J Med Syst
DOI 10.1007/s10916-011-9754-6

ORIGINAL PAPER

Breast Tissue Image Classification Based on Semi-supervised Locality Discriminant Projection with Kernels

Jun-Bao Li · Yang Yu · Zhi-Ming Yang · Lin-Lin Tang

Received: 21 March 2011 / Accepted: 28 June 2011

Abstract Breast tissue classification is an important and effective way for computer aided diagnosis of breast cancer. We present Semi-supervised Locality Discriminant Projections with Kernels for breast cancer classification. The contributions of this work lie in: 1) Semi-supervised learning is used into Locality Preserving Projections (LPP) to enhance its performance using side-information together with the unlabelled training samples, while current algorithms only consider the side-information but ignoring the unlabeled training samples. 2) Kernel trick is applied into Semi-supervised LPP to improve its ability in the nonlinear classification. 3) The framework of breast cancer classification with Semi-supervised LPP with kernels is presented. Many experiments are implemented on four breast tissue databases to testify and evaluate the feasibility and affectivity of the proposed scheme.

Keywords Breast cancer classification · Locality preserving projections · Semi-supervised learning · Kernel learning

J.-B. Li (✉) · Y. Yu · Z.-M. Yang
Department of Automatic Test and Control,
Harbin Institute of Technology,
P. O. Box 3033, Room 426, Building 2A, HIT Sciences Park,
No. 2 Yi-Kuang Street, Nan-Gang District,
Harbin 150080, China
e-mail: junbaolihit@gmail.com

L.-L. Tang
School of Computer Science and Technology,
Harbin Institute of Technology Shenzhen Graduate School,
Shenzhen, China
e-mail: lltang@hitsz.edu.cn

Introduction

As definition of breast cancer from National Cancer Institute, U.S. National Institute of Health, Cancer that forms in tissues of the breast, usually the ducts (tubes that carry milk to the nipple) and lobules (glands that make milk). And breast cancer occurs in both men and women, although male breast cancer is rare. According to the statistical data from American National Cancer Institute, in 2004, the estimated new breast cancer cases was near 215000, but the deaths was estimated more than 40000 [1]. In 2010, the estimated cases and deaths from breast cancer in U. S. are 207,090 (female) and 1,970 (male) respectively, but the population of deaths are 39,840 (female) and 390 (male) respectively. In addition, National Cancer Institute of U.S. estimates that 13.4 percent of women born today will be diagnosed with breast cancer in future [2]. For the diagnosis and prognosis of the breast cancer cases, many techniques have been discussed in the previous work [3–6, 42]. The breast cancer data analysis contains the following three cases: one is to make the decision the health status from the breast tissue data, second is to make the diagnosis to provide a distinction between the malignant and benign breast masses, third is prognostic prediction where the patient is classified as a "recur" case or "non-recur" case.

Machine learning is a feasible method for breast cancer classification based computer aided diagnosis. In the past research, many methods are presented including Principal Component Analysis (PCA) [7] and Linear Discriminant Analysis (LDA) [8]. PCA seeks a linear optimal transformation matrix to minimize the mean squared error criterion, and the optimal matrix is constituted by the largest eigenvectors (called principal components) of the sample covariance. Unlike PCA which has little done with the class information, LDA seeks the optimal projection matrix based on Fisher criterion and the class labels are taken into account for dimensionality reduction. Moreover, PCA is generalized to form the nonlinear curves such as principal curves [9] and principal surfaces [10]. Principal curves and principal surfaces are the nonlinear generalizations of principal components and subspaces respectively. It has turned out that discretized principal curves are essentially equivalent to self-organizing maps (SOM) [11, 12]. Thereof the conception of principal curves provides a new possible viewpoint to the properties of SOM. SOM is a nonparametric latent variable model with a topological constraint, such as lines, squares, or hexagonal grids and its mapping is similar to a discrete self-similarity principle for a principal manifold. SOM is a data driven method to reduce dimensionality and the reduced modes are not always interpretable. SOM serves of an approximation to the principal surface, which converges to it for a large enough number of nodes. As an extended SOM, the visualisation-induced SOM (ViSOM) algorithm directly preserves the distance information on the map along with the topology [13]. ViSOM represents a discrete principal curve or surface and produces a smooth and graded mesh in the data space and captures the nonlinear manifold of data [14]. Other nonlinear manifold algorithms, such as Isomap [15] and Locally Linear Embedding (LLE) [16], were proposed in recent years. The biggest difference between Isomap and LLE lies in the way of dimensionality reduction and data representation. Isomap regards dimensionality reduction problem as a graph problem where data are represented as connected graphs and relationship among data is described through geodesic distances, while LLE views dimensionality reduction as purely geometrical perspective. However, LLE and Isomap are both concerned with preserving the neighborhoods and their geometric and graph relationships respectively. Moreover, Isomap and LLE were well studied on map of the training data but rarely on the test data, while Locality Preserving Projection [17] may be simply applied to any new data point and locate it in the reduced representation space. All these algorithms are based supervised learning or unsupervised learning, while in many applications only few of samples are labeled with the class information or side-information. In these cases, supervised and semi-supervised learning methods do not perform well on classification task. Based on this background, side-information based semi-supervised learning is the popular research topic in machine learning and pattern area, which only knows whether two samples come from the same classes or the different classes but without knowing the class label information. Many algorithms are proposed in the previous work. Shental correlation component analysis based positive constrain information [18], and Bar-Hiller proposed the constrain-based Fisher discriminant analysis [19]. Above two algorithms only use the positive constraint information but ignoring the unlabeled data. Xing [20] and Tang [21], use the positive and negative constraint information but ignoring the unlabeled data. Zhang [22] proposed a semi-supervised dimensionality reduction method through considering the constraint structure and low-dimensional manifold structure, but this method does not consider the local structure. Locality Preserving Projection and its improved methods are used in many areas, such as object recognition [25, 29], face recognition [26–28, 41]. With the applying of the kernel methods in many areas [23, 24], other researchers improved LPP with kernels in the previous works [30–32].

