Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence

Low-Rank Coding with b-Matching Constraint for Semi-Supervised Classification*

Sheng Li
Department of ECE
Northeastern University, Boston, MA, USA
shengli@ece.neu.edu

Yun Fu
Department of ECE and College of CIS
Northeastern University, Boston, MA, USA
yunfu@ece.neu.edu

Abstract

Graph based semi-supervised learning (GSSL) plays an important role in machine learning systems. The most crucial step in GSSL is graph construction. Although several interesting graph construction methods have been proposed in recent years, how to construct an effective graph is still an open problem. In this paper, we develop a novel approach to constructing graph, which is based on low-rank coding and b-matching constraint. By virtue of recent advances in low-rank subspace recovery theory, compact encoding using low-rank representation coefficients allows us to obtain a robust similarity metric between all pairs of samples. Meanwhile, the b-matching constraint helps in obtaining a sparse and balanced graph, which benefits label propagation in GSSL. We build a joint optimization model to learn low-rank codes and balanced graph simultaneously. After using a graph re-weighting strategy, we present a semi-supervised learning algorithm by incorporating our sparse and balanced graph with Gaussian harmonic function (GHF). Experimental results on the Extended YaleB, PIE, ORL and USPS databases demonstrate that our graph outperforms several state-of-the-art graphs, especially when the labeled samples are very scarce.

1 Introduction

In many real world applications, unlabeled data (e.g. images, texts, and videos) are easier to obtain than labeled ones. For classification tasks, traditional unsupervised learning methods, such as principal component analysis (PCA) [Turk and Pentland, 1991], are unable to utilize the class label information, while supervised methods such as linear discriminant analysis (LDA) [Bellhumeur *et al.*, 1997] do not use the unlabeled samples during training. For these reasons, semi-supervised learning (SSL) has attracted much attention in recent years, because SSL learns from both labeled and unlabeled samples [Zhu, 2010; Culp and Michailidis, 2008].

Among various SSL techniques, graph based SSL (GSSL) has been extensively studied and widely applied to many applications due to its inspiring performance on accuracy and speed [Chapelle *et al.*, 2006; Kveton *et al.*, 2010; Liu *et al.*, 2012b]. Zhu *et al.* proposed to use the harmonic property of Gaussian random field over the graph for SSL [Zhu *et al.*, 2003]. Zhou *et al.* performed SSL with the local and global consistency [Zhou *et al.*, 2003]. He *et al.* developed a generative model for GSSL by estimating priors and conditional probabilities [He *et al.*, 2007]. Cai *et al.* proposed a semi-supervised discriminant analysis (SDA) method, which employs a graph based smoothness regularization term [Cai *et al.*, 2007].

Although many GSSL methods have been proposed, limited research focuses on how to construct effective graphs [Wang and Xia, 2012]. k-nearest neighbor (k-NN) and ϵ-neighborhood are two popular graph construction schemes. However, neither methods generate graphs that are balanced, or regular, which hinders learning. b-matching method can solve this problem by learning a symmetric and balanced graph [Jebara *et al.*, 2009]. Unfortunately, these resulting graphs are highly dependant on the similarity function, and are therefore sensitive to sample variation or noise.

To be effective, SSL methods should obey the smoothness, cluster or manifold assumptions [Chen and Wang, 2007]. However, when the samples contain noise or large variations, these assumptions are often violated, and therefore the traditional similarity metric (e.g., Gaussian function) often fails. Fortunately, recent research shows that sparsity and low-rankness could successfully recover signals in noise [Wright *et al.*, 2009; Candès *et al.*, 2011; Liu *et al.*, 2010; 2013; Li *et al.*, 2013]. Some graphs based on sparse representation [Yan and Wang, 2009; He *et al.*, 2011] or low-rank representation [Zhuang *et al.*, 2012] have been proposed for SSL. These graphs have shown promising performance in some applications. Nevertheless, there are still some limitations. First, the reason for using sparse or low-rank representation coefficients as graph weights is not very clear. Second, these graphs are unbalanced, which could seriously degrade the classification performance.

All these limitations in existing graphs motivate us to de-

*This research is supported in part by the NSF CNS award 1135660 and 1314484, Office of Naval Research award N00014-12-1-0125 and N00014-12-1-1028, Air Force Office of Scientific Research award FA9550-12-1-0201, and IC Postdoctoral Research Fellowship award 2011-11071400006.

sign a novel graph construction model, with robust similarity metric and balanced property. Recent advances on low-rank representation suggest that noisy data drawn from different subspaces can be correctly recovered by seeking the lowest-rank encoding matrix for all samples [Liu *et al.*, 2010; 2013]. Inspired by this observation, we propose to measure the similarities between different samples in the low-rank encoding space, instead of original sample space. To generate a sparse and balanced graph, we also incorporate the b-matching constraint. Then we build a joint optimization framework to learn low-rank codes and balance the graph simultaneously. After using a straightforward graph re-weighting strategy, we present a semi-supervised learning algorithm by incorporating our low-rank coding based balanced (LRCB) graph with Gaussian harmonic function (GHF). Experimental results on the Extended YaleB, PIE, ORL and USPS databases demonstrate that our graph outperforms several state-of-the-art graph algorithms, especially when there is very small set of labeled samples.

