Semi-supervised Learning via Adaptive Low-Rank Graph

Mingbo Zhao[1(✉)], Jiang Zhang[1], and Cuili Yang[2]

[1] Donghua University, Shanghai, People's Republic of China
mzhao4@dhu.edu.cn, jiangzhang_2016@163.com
[2] Beijing University of Technology, Beijing, People's Republic of China
clyang5@bjut.edu.cn

Abstract. Graph-based semi-supervised learning (SSL) is one of the most popular topics in the past decades. Most conventional graph-based SSL methods utilize two stage-approach to infer the class labels of the unlabeled data, i.e. it firstly constructs a graph for capturing the geometry of data manifold and then perform SSL for prediction. However, it suffers from three drawbacks: (1) the graph construction and SSL stages are separate. They do not share common information to enhance the performance of classification; (2) the graph construction and SSL should be scalable. However, most methods mainly focus on the improvement of classification accuracy but neglect the computational cost; (3) the graph should also be adaptive and robust to the parameters and datasets. However, this will usually increase computational cost making the efficiency cannot be guaranteed simultaneously. In this paper, we aim to handle the above issues. To achieve adaptiveness of SSL, we adopt a bilinear low-rank model for graph construction, where the coefficient matrix of the low-rank model is calculated through an adaptive and efficient procedure the corresponding constructed graph can capture the global structure of data manifold. Meriting from such a graph, we then propose a unified framework for scalable SSL, where we have involved the graph construction and SSL into a unified optimization problem. As a result, the discriminative information learned by SSL can be provided to improve the discriminative ability of graph construction, while the updated graph can further enhance the classification results of SSL. Simulation indicates that the proposed method can achieve better classification and clustering performance compared with other state-of-the-art graph-based SSL methods.

Keywords: Semi-supervised learning · Unsupervised learning · Spectral clustering · Adaptive low-rank model

1 Introduction

Due to the insufficiency of the labeled set, SSL, which incorporates a small number of labeled data and a large number of unlabeled data into learning, has

X. B. Zhai et al. (Eds.): MLICOM 2019, LNICST 294, pp. 434–443, 2019.
https://doi.org/10.1007/978-3-030-32388-2_37

attracted considerable attention in the artificial intelligence and pattern recognition area. Among different methods for SSL, graph-based SSL approaches, a kind of methods that model the data on a graph, have been extensive study during the past decades. The big advantage for these methods is that the graph can naturally characterize diverse types of the geometry of data manifold. According to the clustering and manifold assumptions, i.e., nearby samples (or samples of the same cluster or data manifold) share the same label [12,17,18], current graph-based SSL methods include Manifold Regularization (MR) [1], Gaussian Fields and Harmonic Functions (GFHF) [18], Learning with Local and Global Consistency (LLGC) [17], and Special Label Propagation (SLP) [10,11]. These methods usually model labeled and unlabeled data by a graph, and then calculate the graph Laplacian matrix to capture the geometrical structure of data manifold [3].

While most conventional graph-based SSL generally perform well in many real-world applications, a good graph-based SSL model should satisfy the following issues: (1) the graph construction should be adaptive and robust to the parameters and datasets. In the conventional graph-based SSL, the graph usually represents a kNN graph associated with weight on it. Many ways are proposed to define the graph weight which include Gaussian function [10,11,17,18], Locally Linear Reconstruction [13,14], Local Regression and Global Alignment [16] and Local Spline Regression [15]. A key limit for these methods is that the number of k needs to be carefully adjusted hence they are not adaptive. Fortunately, this drawback can be solved by the sparse or low-rank representation graph, where SR based graphs can model the data with good properties of adaptiveness, sparsity and high discriminating power, while LRR based graphs can characterize the global structure of data. However, the huge computational cost is needed in order to solve l_1 or trace-norm minimization problem for calculating the sparse or low-rank graph; (2) the graph construction and SSL strategy should also be efficient and scalable to large-scale data. As analyzed in [7,8], the computational cost for searching the k neighbors of data in conventional kNN graph is $O\left(kn^2\right)$. While those for calculating the weight matrix in the SR graph and LRR graph are $O\left(n^3\right)$. None of them is linear with the number of datasets.

