Symmetric Graph Regularized Constraint Propagation

Zhenyong Fu[1,3] and **Zhiwu Lu**[2*] and **Horace H.S. Ip**[1] and **Yuxin Peng**[2] and **Hongtao Lu**[3]
[1]Department of Computer Science, City University of Hong Kong, Hong Kong
[2]Institute of Computer Science and Technology, Peking University, Beijing 100871, China
[3]Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China
zhenyonfu2@student.cityu.edu.hk, luzhiwu@icst.pku.edu.cn

Abstract

This paper presents a novel symmetric graph regularization framework for pairwise constraint propagation. We first decompose the challenging problem of pairwise constraint propagation into a series of two-class label propagation subproblems and then deal with these subproblems by quadratic optimization with symmetric graph regularization. More importantly, we clearly show that pairwise constraint propagation is actually equivalent to solving a Lyapunov matrix equation, which is widely used in Control Theory as a standard continuous-time equation. Different from most previous constraint propagation methods that suffer from severe limitations, our method can directly be applied to multi-class problem and also can effectively exploit both must-link and cannot-link constraints. The propagated constraints are further used to adjust the similarity between data points so that they can be incorporated into subsequent clustering. The proposed method has been tested in clustering tasks on six real-life data sets and then shown to achieve significant improvements with respect to the state of the arts.

Introduction

Pairwise constraints provide prior knowledge on whether two data points belong to the same class or not, known as must-link constraints and cannot-link constraints, respectively. Generally, it is hard to infer instance labels only from pairwise constraints, especially for multi-class data. That is, pairwise constraints are weaker and thus more general than the explicit labels of data. In practice, we can derive pairwise constraints from domain knowledge. Similar to the case that very few data labels are provided for semi-supervised learning, we also suffer from the scarcity of pairwise constraints. It is a challenging task to propagate such scarce pairwise constraints across all the data points.

Pairwise constraints have been widely used in the context of clustering with side information (Xing et al. 2003; Kamvar, Klein, and Manning 2003; Basu, Bilenko, and Mooney 2004; Kulis et al. 2005), where it has been shown that the presence of appropriate pairwise constraints can often improve the performance. For example, in (Kamvar, Klein, and Manning 2003), the similarities between constrained data are trivially adjusted to 1 and 0 for must-link and cannot-link constraints, respectively. This method only adjusts the similarities between constrained data, and does not propagate the pairwise constraints to other data. In contrast, in (Lu and Carreira-Perpinan 2008), the pairwise constraints are propagated to unconstrained data using Gaussian process. However, this constraint propagation method makes certain assumptions to deal with cannot-link constraints specially for two-class problems, although the heuristic approach for multi-class problems is also discussed. Moreover, the pairwise constraint propagation is formulated as a semi-definite programming (SDP) problem in (Li, Liu, and Tang 2008). Although this optimization-based method is not limited to two-class problems, it incurs extremely large computational cost for solving SDP and the authors only report the experimental results on small-scale data sets.

To overcome the above problems, we make attempt to decompose the pairwise constraint propagation problem into a series of two-class semi-supervised learning subproblems. Although we have proposed an exhaustive and efficient constraint propagation method based on the traditional semi-supervised learning in (Lu and Ip 2010), we take a totally different regularization approach into account in this paper. More concretely, we exploit the special symmetric structure of the pairwise constraints and develop a pairwise constraint propagation approach based on symmetric graph regularization. In the following, we will call it as symmetric graph regularized constraint propagation (SRCP).

Interestingly, under our symmetric regularization framework, we show, for the first time, that pairwise constraint propagation is actually equivalent to solving a Lyapunov matrix equation. As a standard continuous-time equation, the Lyapunov equation has been widely used to solve different problems in Control Theory, System Identification and System Stability Analysis (Gajic and Qureshi 1995). It should be noted that the Lyapunov equation has a closed form solution and can be efficiently solved by the Matlab software using a numerical method. More significantly, as an alternative approach, we further formulate pairwise constraint propagation as symmetric information spreading and show that the time invariant solution of this propagation formula also corresponds to that of the special Lyapunov equa-

*Corresponding author.

tion. The present work therefore gives rise to two interesting theoretical insights that link the problem of pairwise constraint propagation to the Lyapunov equation.