The rest of this paper is organized as follows. In Section 2, we briefly review related work. Then we describe the proposed graph construction method in Section 3, and present a GSSL algorithm in Section 4. Experiments are reported in Section 5 before concluding in Section 6.

2 Related Work

Constructing an effective graph is the most important component in GSSL. As we discussed above, there has been some research specifically that targeted graph construction. Given a sample set X with n samples, graph construction aims at building a weighted graph G. Typically, graph construction models consist of three procedures: similarity metric between each pair of samples, sparsification of the graph, and graph re-weighting. Then we have the graph $G = S \circ W$, where S is a binary matrix whose elements indicate the edges in graph. W can be the similarity matrix or other user-defined weighting matrices, and $\circ$ is the Hadamard product.

In the first step, kernel function $k(\cdot)$ (e.g., Gaussian kernel) is usually used to estimate the similarity between samples. After constructing a similarity matrix, a simple and widely-used approach for sparsification and re-weighting is the k nearest neighborhood (k-NN) method [Maier *et al.*, 2008]. It greedily selects k significant neighbors for each node to construct a k-NN graph. As shown in [Jebara *et al.*, 2009], constructing k-NN graph is equivalent to the following optimization problem:

$$\begin{aligned} &\max_{S_{ij}} \sum_{i,j=1}^{n} S_{ij} W_{ij} \\ &\text{s.t.} \sum_{j=1}^{n} S_{ij} = k, S_{ii} = 0, S_{ij} \in \{0, 1\}. \end{aligned} \tag{1}$$

From the objective function (1), we can observe that k-NN method cannot produce a symmetric graph, and therefore it always requires a symmetrization process after obtaining S. However, the resulting k-NN graph is not balanced, which may lead to a performance degradation in SSL. To address this problem, Jebara *et al.* proposed a b-matching method to construct a balanced graph [Jebara *et al.*, 2009]. They employ a b-matching constraint when constructing the graph, and therefore the degree of every node is exactly equal to the constant b. This method simultaneously ensures both symmetry and balance of the graph in learning procedures. However, the performance of k-NN graph and b-matching graph is highly dependent on the similarity metric. If the metric is sensitive to sample variation or noise, the performance would be greatly degraded.

Another class of graph construction methods focuses on weighting the graphs. Yan *et al.* proposed an l_1 graph, which utilizes sparse representation coefficients as the weights on graph [Yan and Wang, 2009]. l_1 graph is constructed in a parameter-free manner, since the adjacency structure and graph weights are determined by the l_1 optimization algorithm automatically. Furthermore, He *et al.* presented a non-negative sparse probability graph (SPG) to further improve the classification performance of SSL [He *et al.*, 2011]. However, the reason for using sparse representation coefficients as graph weights is not quite clear. Especially when the labeled samples are very limited, sparse coefficients are very sensitive to the variation in labeled samples, and then the label propagation results would be affected negatively. In addition, both l_1 graph and SPG graph are not balanced.

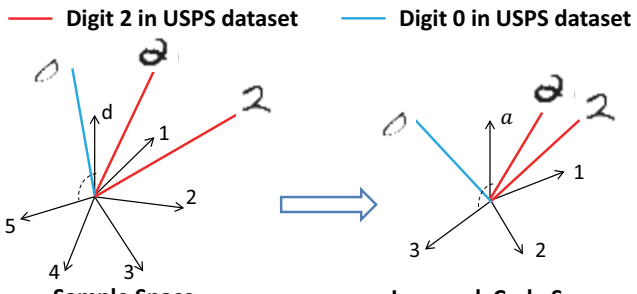
Recently, Zhuang *et al.* proposed a non-negative low-rank and sparse (NNLRS) graph for semi-supervised classification [Zhuang *et al.*, 2012]. Their graph construction enforces low-rankness and sparsity simultaneously, and employs the low-rank representation coefficients as graph weights. There are several key differences between NNLRS and our approach. First, NNLRS directly utilizes low-rank representation coefficients as graph weights, our approach, on the other hand, employs low-rank coefficients of each pair of samples to calculate their similarity. Second, our approach jointly learns a sparse and balanced graph with b-matching constraint, while NNLRS graph is dense and unbalanced, which is undesired for semi-supervised classification in terms of accuracy and computational cost.

3 Low-Rank Coding based Balanced Graph Construction

Assumed that the sample set X contains a set of l labeled samples X_l and a set of u unlabeled samples X_u, $X = [x_1, \cdots, x_l, x_{l+1}, \cdots, x_{l+u}] = [X_l, X_u]$, $n = l + u$ (each column is a sample, $x_i \in \mathbb{R}^d$). The aim of SSL is to infer the missing labels of X_u with the aid of labeled sample set X_l. In this section, we first present a robust similarity metric in the low-rank space. Furthermore, to learn a sparse and balanced graph, we build a joint optimization model to learn the similarity matrix and the balanced graph simultaneously.