A novel breast cancer classification method based on Semi-supervised Locality Discriminant Projections with Kernels is presented. Semi-supervised learning is used into Locality Preserving Projections (LPP) with side-information together with the unlabelled training samples, and we use Kernel trick to improve the ability of semi-supervised LPP in the nonlinear classification. We propose one framework of breast cancer classification with Semi-supervised Locality Discriminant Projection (SLDP) with kernels. We test the feasibility of the framework on popular four breast tissue databases, and evaluate the classification performance compared with other popular classification methods.

Semi-supervised locality discriminant projection with kernels

In this section, we review the Locality Preserving Projection (LPP) algorithm firstly, and then we extends it to Semi-supervised Locality Discriminant Projections (SLDP) with side-information and the unlabelled sample data, finally we enhance SLPP on the nonlinear feature extraction with kernels.

Review of LPP

In this section, we introduce Locality Preserving Projection (LPP) [17], which projects high-dimensional input data into a low-dimensional subspace through preserving the local structure. The definition of LPP is defined as follows.

$$\min \sum_{i,j} \left\| z_i - z_j \right\|^2 S_{ij}, i,j = 1,2,\ldots,n \tag{1}$$

where S is a weight similarity matrix the similarity of two points, and $z_i = w^T x_i$ is the one-dimensional representation of x with a projection vector w. Given the original input data, dimensionality reduction is to find a transformation matrix W to project these input data into the low-dimensional data with one optimal projection criterion. Through minimizing the objective function, the local information of the original space will be preserved under the dimension reduction. The local structure of data in the original space is preserved in the lower dimension space. The definition of LPP is written as

$$\sum_{i,j} \left\| z_i - z_j \right\|^2 S_{ij} = w^T X L X^T w \tag{2}$$

where $X = \{x_1, x_2, \ldots, x_n\}$, $L = D\text{-}S$ is Laplacian matrix, and D is a diagonal matrix with its entries being the column or row (S is symmetric) sums of S. D is the local structure of the data in the original space. With the optimization constraint $w^T XDX^T w = 1$, then

$$\begin{aligned} &\arg\max_w w^T XLX^T w \\ &Subject\ to\ w^T XDX^T w = 1 \end{aligned} \tag{3}$$

The transformation vector w minimizes the objective function through solving the following eigenvalue problem. LPP is a very important approach to extract the feature vector with dimensionality reduction, but it suffers small sample size (SSS) problem, i.e., XDX^T is singular. In order to overcome the singularity of XDX^T, LPP employs PCA to reduce the dimensionality. Firstly, the input vector is transformed to a low-dimensional PCA-transformed vector. Secondly, the nearest-neighbor graph and the similarity matrix S is calculated with the PCA-transformed vectors with nearest-neighbor criterion. Finally, the projection matrix is obtained through solving eigenvalue problem.