The propagated constraints obtained by solving the Lyapunov equation are further used to adjust the similarity between data points so that they can be incorporated into subsequent clustering. To evaluate the effectiveness of the proposed method, we select six real-life data sets for clustering tasks. The experimental results have shown that the proposed method can achieve significant improvements with respect to the state of the arts. Finally, the main contributions of this paper can be summarized as follows:

- This is the first attempt to deal with the challenging problem of pairwise constraint propagation based on a symmetric graph regularization framework.
- Under this framework, we show, for the first time, that pairwise constraint propagation is actually equivalent to solving a Lyapunov equation.
- We also formulate pairwise constraint propagation as symmetric information spreading, which has the same solution as the Lyapunov equation.

The remainder of this paper is organized as follows. Section 2 proposes a symmetric graph regularized constraint propagation algorithm. In Section 3, our method is evaluated on six real-life data sets. Finally, Section 4 gives the conclusions drawn from the experimental results.

Symmetric Graph Regularized Constraint Propagation

In this section, we first give a short review of the graph regularized semi-supervised learning, especially in the situation of the two-class problem, followed by the details of our proposed algorithm, including the propagation of the pairwise constraints and the application of the propagated constraints to data clustering tasks.

Graph Regularized Semi-supervised Learning

Given a point set $\mathcal{X} = \{x_1, \ldots, x_p, x_{p+1}, \ldots, x_n\} \subset \mathbb{R}^m$ and a label set $\mathcal{L} = \{1, \ldots, k\}$, the first p points $x_i (i \leq p)$ are labeled as $y_i \in \mathcal{L}$ and the remaining points $x_u (p+1 \leq u \leq n)$ are unlabeled. Semi-supervised learning (or label propagation) focuses on how to learn from both labeled and unlabeled data (Zhu and Goldberg 2009; Zhou et al. 2004).

Let $G = (V, W)$ be an undirected, weighted graph defined on the data set, i.e., $V = \mathcal{X}$. The similarity matrix W is defined as $W = [w_{ij}]_{n \times n}$, where w_{ij} is the similarity measurement between x_i and x_j. For modeling the local neighborhood relationships between the data points, we construct a k-nearest neighbor graph, in which $w_{ij} = 0$ if x_j is not among the k-nearest neighbors (k-NN) of x_i. We set $w_{ii} = 0$ for $1 \leq i \leq n$ to avoid self-reinforcement, and set $W = (W + W^T)/2$ to ensure that W is symmetric. The graph Laplacian L of G is defined as $L = D - W$, where $D = [d_{ii}]_{n \times n}$ is a diagonal matrix with $d_{ii} = \sum_j w_{ij}$. The normalized graph Laplacian $\bar{L}$ of G is defined as

$$\bar{L} = D^{-1/2} L D^{-1/2} = I - D^{-1/2} W D^{-1/2},$$

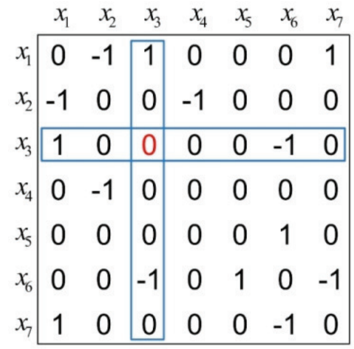


Figure 1: Illustration of the constraint indicator matrix defined by equation (2). When we focus on a single data point, such as x_3 here, the pairwise constraint propagation can be viewed as a two-class semi-supervised learning problem in the row and column simultaneously.

where I is the identity matrix. It should be noted that $\bar{L}$ is symmetric and positive semidefinite, with its eigenvalues being in the interval [0, 2] (Chung 1997).