3.1 Problem Formulation

Many current similarity metrics, such as Gaussian function, are sensitive to noise or large intra-class variations. Inspired by the recent advances on low-rank representation based subspace recovery, we propose to estimate similarity in the low-rank code space to address this problem. Let Z denote the unknown coefficient matrix for sample set X, low-rank method



Similarity	Original space	Low-rank coding space
Intra-class (digit 2)	0.78	0.92
Inter-class (digits 0 and 2)	0.75, 0.32	0.24, 0.10

Figure 1: Similarity metrics in the original space and low-rank code space.

learns a matrix Z which has the lowest $rank(Z)$ and satisfies the constraint $X = AZ + E$, where A is an over-complete dictionary and E is sparse noise component [Liu *et al.*, 2013]. The i-th column vector in Z is the low-rank code vector for x_i. Due to the fact that Z can correctly recover multiple subspaces and its low-rankness, the low-rank codes in Z belonging to one subspace (i.e., one class) should be highly correlated. Liu *et al.* also show that low-rank coefficients are very robust to different kinds of noises [Liu *et al.*, 2013]. Thus, it's reasonable to employ low-rank codes in estimating the similarity between all pairs of samples.

To clearly illustrate our new similarity metric, Fig. 1 compares two similarity metrics of digits images in the original space and low-rank code space. It shows that, in case of large intra-class variation or large inter-class correlation, similarity values calculated in original space may not be reliable. However, the noise-insensitive low-rank codes can correctly recover the subspace structures of multiple classes, and low-rankness means that codes belonging to the same class should have high correlations. Thus, the similarity metric in low-rank space obtains better results as shown in Fig. 1.

Sparsity is another important requirement for effective GSSL, which helps improve the classification accuracy and computational performance of GSSL algorithms [Wang and Xia, 2012]. As discussed above, b-matching constraint is well-suited for this requirement. We propose to build a joint optimization model by learning both low-rank codes and balanced graph. The objective function of our model is:

$$\begin{aligned}&\min_{Z,L,R,E,S} \|Z - LR\|_F^2 + \lambda_1 \|E\|_{2,1} - \lambda_2 \sum_{i,j=1}^{n} S_{ij}(Z^T Z)_{ij} \\ &\text{s.t. } X = AZ + E, \mathbf{1}_a^T Z = \mathbf{1}_n^T, \sum_{j=1}^{n} S_{ij} = b, S_{ij} = S_{ji}.\end{aligned} \tag{2}$$

where A is the dictionary with the size of $n \times a$, L and R are two $a \times m$ and $m \times n$ matrices that are used to approximate Z, $m << n$, $\|.\|_{\mathrm{F}}$ is the Frobenius norm, E is a sparse noise matrix, $\|E\|_{2,1} = \sum_{i=1}^{n} \sqrt{\sum_{j=1}^{n} ([E]_{ij})^2}$ is the $l_{2,1}$-norm, λ_1 and λ_2 are trade-off parameters to balance the effects of other terms, S is a binary balanced graph and b is the constraint on the degree.

In Eq. (2), the first two terms $\|Z - LR\|_F^2 + \lambda_1 \|E\|_{2,1}$ represents the improved version of low-rank representation. Not only does it reduce the computational cost, but also addresses the small sample size problem [Liu *et al.*, 2012a]. The last term $\sum_{i,j=1}^{n} S_{ij}(Z^T Z)_{ij}$ and constraint $\sum_{j=1}^{n} S_{ij} = b$ indicate that each node should has a degree of b and meanwhile the total similarity value should be maximized.

Eq. (2) can be rewritten as:

$$\begin{aligned}&\min_{Z,L,R,E,S} \|Z - LR\|_F^2 + \lambda_1 \|E\|_{2,1} - \lambda_2 \mathbf{1}_n^T (S \circ (Z^T Z)) \mathbf{1}_n \\ &\text{s.t. } X = AZ + E, \mathbf{1}_a^T Z = \mathbf{1}_n^T, \sum_j S_{ij} = b, S_{ij} = S_{ji}.\end{aligned} \tag{3}$$

where $\circ$ is the Hadamard product, i.e., $(A \circ B)_{ij} = A_{ij} B_{ij}$.

3.2 Optimization

The last term in Eq. (3), $\mathbf{1}_n^T (S \circ (Z^T Z)) \mathbf{1}_n$, makes it difficult to solve the objective function. Fortunately, we can relax it to $\mathbf{1}_n^T (S \circ (Z^T J)) \mathbf{1}_n$ with a new constraint $J = Z$. Then we can rewrite (3) as:

$$\begin{aligned}&\min_{Z,L,R,E,S,J} \|Z - LR\|_F^2 - \lambda_2 \mathbf{1}_n^T (S \circ (Z^T J)) \mathbf{1}_n \\ &\qquad +\lambda_1 \|E\|_{2,1} \\ &\text{s.t.} \quad X = AZ + E, \mathbf{1}_a^T Z = \mathbf{1}_n^T, \sum_j S_{ij} = b, \\ &\qquad S_{ij} = S_{ji}, Z = J.\end{aligned} \tag{4}$$

To solve Eq. (4), we introduce several Lagrange multipliers and obtain the augmented Lagrange function:

$$\begin{aligned}L = &\|Z - LR\|_F^2 + \lambda_1 \|E\|_{2,1} - \lambda_2 \mathbf{1}_n^T (S \circ (Z^T J)) \mathbf{1}_n \\ &+ < X - AZ - E, Y_1 > + < \mathbf{1}_a^T Z - \mathbf{1}_n^T, Y_2 > \\ &+ < J - Z, Y_3 > + < \sum_j^n S_{ij} - b, Y_4 > \\ &+ \tfrac{\mu}{2} (\|X - AZ - E\|_F^2 + \left\|\mathbf{1}_a^T Z - \mathbf{1}_n^T\right\|_F^2 + \|J - Z\|_F^2)\end{aligned} \tag{5}$$

where Y_1, Y_2, Y_3 and Y_4 are Lagrange multipliers and $\mu > 0$ is a penalty parameter.

