	0.7337	0.7289	0.7629	**0.8225**
Iris	25%	0.7554	0.6983	0.7723	**0.8196**
	35%	0.7855	0.7187	0.7771	**0.8742**
	15%	0.8282	0.8144	**0.8971**	0.8949
Wine	25%	0.8532	0.8030	0.8894	**0.9452**
	35%	0.8868	0.7803	0.8806	**0.9505**
	15%	0.3286	0.2676	0.2432	**0.3409**
Heart	25%	0.3791	0.2003	0.2478	**0.4451**
	35%	0.4239	0.1731	0.2659	**0.4864**
	15%	0.7388	0.3174	**0.7766**	0.6856
ORL	25%	0.7335	0.3634	**0.7796**	0.7778
	35%	0.7314	0.3724	0.7814	**0.8526**
	15%	0.6742	0.6223	0.6793	**0.7455**
Breast	25%	0.7377	0.6207	0.6918	**0.7900**
	35%	0.7654	0.6243	0.7105	**0.8313**
	15%	0.3401	0.2016	1.0000	**1.0000**
Messidor	25%	0.7139	0.2515	1.0000	**1.0000**
	35%	1.0000	0.3009	1.0000	**1.0000**
	15%	0.0286	0.0219	0.8942	**0.9553**
Bank	25%	0.1342	0.0182	0.9317	**0.9635**
	35%	0.3065	0.0164	0.9390	**0.9743**
	15%	0.7538	0.5253	0.7986	**0.8061**
Isolet	25%	0.7671	0.4920	0.8001	**0.8149**
	35%	0.7734	0.5091	0.8017	**0.8260**
	15%	0.7328	0.7393	0.9201	**0.9451**
Digits	25%	0.7652	0.7529	0.9391	**0.9486**
	35%	0.7769	0.7435	0.9420	**0.9578**
	15%	0.6164	0.6072	0.6496	**0.6815**
Statlog	25%	0.6236	0.5946	0.6616	**0.6966**
	35%	0.6354	0.5797	0.6425	**0.7213**
	15%	0.5101	0.5162	**0.6934**	0.6338
MNIST	25%	0.5201	0.5126	0.7085	**0.7142**
	35%	0.5158	0.4982	0.7113	**0.7204**
ALL Datasets	Average	0.6324	0.4844	0.7633	**0.8067**

Table 5
ARI values for semi-supervised clustering methods with pairwise constraints.

Datasets	percent	COP	CVQE	PCLP	NLPPC
	15%	0.7205	0.7069	0.7339	**0.8137**
Iris	25%	0.7427	0.5987	0.7391	**0.7922**
	35%	0.7865	0.6752	0.7524	**0.8779**
	15%	0.8505	0.8259	0.9210	**0.9118**
Wine	25%	0.8830	0.8117	0.9156	**0.9586**
	35%	0.9123	0.7801	0.9078	**0.9635**
	15%	0.4155	0.3428	0.3078	**0.4216**
Heart	25%	0.4718	0.2476	0.3146	**0.5379**
	35%	0.5201	0.2168	0.3386	**0.5804**
	15%	0.3639	0.0477	**0.4483**	0.2622
ORL	25%	0.3586	0.0632	**0.4567**	0.4414
	35%	0.3034	0.0655	0.4569	**0.6060**
	15%	0.7820	0.7178	0.7689	**0.8271**
Breast	25%	0.8329	0.7192	0.7849	**0.8551**
	35%	0.8564	0.7310	0.8026	**0.8989**
	15%	0.3602	0.2014	1.0000	**1.0000**
Messidor	25%	0.7193	0.2495	1.0000	**1.0000**
	35%	1.0000	0.3017	1.0000	**1.0000**
	15%	0.0379	0.0290	0.9431	**0.9780**
Bank	25%	0.1771	0.0250	0.9633	**0.9817**
	35%	0.3887	0.0210	0.9670	**0.9881**
	15%	0.5275	0.1973	**0.6019**	0.5543
Isolet	25%	0.5302	0.1604	**0.6070**	0.5768
	35%	0.5476	0.1797	**0.6089**	0.5960
	15%	0.6177	0.6444	0.8999	**0.9477**
Digits	25%	0.6741	0.6755	0.9425	**0.9518**
	35%	0.6924	0.6567	0.9456	**0.9613**
	15%	0.5494	0.5242	0.5582	**0.6804**
Statlog	25%	0.5755	0.5104	0.5781	**0.6783**
	35%	0.5974	0.4851	0.5657	**0.7254**
	15%	0.3883	0.4026	**0.5672**	0.4814
MNIST	25%	0.4064	0.4069	0.5918	**0.5942**
	35%	0.3948	0.3836	0.5949	**0.6046**
ALL Datasets	Average	0.5765	0.4123	0.7147	**0.7590**

If a clustering result is close to the true partition, then its NMI and ARI values are high.