Semi-supervised locality discriminant projection

As discussion in "Review of LPP", both PCA and LPP are unsupervised learning methods, LDA is supervised learning method. One of the differences between PCA and LPP lies in the global or local preserving property, that is, PCA seeks to preserve the global property while LPP preserves the local structure. The locality preserving property leads to the fact that LPP outperforms PCA. Also as the global method, LDA utilizes the class information to enhance its discriminant ability which causes LDA to outperform PCA on classification. But the objective function of LPP is to minimize the local quantity, i.e., the local scatter of the projected data. This criterion cannot be guaranteed to yield a good projection for classification purposes. So it is reasonable to enhance LPP on classification using the class information like LDA. In many applications, we must use the semi-supervised learning method to training the samples. So, we proposed semi-supervised locality preserving projection method based graph-based viewpoint including intrinsic graph and cost graph. The intrinsic graph demonstrates the relation of two samples belonging to the same classes, and the cost graph demonstrates the different classes of samples which denotes by the constraint or the non k nearest neighbor. Firstly, we apply the side-information to construct the intrinsic graph and cost graph, and secondly in order to use enough the unlabeled samples we also apply the nearest neighbor criterion to construct the intrinsic graph based k nearest neighbor and the cost based non k nearest neighbor. Finally, we construct the constraint equation through combining the graph of side-information and k nearest neighbor with the weight value.

Firstly, we consider the side-information of samples for training and develop the SLDP-1 (SLDP with side-information training samples). The intrinsic graph is created by positive constraint, and the cost graph is constructed with the negative constraint. We expect to shorten the distance of data points in intrinsic graph and lengthen the distance between the data points in the cost graph. The mathematic expression is shown as follows.

Supposed that P is the positive constraint and N is the negative constraint, P denotes that two samples belong to the same class but the class labels are not known, and N denotes that two different classes of samples. From P and N, we can not know the class labels of samples.

Within-class compactness of samples M is defined as

$$\begin{aligned} M &= \sum_{(x_i,x_j)\in P} \left(w^T x_i - w^T x_j\right)^2 \\ &= 2\sum_i \left(w^T x_i D^P_{ii} x_i w\right) - 2\sum_{ij} \left(w^T x_i S^P_{ij} x_j w\right) \\ &= 2w^T X\left(D^P - S^P\right)X^T w \\ &= 2w^T XL^P X^T w \end{aligned} \tag{4}$$

where $S^P = \begin{cases} 1 & (x_i, x_j) \in P \\ 0 & else \end{cases}$, $D^p_{ii} = \sum_j S^P_{ij}$, and $L^P = D^P - S^P$.

Between-class compactness of samples is defined as

$$\begin{aligned} B &= \sum_{(x_i,x_j)\in N} \left(w^T x_i - w^T x_j\right)^2 \\ &= 2\sum_i \left(w^T x_i D^N_{ii} x_i w\right) - 2\sum_{ij} \left(w^T x_i S^N_{ij} x_j w\right) \\ &= 2w^T X\left(D^N - S^N\right)X^T w \\ &= 2w^T XL^N X^T w \end{aligned} \tag{5}$$

where $S^N = \begin{cases} 1 & (x_i, x_j) \in N \\ 0 & else \end{cases}$, $D^N_{ii} = \sum_j S^N_{ij}$, and $L^N = D^N - S^N$.

The objective function is defined as

$$w^* = \arg\max_w \frac{B}{M} = \arg\max_w \frac{w^T XL^P X^T w}{w^T XL^N X^T w} \tag{6}$$

The above equation aims to maximize the distance between the samples belongs to the different classes but minimize the distance within the samples from the same classes. This equation only consider the side-information but without using the non-labeled data. In order to make enough use of a large of unlabeled samples, we regard the samples belonging to the k nearest neighbor as the positive constraint and the negative constraint if non k nearest neighbor. SLDP-2 (SLDP with both side-information and unlabeled training samples) algorithm is proposed. Then supposed that the samples will be closed in the low-dimensional projection space where its samples are closed in the high-dimensional input space. The improved objective function is defined as

$$w^* = \arg\max_w \frac{B + aB_{knn}}{M + bM_{knn}} = \arg\max_w \frac{w^T X L^P X^T w + a w^T X L^P_{knn} X^T w}{w^T X L^N X^T w + b w^T X L^N_{knn} X^T w} \tag{7}$$

where $B_{knn} = w^T X L^P_{knn} X^T w$, $M_{knn} = w^T X L^N_{knn} X^T w$. The L^P_{knn} and L^N_{knn} are calculated with the positive and negative graph created by the k nearest neighbor criterion. The samples belongs to the k nearest neighbors each other. a and b are weight parameter.

SLDP algorithm is shown as follows.

Input: the samples set $X = \{x_1, x_2, \ldots, x_n\} \in R^{D\times n}$ with the positive constraint P and the negative constraint N.
Output: the transformed feature vector y with projection matrix $W \in R^{D\times d} (d \prec D)$.