Although this problem is not jointly convex for all variables, it is convex for each variable when others are fixed. This equation can now be solved using the inexact ALM algorithm [Lin *et al.*, 2011]. In particular, we alternately update the variables L, R, Z, J, S and E in the $k+1$ iteration as follows:

$$L_{k+1} = Q, \tag{6}$$

$$R_{k+1} = Q^T Z_k, \tag{7}$$

where Q is the QR decomposition on $Z_k R_k^T$.

$$Z_{k+1} = (2\mathrm{I}_n + \mu(A^T A + \mathbf{1}_a^T \mathbf{1}_a + \mathrm{I}_n))^{-1} M_k, \tag{8}$$

where $M_k = 2L_{k+1}R_{k+1} + A^T Y_1 - \mathbf{1}_a Y_2 + \lambda_2 S_k \circ J_k + Y_3 + \mu_k(A^T X - A^T E_k + \mathbf{1}_a \mathbf{1}_n^T + J_k)$.

$$J_{k+1} = \frac{\lambda_2 S_k \circ Z_{k+1} - Y_3}{\mu_k} + Z_{k+1}. \quad (9)$$

$$S_{k+1} = \arg\min_{S_k} -\frac{\lambda_2}{\mu_k}\mathbf{1}_n^T(S_k \circ (Z_{k+1}^T J_{k+1}))\mathbf{1}_n. \quad (10)$$
$$s.t. \sum_j^n (S_k)_{ij} = b, (S_k)_{ij} = (S_k)_{ji}$$

$$E_{k+1} = \arg\min_{E_k} \frac{\lambda_1}{\mu_k}\|E_k\|_{2,1} + \frac{1}{2}\|C_k - E_k\|_{\mathrm{F}}^2, \quad (11)$$

where $C_k = X - AZ_{k+1} + Y_1/\mu_k$.

Eq.(10) can be solved by a fast b-matching algorithm that is recently proposed in [Huang and Jebara, 2011], and the solution to Eq.(11) is presented in [Liu *et al.*, 2010]. The details of the algorithm is outlined in **Algorithm 1**. A similar analysis to show the convergence property of Algorithm 1 can be found in [Lin *et al.*, 2011].

Algorithm 1. *Solving Problem* (5) *by Inexact ALM*

Input: data matrix X, dictionary A, parameter λ_1, λ_2, $L_0 \in \mathbb{R}^{a\times m}$, $R_0 \in \mathbb{R}^{m\times n}$, $Z_0 \in \mathbb{R}^{a\times n}$, $J_0 \in \mathbb{R}^{a\times n}$, $E_0 \in \mathbb{R}^{d\times n}$,$S_0 \in \mathbb{R}^{n\times n}$, $Y_1 \in \mathbb{R}^{d\times n}$, $Y_2 \in \mathbb{R}^{1\times n}$, $Y_3 \in \mathbb{R}^{a\times n}$,$\mu_0 = 10^{-3}$, $\mu_{\max} = 10^5$, $\rho = 1.3$, $k = 1$, $\epsilon = 10^{-5}$

Output: S_k, Z_k, L_k, R_k, E_k

1:**while** *not converged* **do**
2: update L_{k+1} using (6), given others fixed
3: update R_{k+1} using (7), given others fixed
4: update Z_{k+1} using (8), given others fixed
5: update J_{k+1} using (9), given others fixed
6: update S_{k+1} using (10), given others fixed
7: update E_{k+1} using (11), given others fixed
8: update the multipliers Y_1, Y_2 and Y_3
$Y_1 = Y_1 + \mu_k(X - AZ_{k+1} - E_{k+1})$,
$Y_2 = Y_2 + \mu_k(\mathbf{1}_n^T Z_{k+1} - \mathbf{1}_n^T)$,
$Y_3 = Y_3 + \mu_k(J_{k+1} - Z_{k+1})$.
9: update the parameter μ_{k+1} by $\mu_{k+1} = \min(\rho\mu_k, \mu_{\max})$
10: check the convergence conditions $\|X - AZ_{k+1} - E_{k+1}\|_\infty < \epsilon$ and $\|\mathbf{1}_a^T Z_{k+1} - \mathbf{1}_n^T\|_\infty < \epsilon$ and $\|J_{k+1} - Z_{k+1}\|_\infty < \epsilon$.
11: $k = k + 1$
12: **end while**

4 Graph Re-Weighting for Semi-supervised Classification

Given a sample set $X = [X_{\mathrm{l}}, X_{\mathrm{u}}]$, we can learn a regular and binary graph S and low-rank representation coefficients Z using Algorithm 1. As we discussed above, $Z^T Z$ is a good choice for measuring the similarity between each pair of samples. Thus, we can also re-weight the graph S using $Z^T Z$, and finally obtain a sparse, symmetric, weighted and balanced graph W, $W = S \circ (Z^T Z)$, which is named low-rank coding based balanced (LRCB) graph.