The smoothness of a function $\boldsymbol{f} : \mathcal{X} \rightarrow \mathcal{R}$ on the graph can be measured by

$$\Omega(\boldsymbol{f}) = \frac{1}{2} \sum_{i,j} w_{ij} (\frac{f(x_i)}{\sqrt{d_{ii}}} - \frac{f(x_j)}{\sqrt{d_{jj}}})^2 = \boldsymbol{f}^T \bar{L} \boldsymbol{f},$$

where $\boldsymbol{f} = (f(x_1), \ldots, f(x_n))^T$. The smaller is $\Omega(\boldsymbol{f})$, the smoother is $\boldsymbol{f}$. This measurement penalizes large changes between data points that are strongly connected.

If we only consider the two-class problem, we can use the vector $\boldsymbol{y}$ and $\boldsymbol{f}$ to represent the initial labels and the classification results, respectively. Each element y_i of y is defined as $y_i = +1, -1$ or 0, if x_i is labeled as the positive class, the negative class or unlabeled. And each unlabeled data point x_i will be labeled as the positive or negative class according to the sign of $f(x_i)$. In (Zhou et al. 2004), a graph regularization framework is proposed to predict the labels of unlabeled data points. This method simultaneously considers the loss function of labels and the smoothness on the graph. According to (Zhou et al. 2004), the two-class label propagation problem can be formulated as

$$\min_{\boldsymbol{f}} \frac{1}{2} \mu \|\boldsymbol{f} - \boldsymbol{y}\|_2^2 + \frac{1}{2} \boldsymbol{f}^T \bar{L} \boldsymbol{f}, \quad (1)$$

where $\mu > 0$ is a regularization parameter. The classification is performed according to:

$$l(x_i) = \begin{cases} +1, & f(x_i) \geq 0; \\ -1, & f(x_i) < 0, \end{cases}$$

where $l(x_i)$ is the predicted label of data x_i. The value of $|f(x_i)|$ can be viewed as the confidence score of labeling x_i as the positive or negative class.

Symmetric Graph Regularization Framework

We now consider the pairwise constraint propagation problem. The problem definition is very similar to that of the semi-supervised learning. Given a data set of n objects $\mathcal{X} =$

$\{x_1, \ldots, x_n\}$ and two sets of pairwise constraints, denoted respectively by $\mathcal{M} = \{(x_i, x_j)\}$ where x_i and x_j should be in the same class and $\mathcal{C} = \{(x_i, x_j)\}$ where x_i and x_j should be in different classes, our goal is to propagate the sparse pairwise constraints across the entire data set and then classify $\mathcal{X}$ into k classes.

We first represent these two types of pairwise constraints with a single matrix $Y = \{Y_{ij}\}_{n\times n}$:

$$Y_{ij} = \begin{cases} +1, & (x_i, x_j) \in \mathcal{M}; \\ -1, & (x_i, x_j) \in \mathcal{C}; \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

It is obvious that Y is a symmetric matrix. In addition, we define $F = \{F_{ij}\}_{n\times n}$ as the matrix that stores the propagated pairwise constraints.

As shown in Figure 1, when we focus on a single data point x_i in $\mathcal{X}$, we use $Y_{i\cdot}$ and $Y_{\cdot i}$ to respectively denote the i-th row and i-th column with respect to x_i in Y. Similarly, we use $F_{i\cdot}$ and $F_{\cdot i}$ to respectively denote the i-th row and i-th column with respect to x_i in F. It can be observed that $Y_{i\cdot}$ and $Y_{\cdot i}$ is the initial pairwise constraints between x_i and other data points in $\mathcal{X}$ before constraint propagation.

Propagating the constraint relationships related to x_i can be viewed as a two-class semi-supervised learning problem, where the "positive class" is the must-link relationship and the "negative class" is the cannot-link relationship. If we only use $Y_{\cdot i}$, according to equation (1), the constraint propagation with respect to x_i can be formulated as

$$\min_{F_{\cdot i}} \frac{1}{2}\mu\|F_{\cdot i} - Y_{\cdot i}\|_2^2 + \frac{1}{2}F_{\cdot i}^T \bar{L} F_{\cdot i}$$

Similarly, if we only use $Y_{i\cdot}$, the constraint propagation problem is the solution:

$$\min_{F_{i\cdot}} \frac{1}{2}\mu\|F_{i\cdot} - Y_{i\cdot}\|_2^2 + \frac{1}{2}F_{i\cdot} \bar{L} F_{i\cdot}^T$$