Step 1. Implement PCA to reduce the dimension of the samples.
Step 2. Construct the positive and negative graph.
Step 3. Determine the weight parameter α and β, and solve the transformation vector w^*.
Step 4. Calculate the linear transformation matrix. Supposed the linear transformation matrix $W = \{w_1, w_2, \ldots, w_d\}$, where $w_1, w_2, \ldots, w_d$ are the eigenvectors corresponding to the d largest eigenvalues.
Step 5. Reduce the input vector x to the low-dimensional vector y with $y = W^T x$.

Kernel based SLDP (K-SLDP)

Above algorithm is based on the linear projection, and we apply the nonlinear mapping to embedding the nonlinear mapping $\phi(x)$ is used to map the input data space R into the feature space F. Correspondingly, a pattern in the original input space R is mapped into a potentially much higher dimensional feature vector in the feature space F with $z = w^T_{\ker}\phi(x)$. Kernel SLPP aims to seek a set of data points $\{z_1, z_2, \ldots, z_n\}$ with the same local neighborhood structure as $\{\Phi(x_1), \Phi(x_2), \ldots, \Phi(x_n)\}$ in the nonlinear mapping space. The transformation matrix $w_{\ker}$ is that minimizes the objective function can be obtained by solving a generalized eigenvalue problem. According to the definition of kernel method, $k(x_i, x_j) = \phi(x_i)^T \phi(x_j)$, we can have

$$w^*_{\ker} = \arg\max_w \frac{w^T K L^P K w + a w^T K L^P_{knn} K w}{w^T K L^N K w + b w^T K L^N_{knn} K w} \tag{8}$$

With this method, the nonlinear classification is realized based on the Semi-supervised Locality Dicriminant Projection (SLDP). The procedure of Kernel SLDP (K-SLDP) is divided into two stages, KPCA+SLDP. K-SLDP algorithm procedure is shown as follows.

Input: the samples set $X = \{x_1, x_2, \ldots, x_n\} \in R^{D\times n}$ with the positive constraint P and the negative constraint N.
Output: the nonlinear transformation vector y.

Step 1. Implement KPCA to reduce the dimension of the samples, and transform the input vector x into the KPCA-transformed vector z.
Step 2. Construct the positive and negative graph.
Step 3. Determine the weight parameter α and β, and solve the transformation vector w^*.
Step 4. Calculate the linear transformation matrix. Supposed the linear transformation matrix $W = \{w_1, w_2, \ldots, w_d\}$, where $w_1, w_2, \ldots, w_d$ are eigenvectors corresponding to d largest eigenvalues.
Step 5. Reduce the input vector to the low-dimensional vector y with $y = W^T z$.

From the above algorithm procedure, two problems should be emphasized including computing efficiency and the practical applications ability. On the computation efficiency, constructing the positive and negative graph is the main time consuming procedure. The computation time is increasing with the total number of training samples. The dimension of features also is other influence factors for the computing efficiency of construct positive and negative graph, but this influence is not so heavy compared with positive and negative graph construction. Then on the practical applications, it should be emphasize the ability of the construction of positive/negative samples from unlabeled data with the k-nearest neighbor criteria. Because the construction of positive and negative graph is implemented with off-line mode, this method is applicable to the general pattern recognition. The high time consuming of constructing the positive and negative graph has not so heave influence on the on-line application of the proposed algorithm in the practical applications.

Experimental results

In this section, we implement four sets of experiments to testify and evaluate the algorithm compared with the other

Table 1 Class description and number of the database

Class No.	Class decription	Class number
Class 1	Carcinoma	21
Class 2	Fibro-adenoma	15
Class 3	Mastopathy	18
Class 4	Glandular	16
Class 5	Connective	14
Class 6	Adipose	22

popular algorithms on breast tissue classification. For semi-supervised learning, we construct the positive and negative constraints with half number of these training samples according to the labeled class information, and the rest training samples are considered the unlabeled training samples. After training the classifier, we implement the test experiments on the test samples.

Firstly, on the parameters of experimental procedure are selected or set. On kernel function select kernel and its parameters, in current research all most research papers apply the cross-validation method to choose the kernel function including Gaussian kernel and other kernels. The parameters of the kernel function are selected with the same methods. On the dimensionality of PCA and KPCA, in the practical applications, the detail number values in the following experiments on four databases are listed as follows. In the first set of data classification on breast tissue dataset, the dimensionality of KPCA and PCA is 8, and 8-dimensional feature of PCA and KPCA is chosen on Wisconsin Breast Cancer (WBC) Database. The dimensionalities of PCA and KPCA on Wisconsin Diagnostic Breast Cancer (WDBC) dataset and on Wisconsin Prognostic Breast Cancer (WPBC) dataset are 31 and 8. Moreover, in practical applications, we use the signal processing method to de-noise and other relative preprocessing. Our method pays the research emphasis on classification and feature extraction. On the preprocessing procedure, such as denoise, we apply the current method to deal with it.