Our LRCB graph W can be easily combined with the representative label propagation method, Gaussian harmonic function (GHF) [Zhu *et al.*, 2003]. We denote $Y = [Y_{\mathrm{l}}, Y_{\mathrm{u}}]$, where Y_{l} contains the probability vectors for the labeled samples and Y_{u} for the unlabeled samples.

The predicted probability vectors can then be obtained for unlabeled samples by:

$$Y_{\mathrm{u}} = -Y_{\mathrm{l}} L_{\mathrm{lu}} L_{\mathrm{uu}}^{-1}, \quad (12)$$

where L is the Laplacian matrix, $L = D - W$, and D is the diagonal matrix whose diagonal elements are the sum of corresponding columns.

These steps are summarized in **Algorithm 2**. Same as [Liu *et al.*, 2010; 2012a], we use sample matrix, X, as the dictionary. Besides GHF, our LRCB graph can also be combined with other label propagation schemes such as local and global consistency (LGC) [Zhou *et al.*, 2003].

Algorithm 2. *LRCB Graph for Semi-supervised Learning*

Input: data matrix $X = [X_{\mathrm{l}}, X_{\mathrm{u}}] = [x_1, x_2, \ldots, x_n]$, dictionary $A = X$, parameters λ_1 and λ_2

Output: Y_{u}

1. Normalize all the samples x_i to unit-norm, $x_i = x_i / \|x_i\|$.
2. Solve problem (5) by using **Algorithm 1** and obtain optimal solution S and Z.
3. Graph re-weighting: $W = S \circ (Z^T Z)$.
4. Calculate probability vectors Y_{u} for unlabeled samples X_{u} using (12).

5 Experiments

In this section, we evaluate the performance of our graph and related graphs in graph based semi-supervised learning.

5.1 Databases and Settings

In our experiments four image databases are used: Extended YaleB [Georghiades *et al.*, 2001], PIE [Sim *et al.*, 2003], ORL [Samaria and Harter, 1994] and USPS [Hull, 1994].

Extended YaleB Face Database. This database has 38 subjects and around 64 images under different illuminations per subject. We use the images of the first 15 subjects, and we crop and resize each image to the size of 32×32 pixels.

PIE Face Database. The PIE face database consists of 41368 images of 68 subjects. Each subject is shown in different poses, illuminations and expressions. We use the first 15 subjects and five near front poses (C05, C07, C09, C27, C29), and all the different illuminations and expressions. Each images is cropped and resized to the size of 32×32 pixels.

ORL Face Database. The ORL face database contains 400 images of 40 individuals. These images were captured at different times, varying the lighting and facial expressions. We crop and resize each image to 32×32 pixels.

USPS Digit Database. The USPS digit database consists of 9298 handwritten digit images of 10 numbers (0-9). The size of each image is 16×16 pixels.

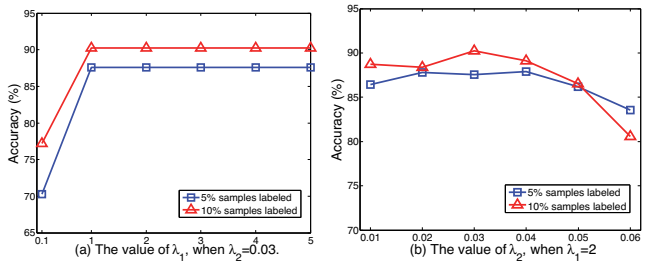


Figure 2: Accuracies of our graph under different values of (a) λ_1 and (b) λ_2 on extended YaleB face database.

In the experiments, we compare the following graphs obtained by different construction methods:

(1) **k-NN graph**. In this graph, two samples are connected if one is among the k nearest neighbors of the other. The k-NN-graph is conducted in two configurations. In **k-NN-I**, the number of nearest neighbors is set to 5; and in **k-NN-II**, this number is set to 8. We use the Gaussian kernel to re-weight the edges, and the parameter δ is adjusted to achieve the best performance on different databases.

(2) **b-matching (BM) graph**. We follow the algorithms described in [Jebara *et al.*, 2009], and construct a weighted and balanced graph. The parameter b is selected to achieve the best results.

(3) **l_1 graph**. We construct the l_1 graph according to [Yan and Wang, 2009], and also symmetrize this graph.

(4) **Sparse probability graph**. Following the algorithms in [He *et al.*, 2011], we construct the non-negative sparse probability graph (SPG).

(5) **LRR graph**. In accordance with [Liu *et al.*, 2010], we construct the LRR graph and symmetrize it. We adopt the same parameters as described in [Liu *et al.*, 2010].

(6) **NNLRS graph**. We construct the non-negative sparse and low-rank graph according to [Zhuang *et al.*, 2012]. We also symmetrize this graph. The parameters are set as described in [Zhuang *et al.*, 2012].

