Available online at www.sciencedirect.com
SCIENCE DIRECT
Pattern Recognition 39 (2006) 980–987
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/patcog

Breast cancer diagnosis using genetic programming generated feature

Hong Guo[1], Asoke K. Nandi*

Signal Processing and Communications Group, Department of Electrical Engineering and Electronics, The University of Liverpool, Brownlow Hill, Liverpool, L69 3GJ, UK

Received 21 July 2005

Abstract

This paper proposes a novel method for breast cancer diagnosis using the feature generated by genetic programming (GP). We developed a new feature extraction measure (modified Fisher linear discriminant analysis (MFLDA)) to overcome the limitation of Fisher criterion. GP as an evolutionary mechanism provides a training structure to generate features. A modified Fisher criterion is developed to help GP optimize features that allow pattern vectors belonging to different categories to distribute compactly and disjoint regions. First, the MFLDA is experimentally compared with some classical feature extraction methods (principal component analysis, Fisher linear discriminant analysis, alternative Fisher linear discriminant analysis). Second, the feature generated by GP based on the modified Fisher criterion is compared with the features generated by GP using Fisher criterion and an alternative Fisher criterion in terms of the classification performance. The classification is carried out by a simple classifier (minimum distance classifier). Finally, the same feature generated by GP is compared with a original feature set as the inputs to multi-layer perceptrons and support vector machine. Results demonstrate the capability of this method to transform information from high-dimensional feature space into one-dimensional space and automatically discover the relationship among data, to improve classification accuracy.

Keywords: Feature extraction; Genetic programming; Fisher discriminant analysis; Pattern recognition

1. Introduction

Feature extraction is one of the most important tasks in pattern recognition problems. Due to the fact that the feature space with high dimensionality requires a large amount of computation, feature extraction should have the capability to project original features into lower feature space for reducing the dimensionality of patterns presented to the classifier, while improving the classification efficiency.

Generally, there are two steps in feature extraction: first, the information relevant for classification is extracted from raw data to a original feature vector with m dimensions; second, extracted feature vector with n dimensions ($n < m$) is created from the parameter vector. The task of linear feature extraction measure transformation algorithm is to reduce the dimensionality of pattern observation space by finding a suitable linear subspace in which the class separability is optimally maintained. Principal component analysis (PCA) and Fisher linear discriminant analysis (FLDA) are the best known linear feature extractors [1]. Both of these methods search optimal directions for the projection of input data onto a lower dimensional space.

In recent years, some variation of linear discriminant analysis (LDA) for feature extraction have been studied due to the limitation of LDA. An alternative FLDA (AFLDA) has been proposed to overcome the limitation of FLDA by replacing the original scatter with a new scatter measure for binary-class problem [2]. A weighted variant of Fisher criterion associated with linear discriminant analysis for multiclass has been introduced by Loog et al. [3], where they proposed an eigenvector-based heteroscedastic LDA for multiclass problem in Ref. [4]. Petridis and Perantonis

* Corresponding author. Tel.: +44 151 794 4525; fax: +44 151 794 4540.
E-mail address: A.Nandi@liverpool.ac.uk (Asoke K. Nandi).

[1] Hong Guo would like to acknowledge the financial support of Overseas Research Studentship committee (UK), the University of Liverpool and the University of Liverpool Graduates Associations (HK).

0031-3203/$30.00 © 2005 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2005.10.001

[5] have proposed a model-independent reformulation of the criteria related to three linear discriminant feature extraction methods that stress their similarities and elucidate their differences. However, these methods have a limitation for the data which are not linearly separable since it is difficult to capture a nonlinear relationship with a linear mapping [6]. To overcome the weakness of linear feature extraction, nonlinear versions of PCA and FLDA have been proposed. Two variations of FLDA for extracting nonlinear feature have been presented in Ref. [7]. A nonlinear feature extraction method based on an analysis of the current kernel Fisher discriminant algorithm has been developed in Ref. [8].