To handle this problem, Liu et al. [7,8] have proposed an efficient anchor graph framework by exploring a set of anchors from data points, where it is first to establish a similarity matrix between data points and anchors, and then to construct the anchor graph for inferring the class labels of anchors instead of the whole data points. As a result, the computational cost can be reduced to be linear with n. Many variants of AGR have been proposed during the past few years. However, the similarity matrix in AGR and its variants are still not adaptive and need to be adjusted according to a different dataset. In other words, it is quite hard to guarantee the adaptiveness and scalability simultaneously; (3) the graph construction and SSL stages are separate. They do not share common information to enhance the performance of classification. In another word, the graph won't be updated once it is constructed. However, the estimated class

labels will include some discriminative information and can further be utilized to update the graph. As a result, the class label inference can further be enhanced.

In this paper, we aim to solve the above problems by developing a new framework for scalable SSL. Specifically, in order to achieve adaptiveness for SSL, we adopt a bilinear low-rank model for graph construction, where the coefficient matrix of the model is calculated through an adaptive and efficient way. We thereby construct the graph based on such weight matrix following the basic concept of AGR. As a result, the adaptiveness and scalability can both be achieved. The corresponding graph can also capture the global structure of the data manifold. Meriting from such a graph, we then propose a unified framework for scalable SSL, where we have involved the graph construction and SSL into a unified optimization problem. As a result, the discriminative information learned by SSL can be provided to improve the discriminative ability of graph construction, while the updated graph can further enhance the classification results of SSL. Simulation indicates that the proposed method can achieve better classification results compared with other state-of-the-art graph-based SSL methods.

The main contributions of this paper are as follows:

(1) We have developed an adaptive bilinear low-rank model for graph construction. With the group sparsity and non-negative constraint, the low-rank value can be automatically determined, and the learned coefficient matrix S is non-negative and can characterize the global structure of data;

(2) We have developed a new graph-based SSL framework, in which the developed low-rank model and SLP are unified into a single optimization problem. In this way, the discriminative information learned by SLP can be provided to improve the discriminative ability of graph construction, while the updated graph can further enhance the classification results of SSL.

(3) We have developed an efficient iterative approach for optimization. Theoretical analysis has guaranteed the convergence and the computational cost is linear with the number of data points. Thereby, the solution is efficient and scalable to large-scale data;

The rest of this paper is organized as follows: In Sect. 1, we will provide some basic notations and reviews of related work; in Sect. 2, we will present the proposed bilinear low-rank model for graph construction. We then develop a unified framework for graph-based SSL. Extensive simulations are conducted in Sect. 3 and final conclusions are drawn in Sect. 4.

2 Adaptive Low-Rank Graph Regularization for Semi-supervised Learning

2.1 Graph Construction via Adaptive Low-Rank Model

Denote $X = [X_l, X_u] \in R^{d\times(l+u)}$ as the data matrix, in which d is the number of features, the first l and the remaining u data points in X form the labeled set X_l and unlabeled set X_u, respectively, $Y = [y_1, y_2, \ldots, y_{l+u}] \in R^{c\times(l+u)}$ is the original class labels of all data that satisfies: $y_{ij} = 1$, given x_j is within the ith class;

otherwise, $y_{ij} = 0$. Accordingly, denote $F = [f_1, f_2, \ldots, f_{l+u}] \in R^{c\times(l+u)}$ be the estimated class label matrix, where f_i is a column vector satisfying $0 \le f_{ij} \le 1$. The AGR has assumed each data point can be approximately reconstructed by its nearby anchors, i.e. $x_j = \sum_{i=1}^{m} a_i Z_{ij}$ or $X \approx AZ$. The coefficient Z_j for each x_j is then calculated by Kernel-defined weights or local reconstructed strategy one by one, i.e.

$$\min_{Z} \frac{1}{2} \|x_j - Az_j\|_F^2 \quad s.t.\ z_{ij} \ge 0,\ \sum\nolimits_{i=1}^{m} z_{ij} = 1 \tag{1}$$