Experiments were divided into three parts according to the types of prior information: (1) using positive labels as prior information, (2) using pairwise constraints as prior information and (3) using both positive and negative labels as prior information. The first part compared the proposed algorithm with four semi-supervised clustering algorithms with positive labels, including LP [9], which is the classical label propagation algorithm, LNP [13], which is an improved label propagation algorithm with modified affinity matrix, CNMF [28], which is an NMF-based constrained clustering method and PLCC [27], which is a k-means-based partition level constrained clustering method. The second part compared the proposed algorithm with three semi-supervised clustering algorithms with pairwise constraints, including COP [29] and CVQE [45], both of which are methods for k-means-based pairwise constraints clustering, and PCLP [19], which is an LP algorithm with pairwise constraints. The third part compared the proposed algorithm with the positive and negative label propagation algorithm (PNLP) [18].

To compare these different algorithms, their related parameters were specified as follows.

- The number of clusters k was set equal to the true number of classes on each dataset.
- For each of the compared algorithms, a parameter similar to μ was set to reflect the importance of prior information. In the comparisons, each compared algorithm was tested by increasing the value of μ in the interval [0,1] in increments of 0.1, and its best clustering result with the highest *NMI* and *ARI* values on a particular dataset was selected.
- Certain algorithms in the comparison, such as LP, LNP, PLCC, PCLP, PNLP and the proposed algorithm, require a Gaussian kernel parameter σ to be specified to construct the similarity matrix. In the comparisons, each of these algorithms was tested with different σ values in the set $\{s/2, s/10, s/20, s/30, s/40, s/50\}$, where s is equal to the covariance of a dataset, and its best clustering result with the highest *NMI* and *ARI* values on the dataset was selected.
- The effectiveness of a semi-supervised clustering result depends on the amount of prior information. Therefore, in each part of the experiment, the amount of prior information data was set to 15, 25, and 35 percent of the number of objects included in a dataset. All the algorithms that were compared were implemented with three sizes of prior information by using six-, four-, and two-fold cross validation, respectively.
- The effectiveness of a semi-supervised clustering result depends on the quality of prior information. Therefore, in each part of the experiment, given the amount of prior information, all the compared algorithms were run with 10 different sets of prior information to compute the average clustering results for *NMI* and *ARI* on each dataset.

4.2. Experimental results

In the first part, the spectral clustering (SC) algorithm was tested without prior information, and involved the LP, LNP, CNMF, and proposed algorithms with positive labels. Tables 2 and 3 list the clustering performance of the different algorithms on all the datasets with positive labels as prior information. These comparisons indicate that the proposed algorithm is superior to the SC algorithm. This suggests that prior information is able to enhance the clustering effectiveness. Furthermore, the results show that

Table 6
NMI and ARI values for semi-supervised clustering methods with both positive and negative labels.

