



ELSEVIER

Contents lists available at ScienceDirect

Neurocomputing

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

Breast cancer discriminant feature analysis for diagnosis via jointly sparse learning



Heng Kong^{a,c}, Zhihui Lai^{b,c,*}, Xu Wang^b, Feng Liu^b

^a Department of Thyroid and Breast Surgery, Shen Zhen Nan Shan District People's Hospital, Shenzhen 518052, China

^b The College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China

^c Cancer Center, Georgia Regents University, Augusta, GA 30912, USA

ARTICLE INFO

Article history:

Received 3 September 2015

Received in revised form

4 November 2015

Accepted 19 November 2015

Communicated by: Yang Tang

Available online 4 December 2015

Keywords:

Breast cancer

Feature analysis

Diagnosis and classification

Sparse learning

ABSTRACT

Breast cancer is one of the most common malignant tumors among female. How to effectively discriminate the category of the cancers using the key features/factors is very important in the diagnosis and prediction. In this paper, Jointly Sparse Discriminant Analysis (JSDA) is proposed to explore the key factors in breast cancer and extract the key features for improving the accuracy in diagnosis and prediction. JSDA introduces the jointly sparse regular term (i.e. $L_{2,1}$ -norms term) to the criterion. A convergent iterative algorithm is designed to solve the optimization problem. It is shown that the proposed JSDA algorithm not only can learn the jointly sparse discriminant vectors to explore the key factors of the breast cancer in cancer pathologic diagnosis, but also can improve the diagnosis accuracy compared with the classical feature extraction and discriminant analysis algorithm. Experimental results on breast cancer datasets indicate that JSDA outperforms some well-known subspace learning algorithms in prediction accuracy, not matter they are non-sparse or sparse, particularly in the cases of small sample sizes. Data analysis shows that the key factors of the breast cancer explored by the JSDA are consistent with the practical experience.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

BREAST cancer feature extraction and feature analysis are of vital importance for cancer diagnosis, classification, and prediction. On the one hand, the information of tumor features can help the doctor discriminate category of the breast tumor (i.e. benign or malignant). On the other hand, since the breast cancer patients often encounter the high recurrence rate, it is of vital importance to predict the possibility of the recurrence. Moreover, it is also very important to explore which factors are the key variables associated with the categories and prediction of the breast cancer. Therefore, we aim to develop a computer-aided diagnosis method that can help the doctor determine the cancer's categories or to predict the recurrence rate [1]. One tractable method is to use the feature extraction and discriminant analysis technique to perform computer-aided diagnosis.

The classical feature extraction methods such as Principle Component Analysis (PCA) [2–4] and Linear Discriminant Analysis (LDA) [5–7] and its variations [8,9] are the most frequently used in the fields of pattern recognition, computer vision and data mining.

PCA and LDA were widely used in various fields including face recognition, palmprint recognition and gene expression data classification. For example, Ma and Kosorok used PCA to identify differential gene pathways [10]. The classical methods (i.e. PCA and LDA) only focus on the global structure of a dataset for dimensionality reduction.

In recent years, classifier design based on regression methods became a hot topic and a number of effective classification methods were proposed [11–13]. Particularly, sparse representation has been paid much attention on classifier design and feature extraction. For example, the sparse representation classifier was proposed for robust face recognition [14,15]. Qiao et al. introduced the L_1 norm sparse learning to learn the optimal representation and proposed Sparsity Preserving Projections (SPP) [16] for face feature extraction, in which the sparse reconstruction relationship was preserved. Gui et al. used the label information and L_1 norm sparse representation [17] for supervised feature extraction. In addition, the sparse L_1 graph was also used in subspace learning [18,19], spectral clustering [20] and label propagation [21].

However, the classical methods (i.e. PCA and LDA) and L_1 graph-based methods (i.e. SPP and its supervised extension [17]) in linear dimensionality reduction can only learn compact projections (i.e. elements in the projections are usually non-zero). Thus, such projections lack reasonable interpretation. In order to

* Corresponding author at: The College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China.

E-mail addresses: generaldoc@126.com (H. Kong), lai_zhi_hui@163.com (Z. Lai), wangxu@szu.edu.cn (X. Wang), feng.liu@szu.edu.cn (F. Liu).

overcome this drawback in these compact subspace learning methods, the L_1 norm based sparse regression methods [22–24] were proposed to learn the sparse projections for feature extraction and classification. It is experimentally shown that these sparse learning methods not only can enhance the prediction accuracy but also strengthen the generalization ability and the robustness on prediction. Therefore, sparse subspace learning methods have attracted much attention.

Among the sparse extensions of the classic feature extraction methods, the well-known ones are Sparse PCA (SPCA) [25], Multilinear SPCA (MSPCA) [26] and Sparse Discriminant Analysis (SDA) [27]. In SPCA and MSPCA, the L_1 norm penalty is imposed on the regression-type optimization problem derived from PCA to learn the sparse projections, which contain most of the variance of the dataset. SDA [27] rewrites the ridge regression using the label matrix for sparse discriminant projection learning. By using the Elastic Net [24] for regression, the most important/contributive variables are selected to form the projection vectors in SPCA, SDA and SLDA.