By considering these two separated propagation processes simultaneously and combining them together, the constraint propagation problem with respect to x_i is equivalent to:

$$\min_{F_{\cdot i}, F_{i\cdot}} \frac{1}{2}\mu\|F_{\cdot i} - Y_{\cdot i}\|_2^2 + \frac{1}{2}\mu\|F_{i\cdot} - Y_{i\cdot}\|_2^2 + \frac{1}{2}F_{\cdot i}^T \bar{L} F_{\cdot i} + \frac{1}{2}F_{i\cdot} \bar{L} F_{i\cdot}^T$$

Again, if we merge all of these subproblems into a single optimization problem, we can get:

$$\min_F \mu\|F - Y\|_F^2 + \frac{1}{2}\mathrm{tr}(F^T \bar{L} F + F \bar{L} F^T) \quad (3)$$

Here, $\mathrm{tr}(Z)$ stands for the trace of a matrix Z.

Let $\mathcal{Q}(F)$ denote the objective function in equation (3). Differentiating $\mathcal{Q}(F)$ with respect to F and setting it to zero, we have the following equation for constraint propagation,

$$\frac{\partial \mathcal{Q}}{\partial F} = 2\mu(F - Y) + \bar{L}F + F\bar{L} = 0, \quad (4)$$

which can be transformed into a symmetric form

$$(\mu I + \bar{L})F + F(\mu I + \bar{L}) = 2\mu Y. \quad (5)$$

It should be noted that the above equation is a standard continuous-time Lyapunov matrix equation (Barnett and Storey 1967) and is of interest in a number of areas of Control Theory such as optimal control and stability analysis (Gajic and Qureshi 1995). According to Proposition 1, the Lyapunov matrix equation (5) has a unique solution. Moreover, this solution is symmetric if Y is.

Proposition 1. *The Lyapunov matrix equation* (5) *has a unique solution.*

Proof. Since the graph Laplacian $\bar{L}$ is positive semidefinite and $\mu > 0$, we know that $\mu I + \bar{L}$ is positive definite and thus all of its eigenvalues are positive. So, for any pair of eigenvalues of $\mu I + \bar{L}$, α_i and α_j, we have $\alpha_i + \alpha_j \neq 0$. According to (Lancaster 1970), the Lyapunov matrix equation (5) has a unique solution. □

Furthermore, the solution of the Lyapunov matrix equation (5) can be explicitly derived as follows. By vectoring on two sides of equation (5), we have

$$\mathrm{vec}[(\mu I + \bar{L})F] + \mathrm{vec}[F(\mu I + \bar{L})] = \mathrm{vec}(2\mu Y),$$

and it is equivalent to

$$[I \otimes (\mu I + \bar{L}) + (\mu I + \bar{L}) \otimes I]\mathrm{vec}(F) = 2\mu \mathrm{vec}(Y),$$

where the symbol $\otimes$ denotes the Kronecker product. Similar to the proof of Proposition 1, we can prove that $I \otimes (\mu I + \bar{L}) + (\mu I + \bar{L}) \otimes I$ is nonsingular. Hence, the closed form solution of equation (5) is

$$F = \mathrm{unvec}(2\mu(I \otimes (\mu I + \bar{L}) + (\mu I + \bar{L}) \otimes I)^{-1}\mathrm{vec}(Y)). \quad (6)$$

However, this method needs to compute an inverse matrix with size $n^2 \times n^2$, and thus is not efficient for solving large-scale problems. Fortunately, many numerical methods have been developed to solve equation (5) efficiently.

Symmetric Information Spreading Perspective

We first give a briefly review of an alternative insight in semi-supervised learning. In (Zhou et al. 2004), the semi-supervised learning problem is formulated as a form of information spreading on the graph. Let $S = D^{-1/2}WD^{-1/2}$. The process of information spreading is defined by the iteration $H(t+1) = \alpha SH(t) + (1-\alpha)Z$, where H stores the predicted labels, Z collects the initial labels, t is the iteration step, and α is a weight parameter.

