Neurocomputing 83 (2012) 1–11

Contents lists available at SciVerse ScienceDirect

Neurocomputing

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

# An ensemble kernel classifier with immune clonal selection algorithm for automatic discriminant of primary open-angle glaucoma

Lijun Cheng [a], Yongsheng Ding [a,b,*], Kuangrong Hao [a,b], Yifan Hu [a]

[a] College of Information Sciences and Technology, Donghua University, Shanghai 201620, PR China
[b] Engineering Research Center of Digitized Textile and Fashion Technology, Ministry of Education, Donghua University, Shanghai 201620, PR China

ARTICLE INFO

*Article history:*
Received 4 April 2011
Received in revised form
17 September 2011
Accepted 17 September 2011
Communicated by Z. Wang
Available online 5 January 2012

*Keywords:*
Ensemble kernel classifier
Kernel principal component analysis
Support vector machine
Immune clonal selection algorithm
Primary open-angle glaucoma

ABSTRACT

An ensemble kernel classifier is proposed in this paper by integrating a kernel principal component analysis (KPCA) with a support vector machine (SVM) as well as an immune clonal selection algorithm (ICSA). The KPCA approach is used to extract features, whereas the SVM technique is employed to deal with classification, and the ICSA is applied to optimize the parameters of the proposed scheme. The proposed ensemble classifier can automatically select the kernel type and optimize its parameter sets, in order to produce various SVM classifiers with different kernels. Regardless of whether the data is linear or nonlinear, an optimum classification result can be obtained. In order to demonstrate the effectiveness of the classifier, it is applied to discriminate the primary open-angle glaucoma (POAG) using a standard classification dataset. Experimental results reveal that the proposed ensemble classifier is accurate and more effective when compared to other approaches in the literature. It is envisaged that ensemble kernel classifier could hold a high potential in classification of pattern recognition problems.



## 1. Introduction

Automatic classification is an important field in pattern recognition. Many classification methods have been proposed in the literature, such as neural networks [1], nearest neighbors [2], decision trees [3], and kernel-based methods (support vector machines) [4–6]. Among these methods, kernel-based methods, although still in its infancy, have attracted greater research interest and have successfully been applied to many areas [7,8]. Many linear learning algorithms have been kernelized successfully by making use of the kernel framework. The inner product is replaced with an appropriate positive definite function (kernel function) so that the resulting nonlinear map can be implicitly implemented by the kernel function from the original data space $X$ to the high-dimensional feature space $F$ [6]. The attractiveness of such algorithms stems from their elegant treatment of nonlinear problems and their efficiency in solving high-dimensional problems.

Support vector machines (SVMs) are kernel classification technique that is established on the unique theory of structural risk minimization principle. SVMs can be robust to over-fitting problem and have been more widely employed in diverse areas ranging from pattern recognition [7–9] to modeling of nonlinear dynamic systems [10].

In a SVM classifier, the first important step is feature selection and extraction. Most data sets contain a certain amount of redundancy or irrelevant features that will not aid the knowledge discovery. In fact, these redundant features will mislead the data analyses, and hinder other relevant features that are present in the data. Consequently, one often tries to reduce the number of features by applying feature selection schemes [9,11,12]. Kernel principal component analysis (KPCA), as a typical feature extraction method, is a nonlinear principal component analysis (PCA) technique that is developed using the kernel method [11], which is closely related to the SVM [6].

To speed up the convergence of SVM and achieve high recognition accuracy, a general kernelization framework used in KPCA is applied to SVM in this paper. KPCA is viewed as a data mapping or transformation to SVM, so that new features are used as inputs to SVM to solve the classification problem. The framework is controlled by two kernel types in the KPCA and SVM. Different kernels in the KPCA and SVM will result in different structures for the classifier. It should be pointed out that the nature of the input data types whether they are linear or nonlinear does not play an important role. The optimal model can then be chosen by the system to fit the data automatically. Once the nonlinear kernel in the KPCA procedure is performed, the

* Corresponding author at: Engineering Research Center of Digitized Textile and Fashion Technology, Ministry of Education, Donghua University, Shanghai 201620, PR China.
*E-mail address:* ysding@dhu.edu.cn (Y. Ding).

0925-2312/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2011.09.030

system can directly perform linear learning algorithm using the SVM subsequently [12,13].