Datasets	percent	NMI		ARI	
		PNLP	NLPPC	PNLP	NLPPC
Iris	15%	0.8049	**0.8354**	0.7862	**0.8283**
	25%	0.8494	**0.8503**	**0.8603**	0.8527
	35%	0.8553	**0.8986**	0.8705	**0.9103**
Wine	15%	0.8354	**0.8877**	0.8661	**0.9085**
	25%	0.8274	**0.9178**	0.8532	**0.9373**
	35%	0.8526	**0.9467**	0.8822	**0.9619**
Heart	15%	0.3739	**0.4224**	0.4463	**0.4986**
	25%	0.4586	**0.5069**	0.5501	**0.5955**
	35%	0.4696	**0.6135**	0.5653	**0.7123**
ORL	15%	0.7414	**0.8048**	0.4001	**0.4835**
	25%	0.8263	**0.8302**	0.5456	**0.5572**
	35%	**0.8852**	0.8732	**0.7014**	0.6523
Breast	15%	0.6955	**0.7582**	0.7809	**0.8361**
	25%	0.7393	**0.8243**	0.8170	**0.8859**
	35%	0.7419	**0.8633**	0.8157	**0.9217**
Messidor	15%	0.9940	**1.0000**	0.9972	**1.0000**
	25%	1.0000	**1.0000**	1.0000	**1.0000**
	35%	1.0000	**1.0000**	1.0000	**1.0000**
Bank	15%	0.9787	**0.9870**	0.9910	**0.9948**
	25%	0.9825	**0.9910**	0.9927	**0.9965**
	35%	0.9837	**0.9917**	0.9933	**0.9968**
Isolet	15%	**0.8321**	0.8189	**0.6867**	0.6172
	25%	**0.8783**	0.8516	**0.7784**	0.7226
	35%	**0.9052**	0.8891	**0.8350**	0.7844
Digits	15%	0.9353	**0.9600**	0.9450	**0.9647**
	25%	0.9601	**0.9661**	0.9670	**0.9709**
	35%	0.9712	**0.9754**	0.9770	**0.9797**
Statlog	15%	**0.7523**	0.7426	**0.7844**	0.7554
	25%	0.7694	**0.7714**	**0.8008**	0.7848
	35%	0.7747	**0.8007**	0.8076	**0.8109**
MNIST	15%	0.6094	**0.8293**	0.6149	**0.7954**
	25%	0.6869	**0.8595**	0.7004	**0.8364**
	35%	0.7406	**0.8948**	0.7580	**0.8937**
ALL Datasets	Average	0.8094	**0.8534**	0.7991	**0.8317**

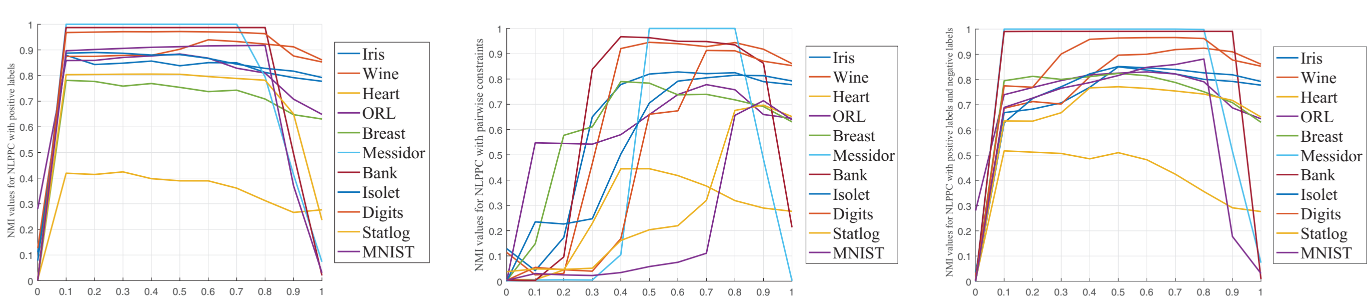
the proposed algorithm is clearly more accurate than LNP, CNMF, and PLCC on these tested datasets. On certain datasets, the proposed algorithm with specific amounts of prior information performed slightly less accurately than LP. However, the performance of the proposed algorithm is superior on most datasets compared to LP. Therefore, the experimental results indicate that the proposed method can enhance the clustering effectiveness by solving the label misalignment problem in the process of label propagation compared to other algorithms.

The second part of the experiments involved testing the COP, CVQE, PCLP, and proposed algorithm with pairwise constraints. The experiment was carried out by setting the number of must-link constraints equal to that of the cannot-link constraints on each dataset. Tables 4 and 5 present the clustering performance of the different algorithms on all the datasets with pairwise constraints. These results show that the proposed algorithm is clearly superior to COP and CVQE on these tested datasets. On 3 of the 11 datasets, the proposed algorithm performed slightly less accurate than PCLP with specific amounts of prior information. However, compared to PCLP, the proposed algorithm is more accurate on most datasets.

The third part of the experiments entailed testing PNLP and the proposed algorithm with both positive and negative labels. The experiment was conducted by setting the number of positive labels equal to the number of negative labels on each dataset. Table 6 lists the clustering performance of the compared algorithms. The performance of the proposed algorithm is obviously superior to that of PNLP on seven of the datasets. On the other datasets, the performance of the proposed algorithm closely approximates that of PNLP.