(7) **Our LRCB graph**. Two major parameters are λ_1 and λ_2. To choose proper values for them, we evaluate the parameter sensitivity of our graph. Fig. 2 shows the accuracies of our graph under different settings of λ_1 and λ_2, respectively. Here, λ_1 is to handle the corruptions or large noise in the samples, while λ_2 is used to balance low-rank approximation and b-matching constraint. Since the images in Extended YaleB database are captured in a relatively controlled environment, there's not much corruptions and our graph is not sensitive to λ_1 over a wide range. Fig. 2 also shows that our graph achieves similar results when λ_2 is varied from 0.01 to 0.04. On other three databases, we obtain similar results. Thus, for all the four databases, λ_1 and λ_2 are separately set as 2 and 0.03. Another parameter b is set as 5, 10, 5, 5 in the Extended YaleB, PIE, ORL and USPS databases, respectively.

5.2 Experimental Results

We first normalize all the images to be unit-norm as shown in **Algorithm 2**. We randomly select 50 images of every class in the PIE and YaleB databases, and 100 images from each class in the USPS database. We use all the images of ORL database. All experiments are repeated 10 times, and each time we randomly select a subset of images for each individual to create labeled sample set. Unlike most existing semi-supervised learning experiments, we test the performance of all compared methods with only a small set of labeled samples, because the goal of semi-supervised learning is to deal with practical tasks that have very limited labeled samples but a large amount of unlabeled ones. For each individual, the percentage of labeled samples on the Extend YaleB, PIE and USPS databases is varied from 5% to 30%. Since there are only 10 images of each individual in ORL database, this percentage varies from 10% to 50%.

Table 1 shows the average accuracies of different graphs combined with GHF label propagation strategy on four databases, when the percentages of labeled samples are 10%, 20% and 30%. Fig. 3 shows the average accuracies versus varying percentages of labeled samples. We can observe from Table 1 and Fig. 3 that:

(a) BM graph usually outperforms k-NN graph, since BM graph emphasises sparsity and balance at the same time;

(b) The advanced sparse representation based graph, SPG, outperforms k-NN graph in many cases on the Extended YaleB and PIE databases, and achieves comparable performance as BM graph;

(c) NNLRS graph, which is a low-rank and sparse graph, performs better than other existing graphs in most cases;

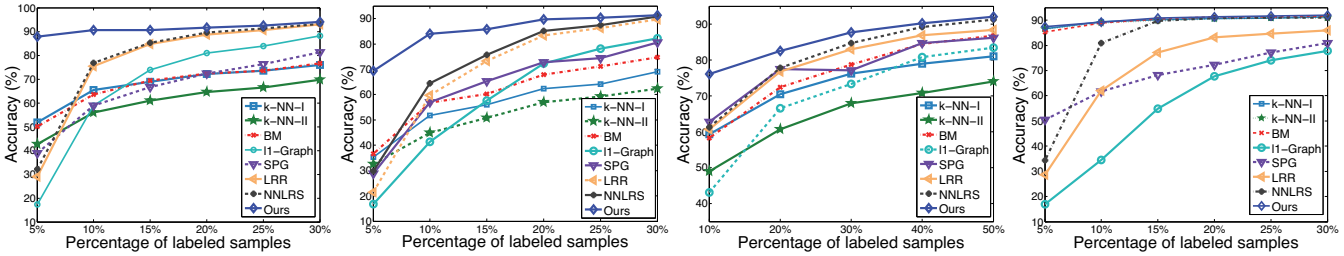
(d) When the percentage of labeled samples is increased, the performance of all compared methods is increased. Our LRCB graph outperforms other compared graphs in almost all cases, and it reaches great performance very quickly. When the labeled samples are very limited, e.g., under 10% of total number of samples, our graph can significantly improve the accuracy over the state-of-the-art graphs on three face databases.

5.3 Discussions

To illustrate why our graph outperforms other compared graphs, Fig. 4 visualizes several weighted graphs including k-NN-graph, LRR graph, NNLRS graph and our graph on the PIE face database. We can observe from Fig. 4 that k-NN graph is sparse, but it's not balanced. LRR produces a very dense graph that is undesirable for GSSL. NNLRS graph is a little sparser than LRR graph, however, why low-rank graph could directly represent the similarities between different samples is unclear. Also, both LRR and NNLRS graphs are not balanced. Fig. 4(d) shows that our graph is much sparser than the aforementioned three graphs. It correctly connects the samples within each class (diagonal blocks in the graph), and meanwhile the misconnections between samples in different classes are fewer than other graphs. These results validate the superiority of our low-rank encoding based similarity metric.