As for pairwise constraint propagation, if we take into account the symmetric structure of the constraints, we can similarly define the process of constraint propagation as

$$F(t+1) = \frac{1}{2}\alpha SF(t) + \frac{1}{2}\alpha F(t)S + (1-\alpha)Y. \quad (7)$$

That is, during each iteration, each data point receives the information from its neighbors along the row and column directions simultaneously as depicted in Figure 1 (see the first two terms of the above equation), and also retains its initial information (see the third term of the above equation). The final results of constraint propagation correspond to the time invariant solution of equation (7), which means $F(t+$

$1) = F(t)$. Let F denote the converged constraint spreading result, and we can get F by solving

$$F = \frac{1}{2}\alpha SF + \frac{1}{2}\alpha FS + (1-\alpha)Y. \quad (8)$$

Let $\mu = (1-\alpha)/\alpha$. Since $\bar{L} = I - S$, we again get the Lyapunov matrix equation (5).

Similarity Adjustment with Constraint Propagation

Once we have obtained the constraint propagation result F, we can consider F_{ij} as the confidence score of the pairwise constraint between x_i and x_j. To incorporate F into the subsequent clustering process, we adjust the similarities between the data points according to the following similarity refinement formula

$$w^*_{ij} = \begin{cases} 1-(1-F_{ij})(1-w_{ij}), & F_{ij} \geq 0; \\ (1+F_{ij})w_{ij}, & F_{ij} < 0. \end{cases} \quad (9)$$

The above refinement can increase the similarity between x_i and x_j when $F_{ij} > 0$ and decrease it when $F_{ij} < 0$. More details can be found in (Lu and Ip 2010).

Symmetric Graph Regularized Constraint Propagation: The Algorithm

Let $W^{=}[w^*_{ij}]_{n\times n}$ be the adjusted similarity matrix according to equation (9). We adopt the spectral clustering algorithm (von Luxburg 2007) with W^* to form k classes. Based on the previous analysis, we develop a constrained clustering algorithm listed in Algorithm 1, in which our constraint propagation is used. In the following, we call it *Symmetric Graph Regularized Constraint Propagation* (SRCP).

Algorithm 1 Symmetric Graph Regularized Constraint Propagation

Input: A data set of n objects $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$, the set of must-link constraints $\mathcal{M} = \{(x_i, x_j)\}$, the set of cannot-link constraints $\mathcal{C} = \{(x_i, x_j)\}$, and the number of classes k.
Output: Cluster labels for the objects in $\mathcal{X}$.

1. Form the symmetric k-NN similarity matrix $W = [w_{ij}]_{n\times n}$.
2. Form the normalized graph Laplacian $\bar{L} = I - D^{-1/2}WD^{-1/2}$, where $D = \text{diag}(d_{ii})$ is the diagonal matrix with $d_{ii} = \sum_{j=1}^{n} w_{ij}$.
3. Solve the Lyapunov matrix equation (5) to obtain the constraint propagation result F.
4. Adjust the similarity matrix using F according to equation (9).
5. Form k classes by performing spectral clustering with the adjusted similarity matrix W^*.

Table 1: Description of the four UCI data sets

	Zoo	WDBC	Ionosphere	Wine
# objects	101	569	351	178
# dimensions	16	30	34	13
# classes	7	2	2	3

Experiments

In this section, we evaluate the performances of the proposed algorithm on a number of real-life data sets. For comparison, the results of three notable and most related algorithms, Lu and Carreira-Perpinan's affinity propagation (AP) (Lu and Carreira-Perpinan 2008), Kamvar et al.'s spectral learning (SL) (Kamvar, Klein, and Manning 2003) and Kulis et al.'s semi-supervised kernel k-means (SSKK) (Kulis et al. 2005), are also reported. In addition, we use the normalized cuts (NCuts) (Shi and Malik 2000), which is effectively a spectral clustering algorithm but without considering pairwise constraints, as the baseline method.

In the following, we first describe the experimental setup, including the performance measure and the parameter selection. Then we compare the proposed algorithm with the other four methods on the six data sets.