However, a key problem for the above strategies is Z is not adaptive since some key parameters (such as the number of anchors m) need to be carefully adjusted. Another problem is that Z cannot preserve the global structure of data manifold since each coefficient vector is only associated with nearby anchor data. On the other hand, the low-rank model can capture the global information as well as achieve data-adaptiveness. We thereby develop an adaptive low-rank model to calculate the weight matrix. Specifically, we first reformulate Eq. (1) as follows:

$$\min_{A,Z} \frac{1}{2} \|X - AZ - O\|_F^2 + \gamma \|O\|_1. \tag{2}$$

where we let $X = AZ + O$ and $O \in R^{d\times n}$ is the additive matrix measuring the corruption of X, $\|O\|_1$ is the sparse l_1-norm of O since we assume the corruptions usually affects some entries of X making O is sparse. In order to grasp the global structure of the whole data and achieve data-adaptiveness, we add a low-rank constraint on AZ, then Eq. (2) can be formulated as:

$$\min_{A,Z} \frac{1}{2} \|X - M - O\|_F^2 + \lambda \|M\|_* + \gamma \|O\|_1 \quad s.t.\ M = AZ. \tag{3}$$

where $\|M\|_*$ is the nuclear norm approximating the rank of M. In addition, as pointed in [], the nuclear norm M can be further reformulated as the penalty of bilinear factorizations, i.e.

$$\|M\|_* = \min_{A,Z} \frac{1}{2} \|A\|_F^2 + \frac{1}{2} \|Z\|_F^2 \quad s.t.\ AZ = M. \tag{4}$$

where M is optimized via the SVD of $M = U\Sigma V^T$ so that $A = U\Sigma^{1/2} \in R^{d\times q}$ and $Z = \Sigma^{1/2} V^T \in R^{q\times n}$, q is the low-rank value of M. Here, let $P = \left[A, Z^T\right]^T \in R^{q\times(d+n)}$ denote a joint matrix, we have:

$$\|P\|_F^2 = \left[A, Z^T\right]^T = \|A\|_F^2 + \|Z\|_F^2. \tag{5}$$

Obviously, the low-rank value is the number of anchors, i.e. $q = m$, then Eq. (5) can be roughly equivalent to the following problem by combining Eqs. (3), (4) and (5):

$$\min_{B,S,O} \frac{1}{2} \|X - AZ - O\|_F^2 + \frac{\lambda}{2} \|P\|_F^2 + \gamma \|O\|_1 \tag{6}$$

Note that optimizing $l_{2,1}$-norm term $\|P\|_{2,1}$ in Eq. (6) enable the columns of P sparse, i.e. some columns of P are non-zero while others are close to zero. Since each column of P, i.e. p_j, is formed by an anchor a_j and its corresponding coefficient z_j, setting the norm of p_j to zero means a_j is less important and can be neglected, while the p_j with non-zero value means the corresponding a_j is more important. As a result, the most important anchors a_j combined with its coefficient z_j can be selected when solving the optimization problem. Therefore, the optimal value of m can be adaptively selected.

After we obtain the weight matrix Z, we can construct the similarity matrix of graph [7,8] in a low-rank form as follows:

$$W_s = Z^T Z. \tag{7}$$

where the inner product is regarded as the adjacent weight between any pairwise x_i and x_j. In other words, if x_i and x_j share common anchors, their s_i and s_j will be similar making W_{ij}^s to be a large value; otherwise, W_{ij}^s will be close to 0, if x_i and x_j do not have any anchors. Hence W^s can also reflect the geometry of data manifold.

2.2 Problem Formulation

It should be noted the labeled information is very effective to improve the discriminative ability of the graph if one can involve limited label information into the graph construction. However, as shown above, the partial label information is not utilized in graph construction. On the other hand, the SLP is to propagate the class label information of labeled set to unlabeled set, where the class labels of unlabeled set can be predicted. This motivates us to consider utilizing the additional labeled information to improve the discriminant of affinity matrix. Motivated by this end, we develop an effective and scalable approach to solve the above problem, where we integrate the adaptive graph construction and SLP into a unified framework. As a result, the discriminative information can be involved to guide the graph construction, while the newly updated graph construction can further improve the classification results for SSL. In addition, both the graph construction and SSL share a unified objective function which can be simultaneously optimized in one step and to guarantee the overall optimum. Specifically, we give our model for SSL as follows:

$$\begin{aligned} \min_{B,S,O} & \tfrac{1}{2}\left\|X - B^T S - O\right\|_F^2 + \tfrac{\beta}{2}\|P\|_{2,1} + \gamma\|O\|_1 \\ & + \alpha\left(\tfrac{1}{2}Tr\left(FL_sF^T\right) + Tr\left(F-Y\right)UD_s(F-Y)^T\right) \\ & s.t.\ B \geq 0,\ S \geq 0,\ O \geq 0,\ \forall j,\ \|s_j\|_0 \leq T_0 \end{aligned} \tag{8}$$

where L_s is the graph Laplacian matrix for W_s, $B \geq 0$, $S \geq 0$, $O \geq 0$ are the non-negative constraints for guaranteeing the non-negativity for $X = B^T S + O$.

2.3 Solution

We will develop an iterative approach to handle the problem of Eq. (8). It can be noted that $\|P\|_{2,1}$ can be written as $\|P\|_{2,1} = Tr\left(S^T G S\right) + Tr\left(B^T G B\right)$,

$G \in R^{q\times q}$ is a diagonal matrix satisfying:

$$[G \leftarrow \begin{bmatrix} \frac{1}{2\|z_1\|_2} & & \\ & \ddots & \\ & & \frac{1}{2\|z_q\|_2} \end{bmatrix} \tag{9}$$

where $\|p_i\|_2$ is the norm of the *i-th* column of P. We also denote $R \in R^{n\times n}$ as a sparse matrix with each element satisfying $R_{ij} = \|f_i - f_j\|_F^2$, so that the following equation holds:

$$\begin{aligned} Tr\left(FL_sF^T\right) &= \sum_{i,j=1}^{n} \|f_i - f_j\|_F^2 (W_s)_{ij} \\ &= \sum_{i,j=1}^{n} (R \odot W_s) = Tr\left(SRS^T\right) \end{aligned} \tag{10}$$

where $\odot$ is the pair-wise product and the third equation is satisfied as $\sum_{i,j=1}^{n} (R \odot W_s) = Tr(RW_s) = Tr\left(SRS^T\right)$. We then develop multiplicative updating rules by formulating the Lagrange function to the problem of Eq. (10) with non-negative constraints as follows:

$$\begin{aligned} \min_{B,S,O} &\tfrac{1}{2}\|X - BS - O\|_F^2 + \tfrac{\beta}{2} Tr\left(S^T GS + B^T GB\right) \\ &+\alpha\left(Tr\left(FL_sF^T\right) + Tr\left(F - Y\right)UD_s(F - Y)^T\right) \\ &+\gamma Tr\left(E^T O\right) + Tr\left(\phi O\right) + Tr\left(\varphi B\right) + Tr\left(\psi S\right) \end{aligned} \tag{11}$$

where ϕ, φ and ψ are three Lagrange multipliers to constrain $O_{ij} \geq 0$, $B_{ij} \geq 0$ and $S_{ij} \geq 0$. Here, by setting the derivative w.r.t. B_{ij}, S_{ij} and O_{ij} to zero and utilizing the Karush-Kuhn-Tuckre (KKT) condition $\phi_{ij}O_{ij} = 0$, $\varphi_{ij}B_{ij} = 0$ and $\psi_{ij}S_{ij} = 0$, O_{ij}, B_{ij} and S_{ij}, can be updated as follows:

$$O_{ij} \leftarrow O_{ij} \frac{\left(X - B^T S\right)_{ij}}{\left(O + \gamma E\right)_{ij}} \tag{12}$$

$$B_{ij} \leftarrow B_{ij} \frac{\left(S(X - O)^T\right)_{ij}}{\left(SS^T B + \beta GB\right)_{ij}} \tag{13}$$

$$S_{ij} \leftarrow S_{ij} \frac{\left(B\left(X - O\right)\right)_{ij}}{\left(BB^T S + \beta GS + \alpha SR\right)_{ij}} \tag{14}$$