Table 2 Description of each feature vector

Elements of feature vector	Element description	Detail description
V[1]	I0	Impedivity (ohm) at zero frequency
V[2]	PA500	Phase angle at 500 KHz
V[3]	HFS	High-frequency slope of phase angle
V[4]	DA	Impedance distance between spectral ends
V[5]	AREA	Area under spectrum
V[6]	A/DA	Area normalized by DA
V[7]	MAX IP	Maximum of the spectrum
V[8]	DR	Distance between I0 and real part of the maximum frequency point
V[9]	P	Length of the spectral curve

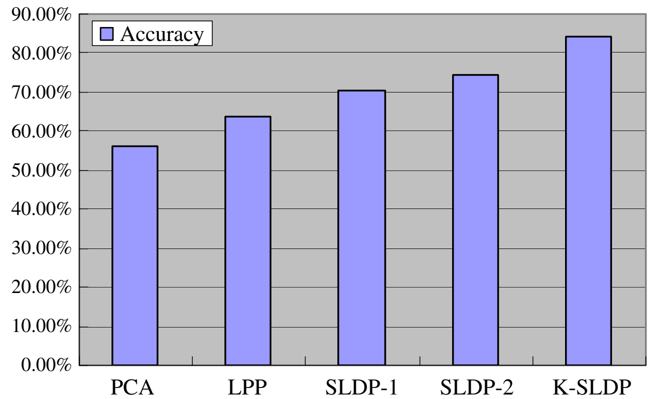

Fig. 1 Data classification performance on Breast Tissue Dataset

Data classification on breast tissue dataset

There are 106 samples on breast tissue dataset with electrical impedance measurements of freshly excised tissue samples from the breast [33, 34]. 6 classes of samples are contained in the database. In the sampled dataset, impedance measurements of freshly excised breast tissue were made at the following frequencies: 15.625, 31.25, 62.5, 125, 250, 500, 1000 KHz. In the experiments, the dataset is used for predicting the classification, and the above measurements plotted in the plane constitute the impedance spectrum from where the breast tissue features are computed. In the above experiments, there are 6 classes, each class description and number of each class of samples is described in Table 1. In the experiments, we randomly choose half number of training samples from the databases

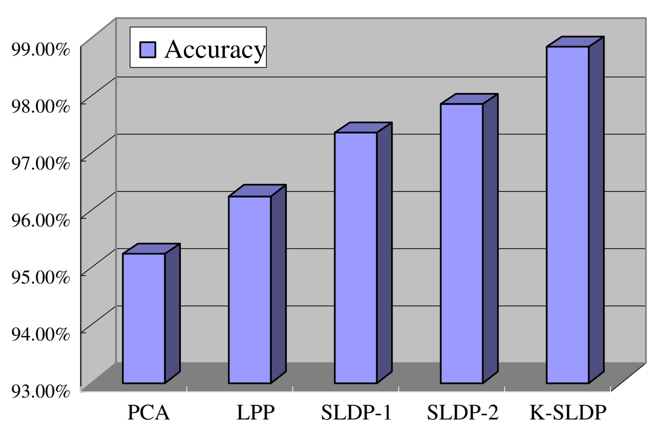

Fig. 2 Data classification performance on Breast Tissue Dataset

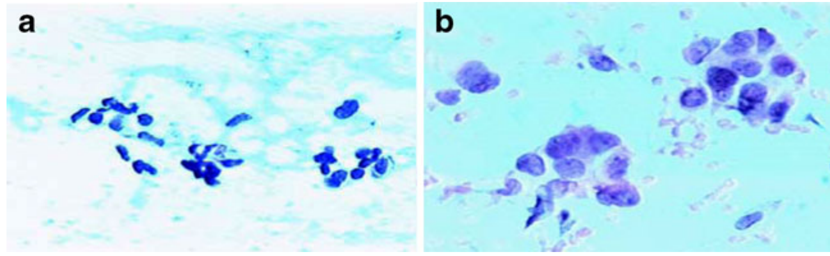

Fig. 3 Example images (**a**) Benign, (**b**) Malignant [36]

as the training samples. For semi-supervised learning, we construct the positive and negative constraints with half number of training samples according to the labeled class information. And the rest training samples are considered the unlabeled training samples.

In the experiments, we construct positive and negative graph with k and non-k nearest neighbor criterion. And the rest samples from the databases as the test samples. We use the positive graph to evaluate the classification performance of the compared algorithm. The dimension of feature vector is 9, Feature = [I0,PA500,HFS,DA,AREA,A/DA,MAX,IP, DR,P], and the corresponding elements of feature vector are described in Table 2.