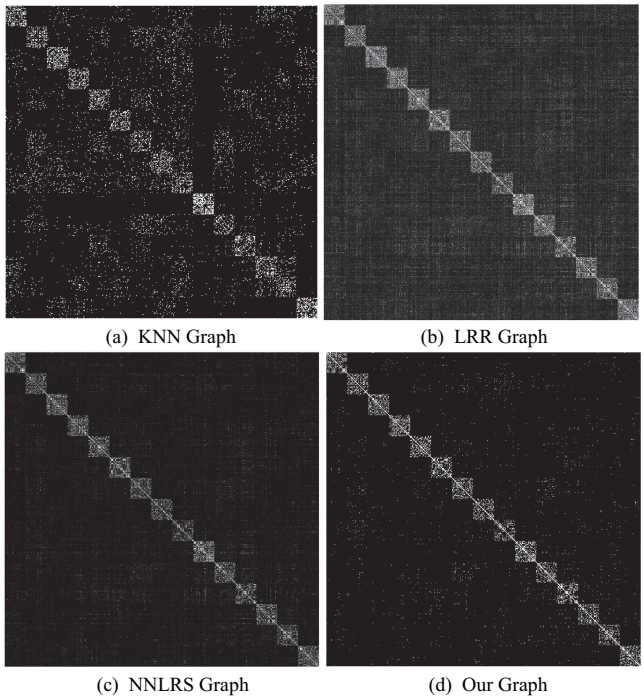
Table 1: Average accuracies of different graphs with standard deviations combined with the GHF label propagation method under different percentages of labeled samples (shown in the parenthesis) on four databases.

Methods	k-NN-I	k-NN-II	BM	l_1-graph	SPG	LRR	NNLRS	Ours
YaleB (10%)	65.41±1.81	56.03±2.27	63.45±2.23	58.81±13.72	58.90±2.18	75.48±4.02	76.89±3.54	**90.67**±0.93
YaleB (20%)	72.10±1.68	64.52±1.85	72.30±2.43	80.93±2.57	72.25±2.04	88.67±1.75	89.58±1.30	**91.61**±0.75
YaleB (30%)	75.92±1.64	69.81±2.36	76.61±2.74	88.11±2.38	81.22±2.18	92.95±1.55	93.20±1.49	**94.02**±0.99
PIE (10%)	51.69±2.69	44.93±3.88	56.84±2.86	41.41±11.79	57.04±3.28	59.67±6.51	64.43±5.12	**84.06**±2.06
PIE (20%)	62.30±1.71	56.90±2.48	67.83±2.39	72.33±5.88	72.83±2.69	83.28±3.56	85.17±2.75	**89.72**±1.91
PIE (30%)	68.99±1.68	62.38±1.75	74.67±1.95	82.25±2.28	80.57±1.98	89.71±1.92	90.88±1.62	**91.30**±1.64
ORL (10%)	59.33±1.44	48.94±2.19	58.33±2.01	43.06±2.74	62.78±3.02	60.69±2.59	61.27±2.76	**76.11**±2.41
ORL (20%)	70.41±2.43	60.69±3.43	72.40±1.69	66.56±3.93	77.50±2.69	76.78±1.91	77.81±2.94	**82.57**±2.23
ORL (30%)	76.21±1.76	67.89±2.98	78.79±2.55	73.36±2.25	77.14±2.18	83.04±2.59	84.75±2.59	**87.70**±1.85
USPS (10%)	**89.40**±0.92	88.88±1.47	88.92±0.91	34.43±7.47	61.64±0.93	62.09±9.91	80.86±5.64	89.16±0.73
USPS (20%)	90.65±0.84	90.51±0.82	91.30±0.92	67.65±5.54	72.35±1.36	83.19±1.82	90.85±2.71	**91.41**±0.79
USPS (30%)	91.31±0.66	90.90±0.72	91.63±0.78	77.79±3.42	80.91±1.30	85.91±1.54	91.01±1.71	**92.06**±0.48



(a) Extended YaleB (b) PIE (c) ORL (d) USPS

Figure 3: Accuracy of our approach and compared methods versus different percentages of labeled samples on four databases.



(a) KNN Graph (b) LRR Graph

(c) NNLRS Graph (d) Our Graph

Figure 4: Visualization of different weighted graphs on PIE face database.

Another consideration is the computational cost of our approach. Although solving low-rank coefficient matrix in **Algorithm 1** is very efficient, the b-matching optimization requires more computational cost than the basic k-NN graph. We will design more efficient implementations in future work. In addition, we simply use sample set X itself as dictionary in **Algorithms 1 and 2**. Learning informative dictionary is expected to further enhance the classification performance, which could be another direction in our future work.

6 Conclusions

In this paper, we propose a novel graph construction approach for semi-supervised learning. By taking advantages of low-rank encoding and b-matching constraint, we jointly learn a symmetric, sparse and balanced graph. GHF method is then used for semi-supervised classification. Experimental results on the Extended YaleB, PIE, ORL and USPS databases demonstrate the effectiveness of our approach compared with several state-of-the-art methods. Moreover, when the labeled samples are very limited (under 10% of total number of samples), our approach demonstrates significant improvement in classification performance.

Acknowledgments

The authors want to thank the anonymous reviewers for their helpful comments, and thank Dmitry Kit and Ming Shao for their suggestions to improve the paper.

References

[Bellhumeur *et al.*, 1997] Peter N. Bellhumeur, Joao P. Hespanha, , and David J. Kriegeman. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(7):711–720, 1997.

[Cai *et al.*, 2007] Deng Cai, Xiaofei He, and Jiawei Han. Semi-supervised discriminant analysis. In *ICCV*, 2007.

[Candès *et al.*, 2011] Emmanuel J. Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *J. ACM*, 58(3):11, 2011.