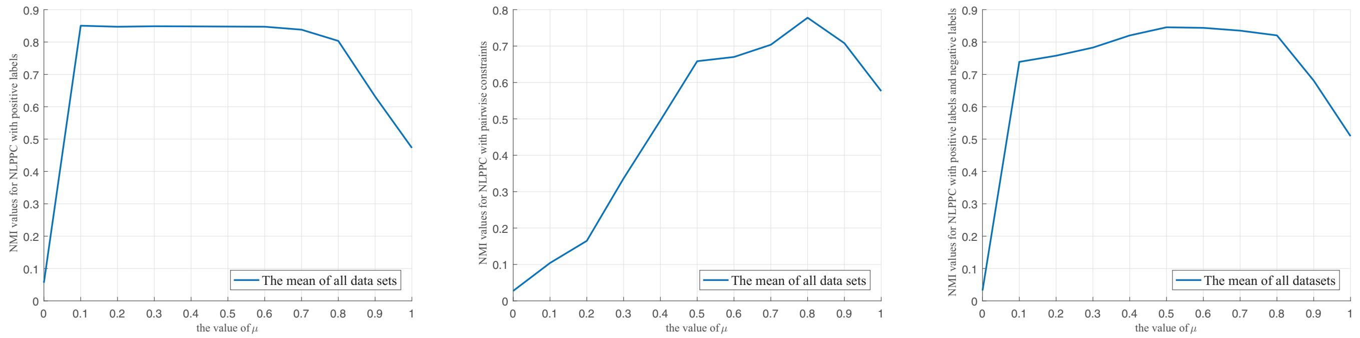
The tables also present the average clustering results of each algorithm on all datasets. The results of the comparison show that our algorithm outperforms the other algorithms with respect to the average clustering results with different types of prior information on all datasets. The experimental analysis indicates that the clustering accuracy of the proposed algorithm is higher than that of the other algorithms on most of the tested datasets. This result is mainly attributed to the ability of the proposed algorithm to solve the misalignment problem between pre-given class labels and cluster labels and obtain a good optimization solution by the eigenvalue decomposition method. It is equally noteworthy that the effectiveness of the proposed algorithm with certain amounts of prior information is slightly less than that of LP, PCLP, and NLPCC on particular datasets. However, compared to LP, PCLP, and PNLP, the advantage of the proposed algorithm is that it can effectively accommodate different types of prior information.

Next, the effect of the parameter μ on the performance of the proposed algorithm on each of datasets was analyzed, as shown in Figs. 3, and 5. The analysis considered the following three scenarios, i.e., the proposed algorithm with positive labels, with pairwise constraints, and with both positive and negative labels. The overall number of constraints was assigned a constant value of 25%N, where N is the number of objects in a dataset. The proposed algorithm was tested by increasing the value of μ in steps of 0.1 in the interval [0,1]. The results in these figures show that the parameter μ has a different effect on each of these datasets. This indicates that it is difficult to select an appropriate μ for the proposed algorithm on each dataset. The effect was further analyzed by plotting the mean values of the evaluation indices for the proposed algorithm on all the tested datasets for each μ in Figs. 4 and 6. The plotted curves show that, for small increases in μ, the average performance of the proposed algorithm is relatively stable.



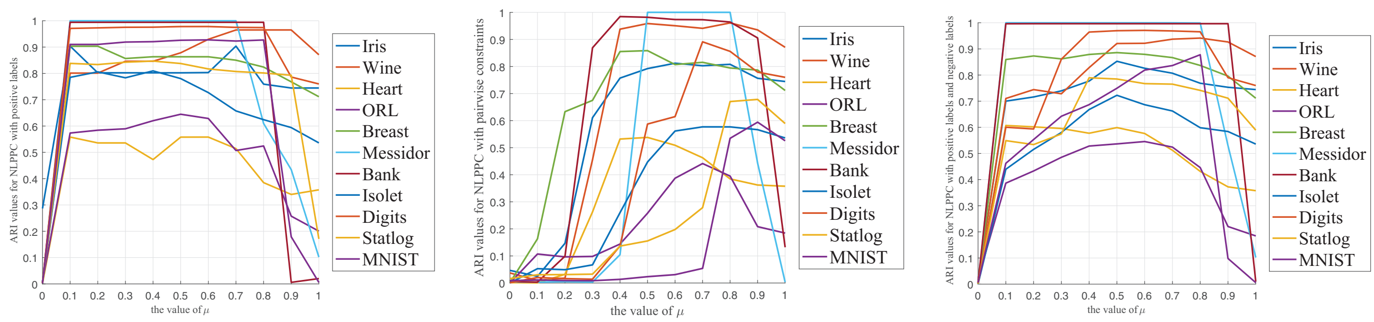
(a) positive labels (b) pairwise constraints (c) positive labels and negative labels

Fig. 3. Effect of parameter μ on the *NMI* values of NLPPC.