Recently, applications of machine learning algorithms for the feature extraction have become increasingly popular with techniques, such as evolutionary programming (EP), genetic algorithms (GAs) and genetic programming (GP). GA based feature selection was carried out in Ref. [9] for the classification of bearing faults using vibration signals. Raymer et al. [10] shows that a hybrid of KNN classifier and GA can considerably improve discrimination accuracy and reduce the dimensionality. GP was first introduced by Koza [11] and has been proposed as a machine learning method in pattern recognition field. GP was tested in six medical diagnosis problems and compared with results obtained by neural networks [12]. The feasibility of applying GP to multi-category pattern classification problems was studied in Ref. [13]. While all the methods mentioned above utilize GP as a classifier, GP-based feature extraction was conducted in Ref. [14] to improve the classification performance and reduce dimensionality in the medical domain. However, this system was unable to adequately sample the search space for high-dimensional problems and the major disadvantage was the computational complexity. A GP-based feature generation method for machine condition monitoring was proposed in Ref. [15], where multiple features were generated using GP and later used as the inputs to ANN classifiers for the classification of different bearing conditions.

In this paper, we develop a new measure to overcome the limitation of Fisher criterion. GP is employed to generate a single nonlinear feature based on this modified Fisher criterion measure to improve the classification accuracy for breast cancer detection. As a machine learning method, GP exhibits intelligent behavior to perform feature generation. During the evolutionary process, the modified Fisher criterion is used to evaluate the effectiveness of each feature in helping GP select the best features by which the pattern vectors from different classes are well separated. Also, this approach provides a solution to obtain a single tree/feature by only a single run of GP. Compared with the framework proposed in Ref. [15], this approach significantly reduces the number of features required to describe the problem and make the classification efficient and reliable in one-dimensional feature space.

The rest of the paper is organized as follows: A new feature extraction measure (modified FLDA (MFLDA)) and some classical linear feature extraction measures PCA, FLDA, AFLDA are addressed in Section 2. Feature generator based on GP and different Fisher criterion is described in Section 3. In Section 4, some simple and classical classifiers are explained. In Section 5, a number of experiments for breast cancer are conducted, the comparison of classification performance using features extracted by linear feature extraction measures (PCA, FLDA AFLDA and MFLDA) and GP generated features based on different Fisher criterion (FLDA, AFLDA and MFLDA) are demonstrated. Finally, based on the experimental results, the advantages and limitations of GP-based feature extraction method are concluded in Section 6.

2. Linear feature extraction measures

Feature extraction/selection is used to project data into a lower dimensional space for data visualization and increasing classifier efficiency. It can be conducted independently or combined with a classifier; linear or nonlinear based on the transformation function; supervised or unsupervised by a prior knowledge.

The four linear measures introduced in this section are all independent feature extraction methods which based on statistical data analysis technique and determine an optimal linear transformation of a real-valued m-dimension feature vector X into another n-dimensional $(n < m)$ transformed vector Y [1]:

$$Y = WX. \tag{1}$$

2.1. Principal component analysis (PCA)

PCA is a well-established unsupervised method for feature extraction and feature dimensionality reduction in terms of a standardized linear transformation which maximizes the variance of the transformed feature space. PCA does not require the information about datasets containing observations labelled by corresponding classes.

The $m \times m$ covariance matrix is given by

$$S = \sum_{i=1}^{N} (x_i - \mu)(x_i - \mu)^{\mathrm{T}}, \tag{2}$$

where x_i is the ith observation, $1 \leqslant i \leqslant N$, N is the number of all observations and μ is the global mean $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$.

The n components of a given observation vector can be obtained by solving the eigenvalue problem

$$SW_i^{\mathrm{T}} = \lambda_i W_i^{\mathrm{T}}, \quad 1 \leqslant i \leqslant m, \tag{3}$$

where λ is a set of eigenvalues and W is a set of the corresponding eigenvectors. The n principal components of projected feature vector Y are obtained by the first largest n eigenvalues and corresponding eigenvectors [1].