Experimental Setup

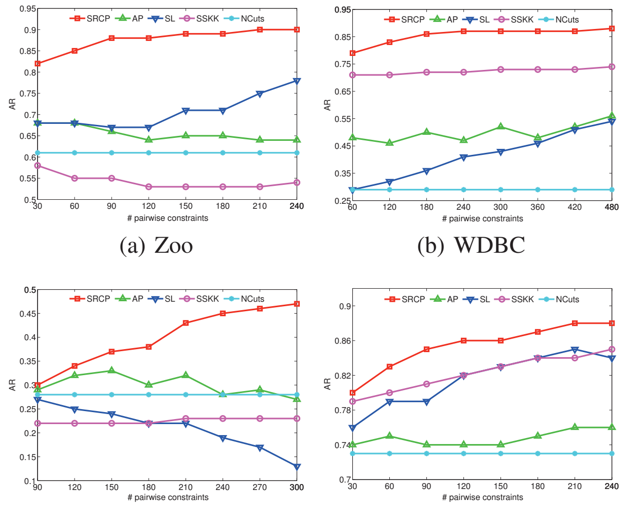
In order to evaluate these algorithms, we compare the clustering results with the available ground-truth data labels, and employ the adjusted Rand (AR) index as the performance measure (Hubert and Arabie 1985). The AR index measures the pairwise agreement between the computed clustering and the ground-truth clustering, and takes a value in the range [-1,1]. The larger is the adjusted Rand index, the better is a clustering result. To evaluate the algorithms under different settings of pairwise constraints, we exploit the ground-truth data labels and generate a varying number of pairwise constraints randomly for each data set. That is, we randomly choose a pair of data points from each data set. If they have the same class labels, we generate a must-link constraint, otherwise a cannot-link constraint. In the following experiments, we run these algorithms 20 times with random initializations, and report the averaged AR index.

Because these algorithms are all graph-based, we adopt the same k-NN graph construction for all the algorithms to ensure a fair comparison. We set $\mu = 0.2$ and $k = 20$ in our experiments. For UCI and image data sets, we construct the graph in different manners. The Gaussian similarity function is used for the UCI data sets, while the spatial Markov kernel (Lu and Ip 2009) is computed on the image data sets. All the algorithms are implemented in Matlab, running on a 2.33 GHz and 2GB RAM PC. The Lyapunov matrix equation for constraint propagation is solved using *lyap* in Matlab.

On UCI data

We first select four data sets from the UCI Machine Learning Repository[1] to test the proposed algorithm. The four UCI data sets are described in Table 1. It should be noted that

[1] http://archive.ics.uci.edu/ml/



(a) Zoo (b) WDBC

(c) Ionosphere (d) Wine

Figure 2: The clustering results on the four UCI data sets with a varied number of pairwise constraints.

the UCI data sets have been widely used to evaluate clustering algorithms in the machine learning community. For each data set, we compute the similarity matrix $W = [w_{ij}]_{n \times n}$ with Gaussian function $\exp(-\|x_i - x_j\|_2^2/(2\sigma^2))$. We simply set $\sigma = 1$ in our experiments.

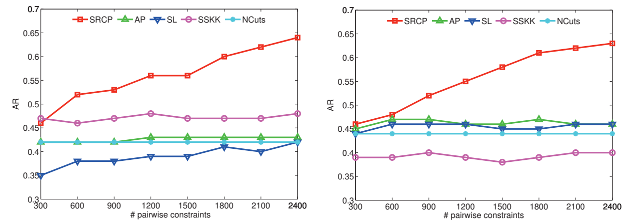
The results are shown in Figure 2, from which two observations can be drawn. Firstly, our algorithm can achieve the best performance on all the four UCI data sets. After incorporating the pairwise constraints into clustering, the improvement achieved by our method is very significant compared with NCuts, the baseline unconstrained method. The other three methods that also consider the constraint information, have inconsistent performance on the four data sets. Especially, SL on Zoo (or SSKK on Zoo and Ionosphere) even does not perform better than NCuts. Secondly, when the number of pairwise constraints grows, we can find a unanimous and obvious improvement in the performance of our SRCP on all the four data sets, but the other three constrained clustering methods (i.e. AP, SL and SSKK) do not present this trend in constrained clustering. In particular, on Ionosphere, more pairwise constraints even decrease the performance of AP and SL. To summarize, since the pairwise constraints can be exploited most effectively for clustering by our SRCP, its performance is the best.