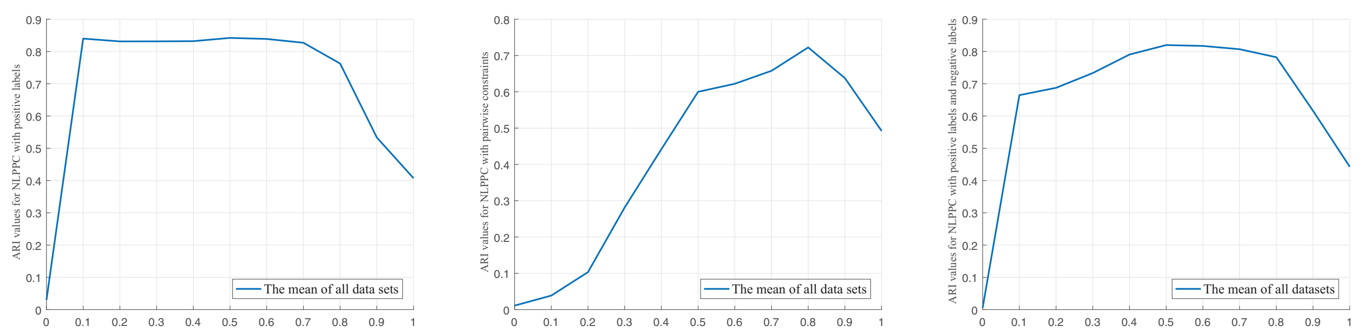
(a) positive labels (b) pairwise constraints (c) positive labels and negative labels

Fig. 4. Effect of parameter μ on the average *NMI* values of NLPPC.



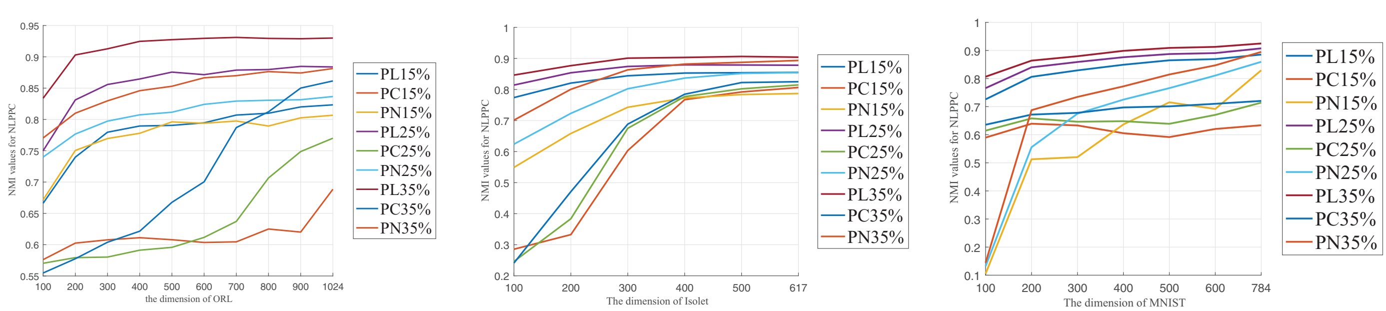
(a) positive labels (b) pairwise constraints (c) positive labels and negative labels

Fig. 5. Effect of parameter μ on the *ARI* values of NLPPC.