Although SPCA and SDA can obtain a set of sparse projections for feature analysis, the non-zero elements in the projections are usually different from their locations. Thus, SPCA and SDA cannot give the consistent interpretation in the features or variables level. To address this problem, Nie et al. proposed the Robust Feature Selection (RFS) [28] via joint $L_{2,1}$ -norm learning. It is shown that the $L_{2,1}$ -norm based regression is robust to outliers and can select features across all data points with joint sparsity. Recently, Gu et al. [29] proposed to use the joint Feature Selection and Subspace Learning (FSSL) method for image recognition. With these efficient methods for solving the $L_{2,1}$ -norm based learning model, joint feature selection and feature extraction methods were widely used in image recognition [30–32], web image annotation [33] and multimedia data understanding [34].

In the past decades, feature selection and discrimination have been very active primarily due to the advances in bioinformatics where a large amount of genomic and proteomic datum are produced for biological and biomedical studies. For example, in genomics, to select a few of relevant genes out of several thousands of genes becomes a key problem. And to identify the meaningful proteomic features from mass spectrometry is crucial for disease diagnosis and protein-based biomarker profiling [35].

As one of the most frequently met diseases, breast cancer has been widely investigated for several decades. In this paper, we focus on the breast cancer feature analysis and discrimination for diagnosis using the newly proposed $L_{2,1}$ -norm based feature selection and discriminant analysis for helping diagnosis and data understanding. Concretely, we explore the important factors in a sparse manner on breast cancer feature extraction and discrimination (i.e. diagnosis), where only the key features related to the discriminant information are jointly extracted in the sparse manner. And thus it is expected to further improve the understanding of the cancer data and strengthen the generalization ability and the robustness for discrimination. By integrating the $L_{2,1}$ -norm penalty term on the Maximal Margin Criterion (MMC) [36,37], a new model is proposed for effective breast cancer diagnosis. The main contributions of this paper are as follows:

(1) We propose a general jointly sparse subspace discriminant analysis method called Jointly Sparse Discriminant Analysis (JSDA) for breast cancer data diagnosis and the key feature/factor extraction and understanding. And the convergence analysis is also presented for the proposed JSDA, which indicates that JSDA will converge to the local optimum.

(2) Extensive experiments show that JSDA outperforms the classical subspace learning methods and the sparse ones for breast cancer identification. Moreover, the key information embedded in the breast cancer data is explored, which will be very important to

the diagnosis. Factor analysis results are consistent with the doctor's practical experience.

The rest of the paper is organized as follows. In Section 2, LDA and MMC are reviewed. In Section 3, SJDA is proposed and theoretical analysis is presented to guarantee the convergence of the proposed algorithm. Experiments are carried out to evaluate SJDA algorithm in Section 4, and the conclusions are given in Section 5.

2. Outline of LDA and MMC

In this section, LDA and MMC are reviewed. Let matrix $X = [x_1, x_2, \dots, x_N]$ be the data matrix including all the training samples $\{x_i\}_{i=1}^N \in R^m$ in its columns. In practice, the feature dimension m is often very high. The goal of feature extraction is to transform the data from the originally high-dimensional space to a low-dimensional one, i.e.

$$y = U^T x \in R^d \tag{1}$$

for any $x \in R^m$ with $d \ll m$, where $U = (u_1, u_2, \dots, u_d)$ and $u_i (i = 1, \dots, d)$ is an m -dimension column vector.

2.1. LDA

LDA is a supervised learning algorithm which aims to maximize the ratio between the between-class scatter and the within-class scatter. Let c denote the total number of classes and l_i denote the number of training samples in the i -th class. Let x_i^j denote the j -th sample in i -th class, \bar{x} be the mean of all the training samples, \bar{x}_i be the mean of the i -th class. The between-class and within-class scatter matrices can be evaluated by:

$$S_b = \sum_{i=1}^c l_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T \tag{2}$$

$$S_w = \sum_{i=1}^c \sum_{j=1}^{l_i} (x_i^j - \bar{x}_i)(x_i^j - \bar{x}_i)^T \tag{3}$$

LDA aims to find an optimal projection U such that the ratio of the between-class scatter to within-class scatter is maximized, i.e.

$$U = \arg \max_U \frac{|U^T S_b U|}{|U^T S_w U|} \tag{4}$$

where $U = \{u_i | i = 1, 2, \dots, d\}$ is the set of generalized eigenvectors of S_b and S_w corresponding to the d largest generalized eigenvalues $\{\lambda_i | i = 1, 2, \dots, d\}$, i.e.

$$S_b u_i = \lambda_i S_w u_i, \quad i = 1, 2, \dots, d \tag{5}$$

Once the optimal projection matrix U is obtained, the high-dimensional patterns can be projected to the low-dimensional subspace for classification.

2.2. MMC

LDA always encounters the small sample size problem when the feature dimensions are larger than the number of samples. In order to overcome the irregularity in computing the inverse of the within-class scatter matrix, MMC uses the differential criterion to compute the optimal projection which has the very similar function as LDA. The objective function of MMC is as follows:

$$\begin{aligned} \max_u & \quad u^T S_b u - u^T S_w u \\ \text{s.t.} & \quad u^T u = 1 \end{aligned} \tag{6}$$

where u is a column vector. By using the Lagrange multiplier method, it is easily shown that the above optimization problem

Download English Version:

<https://daneshyari.com/en/article/408961>

Download Persian Version:

<https://daneshyari.com/article/408961>

[Daneshyari.com](https://daneshyari.com)