[Chapelle *et al.*, 2006] Olivier Chapelle, Bernhard Schlkopf, and Alexander Zien. *Semi-supervised Learning*. MIT Press, Cambridge, Massachusetts, 2006.

[Chen and Wang, 2007] Ke Chen and Shihai Wang. Regularized boost for semi-supervised learning. In *NIPS*, 2007.

[Culp and Michailidis, 2008] Mark Culp and George Michailidis. Graph-based semisupervised learning. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(1):174–179, 2008.

[Georghiades *et al.*, 2001] Athinodoros S. Georghiades, Peter N. Belhumeur, and David J. Kriegman. 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.

[He *et al.*, 2007] Jingrui He, Jaime G. Carbonell, and Yan Liu. Graph-based semi-supervised learning as a generative model. In *IJCAI*, pages 2492–2497, 2007.

[He *et al.*, 2011] Ran He, Weishi Zheng, Baogang Hu, and Xiangwei Kong. Nonnegative sparse coding for discriminative semi-supervised learning. In *CVPR*, pages 2849–2856, 2011.

[Huang and Jebara, 2011] Bert C. Huang and Tony Jebara. Fast b-matching via sufficient selection belief propagation. *Journal of Machine Learning Research - Proceedings Track*, pages 361–369, 2011.

[Hull, 1994] Jonathan J. Hull. A database for handwritten text recognition research. *IEEE Trans. Pattern Anal. Mach. Intell.*, 16(5):550–554, 1994.

[Jebara *et al.*, 2009] Tony Jebara, Jun Wang, and Shih-Fu Chang. Graph construction and *b*-matching for semi-supervised learning. In *ICML*, 2009.

[Kveton *et al.*, 2010] Branislav Kveton, Michal Valko, Ali Rahimi, and Ling Huang. Semi-supervised learning with max-margin graph cuts. *Journal of Machine Learning Research - Proceedings Track*, 9:421–428, 2010.

[Li *et al.*, 2013] Liangyue Li, Sheng Li, and Yun Fu. Discriminative dictionary learning with low-rank regularization for face recognition. In *FG*, 2013.

[Lin *et al.*, 2011] Zhouchen Lin, Risheng Liu, and Zhixun Su. Linearized alternating direction method with adaptive penalty for low-rank representation. In *NIPS*, pages 612–620, 2011.

[Liu *et al.*, 2010] Guangcan Liu, Zhouchen Lin, and Yong Yu. Robust subspace segmentation by low-rank representation. In *ICML*, pages 663–670, 2010.

[Liu *et al.*, 2012a] Risheng Liu, Zhouchen Lin, Fernando. De Torre, and Zhixun Su. Fixed-rank representation for unsupervised visual learning. In *CVPR*, pages 598–605, 2012.

[Liu *et al.*, 2012b] Wei Liu, Jun Wang, and Shih-Fu Chang. Robust and scalable graph-based semisupervised learning. *Proceedings of the IEEE*, 100(9):2624–2638, 2012.

[Liu *et al.*, 2013] Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yong Yu, and Yi Ma. Robust recovery of subspace structures by low-rank representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, pages 171–184, 2013.

[Maier *et al.*, 2008] Markus Maier, Ulrike von Luxburg, and Matthias Hein. Influence of graph construction on graph-based clustering measures. In *NIPS*, pages 1025–1032, 2008.

[Samaria and Harter, 1994] Ferdinando Samaria and Andy Harter. Parameterisation of a stochastic model for human face identification. In *Proceedings of the Second IEEE Workshop on Applications of Computer Vision*, pages 138–142, 1994.

[Sim *et al.*, 2003] Terence Sim, Simon Baker, and Maan Bsat. The cmu pose, illumination, and expression database. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(12):1615–1618, 2003.

[Turk and Pentland, 1991] Matthew A. Turk and Alex P. Pentland. Eigenfaces for face recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[Wang and Xia, 2012] Jun Wang and Yinglong Xia. Fast graph construction using auction algorithm. In *UAI*, pages 873–882, 2012.

[Wright *et al.*, 2009] John Wright, Allen Y. Yang, Arvind Ganesh, S.Shankar Sastry, and Yi Ma. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(2):210–227, 2009.

[Yan and Wang, 2009] Shuicheng Yan and Huan Wang. Semi-supervised learning by sparse representation. In *SIAM International Conference on Data Mining (SDM)*, pages 792–801, 2009.

[Zhou *et al.*, 2003] Dengyong Zhou, Olivier Bousquet, Thomas Navin Lal, Jason Weston, and Bernhard Schölkopf. Learning with local and global consistency. In *NIPS*, 2003.

[Zhu *et al.*, 2003] Xiaojin Zhu, Zoubin Ghahramani, and John D. Lafferty. Semi-supervised learning using gaussian fields and harmonic functions. In *ICML*, pages 912–919, 2003.

[Zhu, 2010] Xiaojin Zhu. Semi-supervised learning. *Encyclopedia of Machine Learning*, pages 892–897, 2010.

[Zhuang *et al.*, 2012] Liansheng Zhuang, Haoyuan Gao, Zhouchen Lin, Yi Ma, Xin Zhang, and Nenghai Yu. Non-negative low rank and sparse graph for semi-supervised learning. In *CVPR*, pages 2328–2335, 2012.