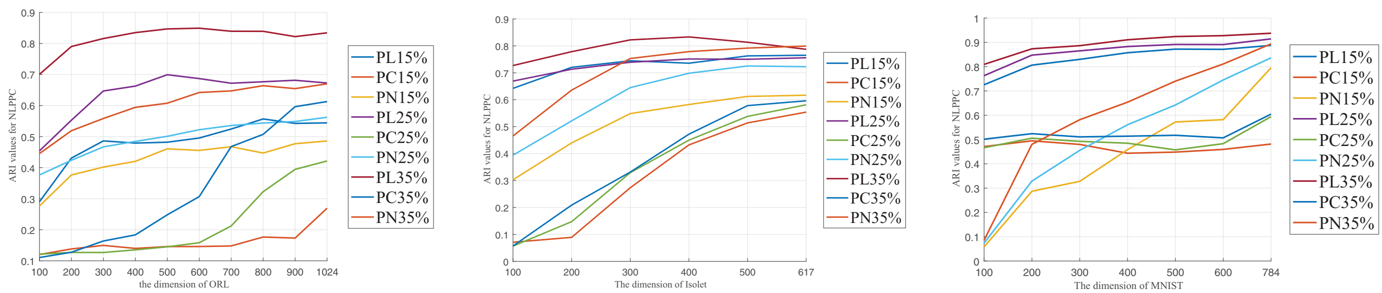
(a) positive labels (b) pairwise constraints (c) positive labels and negative labels

Fig. 6. Effect of parameter μ on the average *ARI* values of NLPPC.



(a) ORL dataset (b) Isolet data set (c) MNIST data set

Fig. 7. Effect of the number of dimensions on the *NMI* value of NLPPC.



(a) ORL data set (b) Isolet data set (c) MNIST data set

Fig. 8. Effect of the number of dimensions on the *ARI* value of NLPPC.

Finally, the effect of the number of dimensions on the performance of the proposed algorithm was analyzed. Three datasets containing high-dimensional images, including *ORL, Isolet*, and *MNIST*, were selected to assess the performance of the proposed algorithm with different dimensions. First, each dataset was randomly disordered in the dimension direction. Then, the *NMI* values of the proposed algorithm were compared with the different types of prior information on each dataset with different dimensions, as shown in Figs. 7 and 8. The performance of the proposed algorithm tended to be stable as the number of dimensions increased for each dataset. The experimental results suggest that the proposed algorithm can effectively accommodate high-dimensional datasets.

5. Conclusions

In this paper, we propose a new label propagation algorithm with pairwise constraints, named NLPPC. In the proposed algorithm, the divergence between the pairwise relations of the prior information and clustering result is used as a penalty factor, instead of the difference between partition matrices. An eigenvalue decomposition method is used to optimize its solution. The proposed algorithm not only solves the misalignment problem of the LP algorithm, it is also able to process pairwise constraints. Finally, extensive experiments were conducted to demonstrate the effectiveness of the proposed algorithm compared to other semi-supervised clustering algorithms. The experimental results indicated that the effectiveness of the proposed algorithm with certain amounts of prior information is slightly less than that of LP, PCLP, and PNLP on particular datasets. However, the clustering performance of the proposed algorithm is superior to that of the other algorithms on most of the tested datasets.

This study mainly focused on label propagation with pairwise constraints. In future, we aim to investigate the effect of noisy prior information on semi-supervised clustering. Furthermore, we plan to develop a semi-supervised clustering framework with high robustness which will make use of the ensemble learning technique to reduce the effect of noisy prior information.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Acknowledgement

The authors are very grateful to the editors and reviewers for their valuable comments and suggestions. This work is supported by the National Natural Science Foundation of China (Nos. 61773247, 61876103, 61976128, 61902227), the Technology Research Development Projects of Shanxi (No. 201901D211192) and the 1331 Engineering Project of Shanxi Province, China.

References

[1] X. Wu, V. Kumar, J.R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G.J. Mclachlan, A. Ng, B. Liu, Yu., Top 10 algorithms in data mining, Knowl. Inf. Syst. 14 (1) (2008) 1–37.

[2] A.K. Jain, Data clustering: 50 years beyond k-means, Pattern Recognit. Lett. 31 (8) (2010) 651–666.

[3] A.-J. Gallego, J. Calvo-Zaragoza, J.J. Valero-Mas, J.R. Rico-Juan, Clustering-based k-nearest neighbor classification for large-scale data with neural codes representation, Pattern Recognit. 74 (2018) 531–543.

[4] S. Paterlini, T. Krink, Differential evolution and particle swarm optimisation in partitional clustering, Comput. Stat. Data Anal. 50 (5) (2006) 1220–1247.

[5] Y. Lu, Y. Wan, Pha: a fast potential-based hierarchical agglomerative clustering method, Pattern Recognit. 46 (5) (2013) 1227–1239.

[6] K.M. Kumar, A.R.M. Reddy, A fast dbscan clustering algorithm by accelerating neighbor searching using groups method, Pattern Recognit. 58 (2016) 39–48.