For the purpose of comparison, we also implement the other semi-supervised and unsupervised learning algorithm, including Principal Component Analysis (PCA), Locality Preserving Projection (LPP) and Semi-supervised Locality Preserving Projection (SLPP). Since there is no labeled class information, we have not implemented the supervised learning for the comparison in the experiments for comparison. The experiment parameters including the weight parameters α, β and k are determined with the cross-validation method, moreover other algorithms achieves their optimal procedural parameters with cross-validation method.

As shown in Fig. 1, the proposed algorithm outperforms other algorithm for the same training and test sets. Although the recognition rate is not high, it indicates that it is feasible to improve the performance with unlabeled classes together with the side-information. Moreover, semi-supervised learning methods perform better than the unsupervised learning methods, for example, SLPP outperforms LPP, where SLDP-1: SLDP with side-information training samples), SLDP-2: SLDP with both side-information and unlabeled training samples.

Table 3 Data classification performance on WPBC dataset

Algorithms	Accuracy
PCA	90.34%
LPP	91.25%
SLDP-1	92.37%
SLDP-2	93.25%
K-SLDP	95.67%

Diagnostic experiments on Wisconsin Breast Cancer (WBC) database

Wisconsin Breast Cancer (WBC) Database [35] was obtained from the University of Wisconsin Hospitals in 1992. These samples from this database are collected periodically in the clinical cases, and therefore it reflects this chronological grouping of the data. The original data contains 699 samples, but the elements of feature vector of 16 samples are empty, so we delete these samples in our experiments. That is, there are 683 samples in the database in our experiments. There are two classes, i.e., Benign and Malignant, in the experiments. The features of these images are extracted, and the dimension of the feature is 9. Each elements of the feature are Clump Thickness, Uniformity of Cell Size, Uniformity of Cell Shape, Marginal Adhesion, Single Epithelial Cell Size, Bare Nuclei, Bland Chromatin, Normal Nucleoli, Mitoses. The data classification performance is shown as Fig. 2. Kernel based SLDP achieves the highest recognition accuracy compared with other algorithms including SLDP-1 (SLDP with side-information training samples), SLDP-2 (SLDP with both side-information and unlabeled training samples), LPP (Locality Preserving Projection) and PCA (Principal Component Analysis) under the same training and test set. The kernel parameters of K-SLDP are chosen with the cross-validation methods. Also, the side-information using is feasible to enhance the recognition performance compared with the unlabeled information with the unsupervised learning.

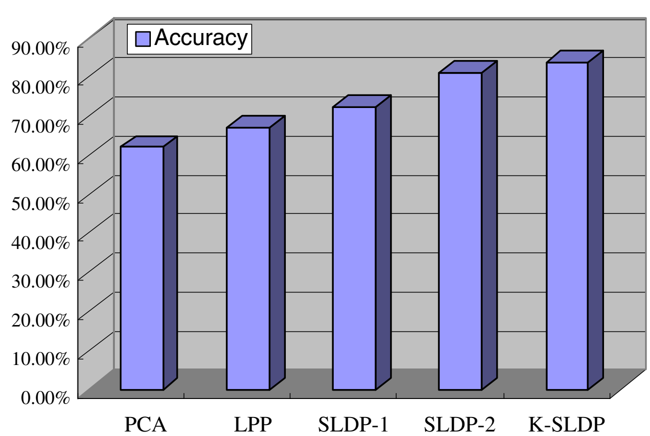

Fig. 4 Data classification performance on WPBC Dataset

Diagnostic experiments on Wisconsin Diagnostic Breast Cancer (WDBC) dataset

In 1995, *Wisconsin Diagnostic Breast Cancer* (*WDBC*) dataset [36] consist of 569 instances including 357 benign samples and 212 malignant samples. And each one represents FNA test measurements for one diagnosis case. For this dataset each instance has 32 attributes, where the first two attributes correspond to a unique identification number and the diagnosis status (benign / malignant). The rest 30 features are computations for ten real-valued features, along with their mean, standard error and the mean of the three largest values ("worst" value) for each cell nucleus respectively. These ten real values, which are depicted at Table 1, are computed from a digitized image of a fine needle aspirate (FNA) of breast tumour, describing characteristics of the cell nuclei present in the image and are recorded with four significant digits. Figure 3 depicts two images, which were taken from fine needle biopsies of breast tumours [36]. As shown in Table 3, the highest recognition accuracy is 95.67% with the proposed algorithm. Note: SLDP-1 (SLDP with side-information training samples), SLDP-2 (SLDP with both side-information and unlabeled training samples).