It should be noted that the main computation for calculating the optimal solution F is to perform the inverse of $L_s + UD_s$, in which the complexity is $O\left(n^3\right)$. However, such computational complexity for F can be dramatical given the data is large-scale. Fortunately, by the form of $W_s = S^T S$, Eq. (14) can be rewritten as follows:

$$\begin{aligned} F &= YUD_s(L_s + UD_s)^{-1} \\ &= YU\left(I - S^T SD_s^{-1} + U\right)^{-1} \\ &= YI_\alpha\left(I - S^T SD_s^{-1} I_\beta\right)^{-1} \\ &= YI_\alpha\left(I + S^T\left(I_q + SI_\beta D_s^{-1} S^T\right)^{-1} SD_s^{-1} I_\beta\right) \end{aligned} \tag{15}$$

Following Eq. (15), we can observe that the computational complexity for performing the inverse of F reduces from $O\left(n^3\right)$ to $O\left(q^3\right)$. Given $q \ll n$, the calculation for F can be significantly speeded up by Eq. (15), which is good for dealing with large-scaled data.

3 Simulations

3.1 Dataset Description

We in this section conduct extensive simulations on three synthetic data as well as several real-world datasets to evaluate the effectiveness of the proposed method. In the synthetic dataset, we evaluate the proposed method based on two-swiss-roll and two-moon datasets, where each dataset has two classes and each class per dataset follows a two-swiss-roll and two-moon distribution. In real-world datasets, we illustrate the performance of the proposed method as well as compare with those of other state-of-the-art SSL methods based on six real-world datasets, which include Extended Yale-B dataset [2], COIL100 [9], ETH80 [6].

For each dataset, we randomly annotate 5%, 10%, 15% and 20% data from each class to form labeled set while the remaining data is selected as an unlabeled set and 20% data per class is selected as a test set.

3.2 Image Classification

We in this subsection evaluate our method for image classification and compare the results with other graph-based SSL methods, which include LGC, SLP, LNP, AGR, EMR and Manifold Regularization (MR). We also choose SVM as a baseline in our simulation. For the parameter k in LGC, SLP, LNP, AGR, EMR and MR to formulate the k neighborhood graph, we use five-fold cross-validation to determine the best value, where the candidates are were chosen from 6 to 20. For the parameter σ used in LGC, SLP EMR and MR as the Gaussian variance, we utilize the same approach in [4] to choose its best value. For LGC, LNP AGR, EMR and the proposed method, they need to determine the regularized parameter, and we set the candidates from $\left\{10^{-6}, 10^{-3}, 10^{-1}, 1, 10, 10^3, 10^6\right\}$ by using five-fold cross-validation.

The simulation results over 20 random splits with varied numbers of labeled data for different methods are shown in Table 1. We can have the following results:

(1) The classification accuracies become higher given the number of labeled data is increased. In detail, the accuracy of the proposed method is increased by approximately 15% given the labeled data varies from 5% to 20% in most cases. This can even be achieved by about 17% for the CASIA-HWDB dataset. This indicates that the labeled data actually be useful for image classification. In addition, we can also see that the accuracies will not change any more given sufficient labeled data;

(2) among all SSL methods, the proposed method can almost obtain the best results in all cases. For example, the proposed method can achieve 4%–7% superiority to LGC, SLP and LNP, in most cases. This improvement can even achieve 8% in the CASIA-HWDB dataset. In addition, AGR and EMR can obtain competitive accuracy to the proposed method in most cases. But the results of AGR and EMR are achieved by carefully adjusting the parameters, while the proposed method can adaptively calculate the weight matrix for constructing the graph;

(3) another observation is that similar to the unlabeled data, the accuracies of test data of the proposed method, LNP, AGR and MR increase given that the number of labeled data increases. Specifically, the accuracy of the test set will increase by approximately 10% given the number of labeled data increases from 5% to 20% for most cases. The accuracies of unlabeled data are better than those of test data. This can be natural due to the reason that the test data are not be used in training the method as unlabeled data.