On Image Data

We further test the proposed algorithm on two different image data sets. The first one contains 8 scene categories from MIT (Oliva and Torralba 2001), including four man-made scenes and four natural scenes. The total number of images is 2,688. The size of each image in this Scene data set is 256×256 pixels. The second data set contains images from a Corel collection. We select 15 categories including bus, sunrise/sunset, plane, foxes, horses, coins, gardens, eagles, models, sailing, stream trains, racing car, pumpkins, rockies and fields. Each of these categories contains 100 images.

Table 2: Description of the two image data sets

	Scene	Corel
# objects	2688	1500
# dimensions	256×256	256×384
# classes	8	15



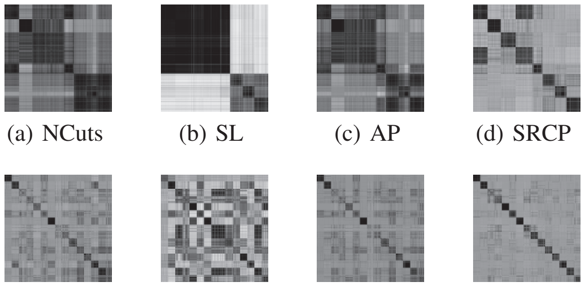
(a) Scene (b) Corel

Figure 3: The clustering results on the two image data sets with a varied number of pairwise constraints.

Therefore, this selected Corel data set has totally 1,500 images. The size of each image in this data set is 384×256 or 256×384 pixels. We summarize the description of the above two image data sets in Table 2.

For the two image data sets, we choose two different feature sets which are introduced in (Bosch, Zisserman, and noz 2006) and (Lu and Ip 2009), respectively. That is, as in (Bosch, Zisserman, and noz 2006), the SIFT descriptors are used for the Scene data set, while, similar to (Lu and Ip 2009), the joint color and Gabor features are used for the Corel data set. These features are chosen to ensure a fair comparison with the state-of-the-art techniques. More concretely, for the Scene data set, we extract SIFT descriptors of 16×16 pixel blocks computed over a regular grid with spacing of 8 pixels. As for the Corel data set, we divide each image into blocks of 16×16 pixels and then extract a joint color/texture feature vector from each block. Here, the texture features are represented as the means and standard deviations of the coefficients of a bank of Gabor filters (with 3 scales and 4 orientations), and the color features are the mean values of HSV color components. Finally, for the two data sets, we perform k-means clustering on the extracted feature vectors to form a vocabulary of 400 visual keywords. Based on this visual vocabulary, we then define a spatial Markov kernel (Lu and Ip 2009) as the similarity matrix for graph construction.

In the experiments, we compare the clustering performance of the five algorithms with a varied number of pairwise constraints. The clustering results are shown in Figure 3, from which we can see that the proposed SRCP algorithm consistently and significantly outperforms the other four algorithms on both of the two image data sets under different settings of pairwise constraints. As the number of constraints grows, the performance of our SRCP algorithm improves more significantly than those of the other three constrained clustering methods (i.e. AP, SL and SSKK). Here, it is worth noting that AP, SL and SSKK perform rather unsatisfactorily, and in some cases, their performances have even been degraded to that of NCuts.



(a) NCuts (b) SL (c) AP (d) SRCP

(e) NCuts (f) SL (g) AP (h) SRCP

Figure 4: Distance matrices of the low-dimensional representations for the two image datasets (first row: Scene; second row: Corel) obtained by NCuts, SL, AP, and SRCP, respectively. The darker is a pixel, the smaller is the distance.

To make it clearer how our SRCP algorithm exploits the pairwise constraints for clustering, we show the distance matrices of the low-dimensional data representations (generated during spectral clustering) obtained by NCuts, SL, AP and SRCP in Figure 4. We can find that the block structure of the distance matrices of the data representations obtained by our SRCP algorithm on each data set is significantly more obvious, as compared to those of the data representations obtained by NCuts, SL, and AP. This means that after being adjusted by our SRCP algorithm, each cluster associated with the new data representation becomes more compact and different clusters become more separated.