Prognostic experiments on Wisconsin Prognostic Breast Cancer (WPBC) dataset

In this section, we evaluate the proposed algorithm and the relative algorithms on Wisconsin Prognostic Breast Cancer (WPBC) dataset, which contains 198 samples and each sample represents one breast cancer case from 1984 to 1995, developed from University of Wisconsin Hospitals. In our experiments, we divide all these samples into recurrent and nonrecurrent classes. In the database, there are 198 samples, and 151 samples are nonrecurrent and the others are recurrent. The feature extraction methods on these images were proposed in the previous work [37–40]. Following this fruits, we also choose the following feature elements including Radius, Texture, Perimeter, Area, Smoothness, Compactness, Concavity, Concave points, Symmetry and Fractal dimension. As shown in Fig. 4, K-SLDP outperforms PCA, LPP, SLDP-1 and SLDP-2.

Conclusion

In this paper, we present an improved Semi-supervised Locality Discriminant Projections with Kernels and apply it into breast tissue classification for computer aided diagnosis of breast cancer. Firstly, we extend LPP with semi-supervised learning to enhance the classification performance and this method uses both side-information and unlabelled training samples. Secondly, we improve Semi-supervised LDP on the nonlinear classification performance with kernel trick. Finally, the breast tissue classification framework is presented based kernel based SLDP, and its feasibility is testified and performance is evaluated compared other popular methods. The proposed algorithm is also used in other research area, such as face recognition, image retrieval, video classification, and so on. Moreover, the procedural parameters of the proposed algorithm are chosen with cross-validation methods, so whether there are other methods to choose these parameters.

Acknowledgments This work is supported by National Science Foundation of China under Grant No. 61001165, Natural Science Foundation of Heilongjiang Province under Grant No. QC2010066. HIT Young Scholar Foundation of 985 Project.

References

1. http://seer.cancer.gov/cgi-bin/csr/1975_2001/search.pl#results, Estimated New Cancer Cases and Deaths for 2004.
2. U.S. National Institutes of Health, National Cancer Institute, http://cancernet.nci.nih.gov/.
3. Huo, Z., Giger, M., Vyborny, C., Wolverton, D., Schmidt, R., and Doi, K., Automated computerized classification of malignant and benign mass lesions on digital mammograms. *Acad. Radiol.* 5:155–168, 1998.
4. Setiono, R., Generating concise and accurate classification rules for breast cancer diagnosis. *Artif. Intell. Med.* 18:205–219, 2000.
5. Tourassi, G. D., Markey, M. K., Lo, J. Y., and Floyd, C. E., Jr., A neural network approach to breast cancer diagnosis as a constraint satisfaction problem. *Med. Phys.* 28:804–811, 2001.
6. Chen, D., Chang, R. F., and Huang, Y. L., Breast cancer diagnosis using self-organizing map for sonography. *Ultrasound Med. Biol.* 26:405–411, 2000.
7. Belhumeur, P. N., Hespanha, J. P., and Kriegman, D. J., Eigenfaces vs fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.* 19(7):711–720, 1997.
8. Batur, A. U., and Hayes, M. H., Linear subspace for illumination robust face recognition. *Proc. IEEE Int'l Conf. Comput. Vis. Pattern Recogn.* 296–301, 2001.
9. Hastie, T., and Stuetzle, W., Principal curves. *J. Am. Stat. Assoc.* 84:502–516, 1989.
10. Chang, K.-Y., and Ghosh, J., A unified model for probabilistic principal surfaces. *IEEE Trans. Pattern Anal. Mach. Intell.* 23 (1):22–41, 2001.
11. Mulier, F., and Cherkassky, V., Self-organization as an iterative kernel smoothing process. *Neural Comput.* 7:1165–1177, 1995.
12. Ritter, H., Martinetz, T., and Schulten, K., Neural computation and self-organizing maps. Addison-Wesley, 64–72, 1992.
13. Zhu, Z., He, H., Starzyk, J. A., and Tseng, C., Self-organizing learning array and its application to economic and financial problems. *Inform. Sci.* 177(5):1180–1192, 2007.
14. Yin, H., Data visualisation and manifold mapping using the ViSOM. *Neural Netw.* 15(8):1005–1016, 2002.
15. Tenenbaum, J. B., de Silva, V., and Langford, J. C., A global geometric framework for nonlinear dimensionality reduction. *Science* 290:2319–2323, 2000.
16. Roweis, S. T., Saul, L. K., and Dimensionality, N., Reduction by locally linear embedding. *Science* 290:2323–2326, 2000.