Table 1. Classification accuracies of different datasets

Datasets	Methods	5% Training Labeled		10% Training Labeled		15% Training Labeled		20% Training Labeled	
		Unlabeled	Test	Unlabeled	Test	Unlabeled	Test	Unlabeled	Test
Yale-B [5]	SVM	53.1±1.1	52.7±1.0	68.8±2.0	67.7±0.6	75.2±1.1	73.7±1.3	80.0±1.8	78.8±1.2
	MR	59.0±1.2	58.5±1.3	70.3±1.1	69.4±0.5	76.4±1.3	74.9±1.5	80.7±1.3	79.0±1.1
	LGC	64.7±1.0		71.8±1.1		76.4±4.2		80.8±1.0	
	SLP	65.6±2.3		73.9±1.0		78.0±1.8		81.8±1.0	
	LNP	64.9±1.3	53.8±2.7	72.0±1.2	71.2±0.4	78,0±2.4	76.6±2.1	81.6±1.0	80.0±1.4
	AGR	66.6±1.5	65.8±1.3	74.3±1.2	72.2±0.4	78.1±1.5	77.3±1.7	83.0±1.2	80.0±4.5
	EMR	66.9±0.8		74.4±1.1		78.0±1.5		84.4±2.4	
	EAGR	69.9±0.4	67.2±1.0	75.7±1.1	74.0±3.3	79.4±1.0	78.3±1.1	86.3±2.5	82.8±2.4
	ALG	**69.9±0.4**	**67.2±1.0**	**75.7±1.1**	**74.0±3.3**	**79.4±1.0**	**78.3±1.1**	**86.3±2.5**	**82.8±2.4**
COIL100 [9]	SVM	83.6±0.9	83.2±0.8	88.5±0.8	86.6±0.8	91.8±0.8	91.4±0.7	95.3±0.8	94.5±1.6
	MR	83.7±1.0	83.4±0.9	89.0±0.9	87.3±0.9	92.1±0.8	91.6±0.9	95.3±0.7	94.7±1.3
	LGC	85.5±0.8		89.3±0.9		92.4±0.8		95.5±0.6	
	SLP	86.4±0.7		89.3±0.9		92.8±0.6		95.6±0.8	
	LNP	86.5±0.7	85.6±0.7	89.6±0.9	88.7±0.7	92.9±0.7	92.4±0.8	95.8±0.7	95.1±1.3
	AGR	86.5±0.6	85.8±0.9	90.9±0.9	88.8±0.8	93.3±0.6	92.7±0.9	95.8±0.7	95.3±1.4
	EMR	86.6±0.7		89.9±0.9		93.2±0.6		96.0±0.7	
	ALG	**87.0±0.6**	**86.7±1.0**	**91.8±0.9**	**89.7±0.8**	**94.7±0.6**	**93.2±0.8**	**97.0±0.6**	**95.6±0.9**
ETH80 [6]	SVM	61.1±1.3	59.4±0.3	71.1±1.9	70.2±2.0	75.9±1.5	75.3±3.1	78.9±2.0	77.9±2.5
	MR	62.3±0.8	60.0±0.2	71.7±2.0	71.0±2.7	76.2±1.0	75.3±2.8	78.9±1.9	78.3±2.5
	LGC	65.7±1.4		73.5±1.4		76.8±1.5		79.0±1.7	
	SLP	65.9±1.5		73.9±1.2		76.9±1.6		79.3±1.8	
	LNP	64.9±0.9	62.2±0.2	73.4±2.0	71.4±2.6	76.7±1.1	76.0±2.6	79.0±1.8	78.5±2.0
	AGR	66.4±1.6	65.1±0.2	75.0±1.7	72.2±2.2	76.9±1.7	76.1±2.5	79.6±2.0	78.9±1.9
	EMR	68.2±1.7		74.9±1.4		77.3±1.7		80.0±2.2	
	ALG	**69.4±1.9**	**67.2±0.1**	**74.0±1.3**	**74.2±2.2**	**77.5±1.9**	**77.3±1.8**	**79.8±2.2**	**79.0±2.2**

4 Conclusion

In this paper, we develop a graph-based semi-supervised learning framework for image classification and clustering. According to the theoretical analysis and simulation results, we can draw the following conclusions: (1) we have developed an adaptive bilinear low-rank model for graph construction. With the group sparsity and non-negative constraint, the low-rank value can be automatically determined, and the learned coefficient matrix S is non-negative and can grasp the global structure of whole data; (2) We have developed a new graph-based

SSL framework, in which the developed low-rank model and SLP are unified into a single optimization problem. In this way, the discriminative information learned by SLP can be provided to improve the discriminative ability of graph construction, while the updated graph can further enhance the classification results of SSL.