[7] D. Dinler, M.K. Tural, Robust semi-supervised clustering with polyhedral and circular uncertainty, Neurocomputing 265 (2017) 4–27.

[8] X. Zhou, M. Belkin, Semi-supervised learning, Acad. Press Libr. Signal Process. 1 (2014) 1239–1269.

[9] D. Zhou, O. Bousquet, T.N. Lal, J. Weston, B.S. Olkopf, Learning with local and global consistency, in: International Conference on Neural Information Processing Systems, 2003.

[10] H. Wang, T. Li, T. Li, Y. Yang, Constraint neighborhood projections for semi-supervised clustering, IEEE Trans. Syst. Man Cybern. 44 (5) (2014) 636–643.

[11] X. Zhu, A.B. Goldberg, Introduction to semi-supervised learning, Synth. Lect. Artif. Intell.Mach. Learn. 3 (1) (2009) 1–130.

[12] R. Johnson, T. Zhang, On the effectiveness of Laplacian normalization for graph semi-supervised learning, J. Mach. Learn. Res. 8 (2007) 1489–1517.

[13] F. Wang, C. Zhang, Label propagation through linear neighborhoods, IEEE Trans. Knowl. Data Eng. 20 (1) (2007) 55–67.

[14] J. Tang, R. Hong, S. Yan, T.S. Chua, G.J. Qi, R. Jain, Image annotation by kNN-sparse graph-based label propagation over noisily tagged web images, ACM Trans. Intell. Syst. Technol. 2 (2) (2011) 1–15.

[15] M. Belkin, I. Matveeva, P. Niyogi, Regularization and semi-supervised learning on large graphs, in: Conference on Learning Theory, 2004, pp. 624–638.

[16] X. Zhu, Z. Ghahramani, J.D. Lafferty, Semi-supervised learning using gaussian fields and harmonic functions, in: International Conference on Machine Learning, 2003, pp. 912–919.

[17] X. Zhu, J. Lafferty, Z. Ghahramani, Semi-Supervised Learning: From Gaussian Fields to Gaussian Processes, Technical Report, School of CS, CMU, 2003.

[18] O. Zoidi, A. Tefas, N. Nikolaidis, I. Pitas, Positive and negative label propagations, IEEE Trans. Circuits Syst. Video Technol. 28 (2) (2018) 342–355.

[19] Z. Lu, Y. Peng, Exhaustive and efficient constraint propagation: agraph-based learning approach and its applications, Int. J. Comput. Vis. 103 (3) (2013) 306–325.

[20] M.S. Baghshah, S.B. Shouraki, Semi-supervised metric learning using pairwise constraints, in: International Joint Conference on Artificial Intelligence, 2009, pp. 1217–1222.

[21] Z.-H. Zhou, Ensemble methods: Foundations and Algorithms, Chapman and Hall/CRC, 2012.

[22] X. Yin, S. Chen, E. Hu, D. Zhang, Semi-supervised clustering with metric learning: an adaptive kernel method, Pattern Recognit. 43 (4) (2010) 1320–1333.