We also look at the computational cost of different clustering algorithms on the two image data sets. For example, for each run on Corel (of size 1500) with 2400 pairwise constraints, our SRCP takes about 38 seconds, while SL takes about 4 seconds, and both AP and SSKK take about 8 seconds. The main computational cost of our SRCP is incurred by solving the Lyapunov matrix equation.

Conclusions

In this paper, we have proposed a novel constraint propagation approach, called Symmetric Graph Regularized Constraint Propagation (SRCP), to propagate the sparse pairwise relationships, including must-link and cannot-link constraints, across the entire data set. This is achieved by decomposing the problem of pairwise constraint propagation into a series of two-class label propagation subproblems and considering the special symmetric structure of the pairwise relationships. More importantly, it has been shown that pairwise constraint propagation is actually equivalent to solving a Lyapunov equation which is commonly used to deal with different problems in Control Theory such as System Identification. When the constraint propagation problem is viewed in terms of information spreading over a graph, the resulting time invariant solution of the iteration propagation formula is shown to exactly be the solution of the Lyapunov equation. Experimental results on a variety of real-life data sets have demonstrated the superiority of our proposed algorithm over the state-of-the-art techniques.

Acknowledgements

This work was supported in part by the City University of Hong Kong under Grant No. 7008040, by CityU matching grant to Australian Linkage grant No. 9678017, and by the National Natural Science Foundation of China under Grant Nos. 60873154 and 61073084.

References

Barnett, S., and Storey, C. 1967. On the general functional matrix for a linear system. *IEEE Trans. Automatic Control* 12(4):436–438.

Basu, S.; Bilenko, M.; and Mooney, R. 2004. A probabilistic framework for semi-supervised clustering. In *SIGKDD*, 59–68.

Bosch, A.; Zisserman, A.; and noz, X. M. 2006. Scene classification via pLSA. In *ECCV*, 517–530.

Chung, F., ed. 1997. *Spectral Graph Theory*. American Mathematical Society.

Gajic, Z., and Qureshi, M., eds. 1995. *Lyapunov Matrix Equation in System Stability and Control*. Academic Press.

Hubert, L., and Arabie, P. 1985. Comparing partitions. *Journal of Classification* 2(1):193–218.

Kamvar, S.; Klein, D.; and Manning, C. 2003. Spectral learning. In *IJCAI*, 561–566.

Kulis, B.; Basu, S.; Dhillon, I.; and Mooney, R. 2005. Semi-supervised graph clustering: A kernel approach. In *ICML*.

Lancaster, P. 1970. Explicit solutions of linear matrix equations. *SIAM Review* 12(4):544–566.

Li, Z.; Liu, J.; and Tang, X. 2008. Pairwise constraint propagation by semidefinite programming for semi-supervised classification. In *ICML*, 576–583.

Lu, Z., and Carreira-Perpinan, M. 2008. Constrained spectral clustering through affinity propagation. In *CVPR*.

Lu, Z., and Ip, H. 2009. Image categorization by learning with context and consistency. In *CVPR*, 2719–2726.

Lu, Z., and Ip, H. 2010. Constrained spectral clustering via exhaustive and efficient constraint propagation. In *ECCV*, volume 6, 1–14.

Oliva, A., and Torralba, A. 2001. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision* 42(3):145–175.

Shi, J., and Malik, J. 2000. Normalized cuts and image segmentation. *IEEE Trans. PAMI* 22(8):888–905.

von Luxburg, U. 2007. A tutorial on spectral clustering. *Statistics and Computing* 17(4):395–416.

Xing, E.; Ng, A.; Jordan, M.; and Russell, S. 2003. Distance metric learning, with application to clustering with side-information. In *NIPS 15*, 505–512.

Zhou, D.; Bousquet, O.; Lal, T.; Weston, J.; and Schölkopf, B. 2004. Learning with local and global consistency. In *NIPS 16*, 321–328.

Zhu, X., and Goldberg, A., eds. 2009. *Introduction to Semi-Supervised Learning*. Morgan and Claypool Publishers.