17. He, X., and Niyogi, P., Locality preserving projections. Proc. Conf. Advances in Neural Information Processing Systems pp. 585–591, 2003.
18. Shental, N., Hertz, T., Weinshall, D., and Pavel, M., Adjustment learning and relevant component analysis. Proceeding of the 7th European conference on computer vision, pp. 776–792, 2002.
19. BarHillel, A., Hertz, T., Shental, M., and Weinshall, D., Learning a Mahalanobis metric from equivalence constrains. *J. Mach. Learn. Res.* 6(6):937–965, 2005.
20. Xing, E. P., Jordan, M. I., and Russell, S., Distance metric learning with application to clustering with side-information. *Adv. Neural Inform. Process. Syst.* pp 505–512, MIT Press, 2003:
21. Tang, W., and Zhong, S., Pairwise constraints-guided dimensionality reduction. Proc. Data mining workshop on feature selection for data mining. pp. 59–66, 2006.
22. Zhang, D. Q., Zhou, Z. H., and Chen, S. C., Semi-supervised dimensionality reduction. Proc 7th SIAM International Conference on Data Mining, pp. 629–634. 2007.
23. Van Gestel, T., Baesens, B., and Martens, D., From linear to nonlinear kernel based classifiers for bankruptcy prediction. *Neurocomputing* 73(16–18):2955–2970, 2010.
24. Zhua, Q., Reformative nonlinear feature extraction using kernel MSE. *Neurocomputing* 73(16–18):3334–3337, 2010.
25. Veerabhadrappa, and Rangarajan, L., Diagonal and secondary diagonal locality preserving projection for object recognition. *Neurocomputing* 73(16–18):3328–3333, 2010.
26. Zhang, L., Qiao, L., and Chen, S., Graph-optimized locality preserving projections. *Pattern Recogn.* 43(6):1993–2002, 2010.
27. Wang, J., Zhang, B., Wang, S., Qi, M., and Kong, J., An adaptively weighted sub-pattern locality preserving projection for face recognition. *J. Netw. Comput. Appl.* 33(3):323–332, 2010.
28. Gui-Fu, Lu, Lin, Z., and Jin, Z., Face recognition using discriminant locality preserving projections based on maximum margin criterion. *Pattern Recogn.* 43(10):3572–3579, 2010.
29. Wang, X., Chung, Fu-lai, and Wang, S., On minimum class locality preserving variance support vector machine. *Pattern Recogn.* 43(8):2753–2762, 2010.
30. Cheng, J., Liu, Q., Lua, H., and Chen, Y. W., Supervised kernel locality preserving projections for face recognition. *Neurocomputing* 67:443–449, 2005.
31. Li, J. B., Pan, J. S., and Chu, S. C., Kernel class-wise locality preserving projection. *Inform. Sci.* 178(7):1825–1835, 2008.
32. Zhao, H., Sun, S., Jing, Z., and Yang, J., Local structure based supervised feature extraction. *Pattern Recogn.* 39(8):1546–1550, 2006.
33. Jossinet, J., Variability of impedivity in normal and pathological breast tissue. *Med. Biol. Eng. Comput.* 34:346–350, 1996.
34. Silva, J. E., Marques de Sá, J. P., and Jossinet, J., Classification of breast tissue by electrical impedance spectroscopy. *Med. Biol. Eng. Comput.* 38:26–30, 2000.
35. Mangasarian, O. L., and Wolberg, W. H., Cancer diagnosis via linear programming. *SIAM News* 23(5):1–18, 1990.
36. Wolberg, W. H., Street, W. N., Heisey, D. M., and Mangasarian, O. L., Computer-derived nuclear features distinguish malignant from benign breast cytology. *Hum. Pathol.* 26:792–796, 1995.
37. Street, W. N., Mangasarian, O. L., and Wolberg, W. H., An inductive learning approach to prognostic prediction. In: Prieditis, A., and Russell, S. (Eds.), *Proceedings of the Twelfth International Conference on Machine Learning*. Morgan Kaufmann, San Francisco, pp. 522–530, 1995.
38. Mangasarian, O. L., Street, W. N., and Wolberg, W. H., Breast cancer diagnosis and prognosis via linear programming. *Oper. Res.* 43(4):570–577, 1995.
39. Wolberg, W. H., Street, W. N., Heisey, D. M., and Mangasarian, O. L., Computerized breast cancer diagnosis and prognosis from fine needle aspirates. *Arch. Surg.* 130:511–516, 1995.
40. Wolberg, W. H., Street, W. N., and Mangasarian, O. L., Image analysis and machine learning applied to breast cancer diagnosis and prognosis. *Anal. Quant. Cytol. Histol.* 17(2):77–87, 1995.
41. Krinidis, S., and Pitas, I., Statistical analysis of human facial expressions. *Journal of Information Hiding and Multimedia Signal Processing* 1(3):241–260, 2010.
42. Jun-Bao Li, Mammographic image based breast tissue classification with kernel self-optimized fisher discriminant for beast cancer diagnosis. *J. Med. Syst.* doi:10.1007/s10916-011-9691-4)