[23] J.V. Davis, B. Kulis, P. Jain, S. Sra, I.S. Dhillon, Information-theoretic metric learning, in: Proceedings of the 24th International Conference on Machine Learning, 2007, pp. 209–216.
[24] L. Wu, S.C. Hoi, R. Jin, J. Zhu, N. Yu, Learning Bregman distance functions for semi-supervised clustering, IEEE Trans. Knowl. Data Eng. 24 (3) (2012) 478–491.
[25] S. Basu, A. Banerjee, R.J. Mooney, Semi-supervised clustering by seeding, in: International Conference on Machine Learning, 2002, pp. 27–34.
[26] Z. Yu, D. Wang, J. You, H. Wong, S. Wu, J. Zhang, G. Han, Progressive subspace ensemble learning, Pattern Recognit. 60 (2016) 692–705.
[27] H. Liu, Z. Tao, Y. Fu, Partition level constrained clustering, IEEE Trans. Pattern Anal. Mach. Intell. 40 (10) (2018) 2469–2483.
[28] H. Liu, Z. Wu, Non-negative matrix factorization with constraints, in: National Conference on Artificial Intelligence, 2010, pp. 506–511.
[29] K. Wagstaff, C. Cardie, S. Rogers, S. Schrodl, Constrained k-means clustering with background knowledge, in: Eighteenth International Conference on Machine Learning, 2001.
[30] Y. Yang, W. Tan, T. Li, D. Ruan, Consensus clustering based on constrained self--organizing map and improved Cop-Kmeans ensemble in intelligent decision support systems, Knowl. Based Syst. 32 (2012) 101–115.
[31] S. Wei, Z. Li, C. Zhang, A semi-supervised clustering ensemble approach integrated constraint-based and metric-based, in: International Conference on Internet Multimedia Computing and Service, 2015, p. 26.
[32] T. Li, C.H.Q. Ding, M.I. Jordan, Solving consensus and semi-supervised clustering problems using nonnegative matrix factorization, in: International Conference on Data Mining, 2007, pp. 577–582.
[33] S.D. Kamvar, D. Klein, C.D. Manning, Spectral learning, in: International Joint Conference on Artificial Intelligence, 2003, pp. 561–566.
[34] Q. Xu, M. Desjardins, K. Wagstaff, Constrained spectral clustering under a local proximity structure assumption, in: In FLAIRS Conference, 2005.
[35] X. Ji, W. Xu, Document clustering with prior knowledge, in: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2006, pp. 405–412.
[36] F. Wang, C. Ding, T. Li, Integrated KL (K-means–Laplacian) clustering: a new clustering approach by combining attribute data and pairwise relations, in: Proceedings of the 2009 SIAM International Conference on Data Mining, 2009, pp. 38–48.
[37] B. Kulis, S. Basu, I.S. Dhillon, R.J. Mooney, Semi-supervised graph clustering: a kernel approach, Mach. Learn. 74 (1) (2009) 1–22.
[38] J. Yu, D. Tao, J. Li, J. Cheng, Semantic preserving distance metric learning and applications, Inf. Sci. 281 (2014) 674–686.
[39] N.M. Portela, G.D.C. Cavalcanti, T.I. Ren, Semi-supervised clustering for mr brain image segmentation, Expert Syst. Appl. 41 (4) (2014) 1492–1497.
[40] R. Huang, W. Lam, An active learning framework for semi-supervised document clustering with language modeling, Data Knowl. Eng. 68 (1) (2009) 49–67.
[41] L. Yang, X. Cao, D. Jin, X.Y. Wang, D. Meng, A unified semi-supervised community detection framework using latent space graph regularization, IEEE Trans. Syst. Man Cybern. 45 (11) (2015) 2585–2598.
[42] D. You, O.C. Hamsici, A.M. Martinez, Kernel optimization in discriminant analysis, IEEE Trans. Pattern Anal. Mach. Intell. 33 (3) (2011) 631–638.
[43] B. Dom, An information-theoretic external cluster-validity measure, Uncertain. Artif. Intell. (2002) 137–145.
[44] L. Hubert, P. Arabie, Comparing partitions, J. Classif. 2 (1) (1985) 193–218.
[45] I. Davidson, S.S. Ravi, Clustering with constraints: feasibility issues and the k-means algorithm, in: SIAM International Conference on Data Mining, 2005, pp. 138–149.

Liang Bai received his Ph.D degree in computer science from Shanxi University in 2012. He is currently a Professor with the School of Computer and Information Technology, Shanxi University. His research interest is in the areas of cluster analysis. He has published several journal papers in his research fields, including IEEE TPAMI, IEEE TKDE, DMKD, IEEE TFS, PR and INS and so on.

Junbin Wang is a M.S. candidate from the School of Computer and Information Technology at Shanxi University, China. His research interest is semi-supervised learning.

Jiye Liang received the M.S. and Ph.D degrees from Xi'an Jiaotong University, Xi'an, China, in 1990 and 2001, respectively. He is currently a Professor with the School of Computer and Information Technology and the Key Laboratory of Computational Intelligence and Chinese Information Processing of the Ministry of Education, Shanxi University, Taiyuan, China. His current research interests include computational intelligence, rough set theory, granular computing and so on. He has published more than 130 journal papers in his research fields, including AI, IEEE TPAMI, IEEE TKDE, IEEE TFS, DMKD, IEEE TSMC and so on.

Hangyuan Du received the B.S. degree from the Heilongjiang Institute of Technology, in 2008, and the Ph.D. degree in control science and engineering from Harbin Engineering University, in 2012. He is currently a member of the School of Computer and Information Technology, Shanxi University, as a lecturer. His research interests lie in the areas of cluster analysis.