Acknowledgment. The work is supported by the National Natural Science Foundation of China (61601112). It is also supported by the Fundamental Research Funds for the Central Universities and DHU Distinguished Young Professor Program.

References

1. Belkin, M., Niyogi, P., Sindhwani, V.: Manifold regularization: a geometric framework for learning from labeled and unlabeled samples. J. Mach. Learn. Res. **7**, 2399–2434 (2006)
2. Georghiades, A.S., Belhumeur, P.N., Kriegman, D.J.: From few to many: illumination cone models for face recognition under variable lighting and pose. IEEE Trans. Pattern Anal. Mach. Intell. **23**(6), 643–660 (2001)
3. He, X., Yan, S., Hu, Y., Niyogi, P., Zhang, H.: Face recognition using Laplacian faces. IEEE Trans. Pattern Anal. Mach. Intell. **27**(3), 328–340 (2005)
4. Hou, C., Nie, F., Wang, F., Zhang, C., Wu, Y.: Semisupervised learning using negative labels. IEEE Trans. Neural Netw. **22**(3), 420–432 (2011)
5. Lee, K., Ho, J., Kriegman, D.: Acquiring linear subspaces for face recognition under variable lighting. IEEE Trans. Pattern Anal. Mach. Intell. **27**(5), 947–963 (2005)
6. Leibe, B., Schiele, B.: Analyzing appearance and contour based methods for object categorization. In: Proceedings of 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, p. II-409. IEEE (2003)
7. Liu, W., He, J., Chang, S.F.: Large graph construction for scalable semi-supervised learning. In: Proceedings of the 27th International Conference on Machine Learning (ICML-2010), pp. 679–686 (2010)
8. Liu, W., Wang, J., Chang, S.F.: Robust and scalable graph-based semisupervised learning. Proc. IEEE **100**(9), 2624–2638 (2012)
9. Nene, S.A., Nayar, S.K., Murase, H.: Columbia object image library (coil-100). Technical report, Technical Report CUCS-005-96, Columbia University (1996)
10. Nie, F., Xiang, S., Liu, Y., Zhang, C.: A general graph based semi-supervised learning with novel class discovery. Neural Comput. Appl. **19**(4), 549–555 (2010)
11. Nie, F., Xu, D., Li, X., Xiang, S.: Semisupervised dimensionality reduction and classification through virtual label regression. IEEE Trans. Syst. Man Cybern. Part B (Cybern.) **41**(3), 675–685 (2011)
12. Szummer, M., Jaakkola, T.: Patially labeled classification with Markov random walks. In: NIPS (2002)
13. Wang, F., Zhang, C.: Label propagation through linear neighborhoods. IEEE Trans. Knowl. Data Eng. **20**(1), 55–67 (2008)
14. Wang, J., Wang, F., Zhang, C., Shen, H.C., Quan, L.: Linear neighborhood propagation and its applications. IEEE Trans. Pattern Anal. Mach. Intell. **31**(9), 1600–1615 (2009)
15. Xiang, S., Nie, F., Zhang, C.: Semi-supervised classification via local spline regression. IEEE Trans. Pattern Anal. Mach. Intell. **32**(11), 2039–2053 (2010)

16. Yang, Y., Nie, F., Xu, D., Luo, J., Zhuang, Y., Pan, Y.: A multimedia retrieval framework based on semi-supervised ranking and relevance feedback. IEEE Trans. Pattern Anal. Mach. Intell. **34**(4), 723–742 (2012)
17. Zhou, D., Bousquet, O., Lal, T.N., Weston, J., Scholkopf, B.: Learning with local and global consistency. In: NIPS (2004)
18. Zhu, X., Ghahramani, Z., Lafferty, J.D.: Semi-supervised learning using gaussian fields and harmonic functions. In: ICML (2003)