A highly effective model can only be built after its parameters are carefully determined. The researchers in [14–16] have pointed out that the search for optimal parameters of the SVM plays a crucial role in building a prediction model with a high prediction accuracy and stability. However, few studies have been devoted to optimizing the overall dynamic parameters of the KPCA and SVM integrated framework. The overall performance of the integrated KPCA and SVM scheme depends on not only kernel parameters in the SVM, but also on different kernels and identified number of the principal components in the KPCA, as well as on the penalty parameter $C$ in the SVM algorithm. These parameters can control both the model complexity as well as the training error.

In this paper, in order to automatically optimize the integrated KPCA and SVM parameters, an adaptive artificial immune algorithm, designated as immune clonal selection algorithm (ICSA) is proposed. Artificial immune algorithm is an effective method in solving a wide range of optimization problems [17,18] and has been widely and successfully applied to complicated real-world problems in science and engineering, such as multivariable function optimization [26], and designing stack filters [27]. Compared with the commonly used standard genetic algorithm (GA), ICSA's antibody fitness calculation maintains a good diversity of the population. It overcomes the premature convergence of GA and limits its random roaming. In order to demonstrate the effectiveness of the proposed ensemble kernel model, we use it to identify primary open-angle glaucoma (POAG) disease as an auxiliary diagnosis tool, and compare it with the SVM machine learning model. Furthermore, other linear and nonlinear experiments in certain general data sets have also been conducted to test the performance of as proposed method.

The main contribution of this work is on developing a hybrid KPCA+SVM+ICSA framework, which can classify the data automatically. The model is then used for the first time to identify the POAG as an auxiliary diagnosis tool. The proposed ensemble kernel model is shown to be effective by conducting a number of experiments and is shown to possess a high potential as a useful tool in classification of pattern recognition problems.

The remainder of the paper is organized as follows: in Section 2, the SVM classification and the KPCA dimension reduction methods are reviewed and the basic structure of the kernel optimization classifier is presented. In Section 3, we provide the details on the parameter optimization using the ICSA. In Section 4, the proposed optimization classifier is applied to the POAG discriminant and is tested using various data sets. Finally, conclusions are stated in Section 5.

## 2. Basic methods

### 2.1. Kernel principal component analysis

The kernel principal component analysis (KPCA) [9,11] aims to map the given data points $x=\{x_1,x_2,\ldots,x_n\}$, $x_i\in R^n$, $i=1,2,\ldots,n$, from the input space $R^n$ to a high dimensional (possibly infinite-dimensional) feature space $F$ by a nonlinear function $\varphi: R^n\rightarrow F$. The training data matrix is denoted by $\varphi(x)=\{\varphi(x_1),\varphi(x_2),\ldots,\varphi(x_n)\}$.

In fact, one can avoid performing the nonlinear mapping and only compute the dot products $\varphi(x)^T\cdot\varphi(y)$ in the feature space $F$ by introducing a kernel function $k(x,y)$ [11]. Some of the most widely used kernel functions are shown in Table 1.

In Table 1, $d$ and $\delta$ denote the parameters that affect the performance of the KPCA. The polynomial kernel and the radial basis function (RBF) kernel always satisfy Mercer's theorem. The polynomial kernel can be simplified as a linear kernel when the parameter $d=1$. In practice, the radial basis kernel function is most commonly used [4]. Different kernel functions implicitly define the form of the mapping and the feature space, thus actually determine how well the nonlinearity of a system can be captured.

**Table 1**
Some of the most widely used kernel functions.

| Type of classifier | Kernel function |
|---|---|
| Linear kernel | $k(x,y)=(x\cdot y)$ |
| Polynomial kernel | $k(x,y)=(x\cdot y+1)^d$ |
| Radial basis kernel | $k(x,y)=\exp(-\lvert x-y\rvert^2/\delta)$ |

Given the mapping data set $\varphi(x)$, which has been centralized $\sum_{k=1}^{n}\varphi(x_k)=0$ in the feature space $F$ and the covariance matrix

$$C^F=\frac{1}{n}\sum_{i=1}^{n}\varphi(x_i)\cdot\varphi(x_i)^T \tag{1}$$

one has to solve the following eigenvalue problem in the feature space to diagonalize the covariance matrix, namely

$$\lambda v=C^F v \tag{2}$$

where the eigenvalue $\lambda\geq 0$ and the principal component $v\in F\backslash\{0\}$. According to the reproducing kernel theory, there exists a set of coefficients $\alpha_i$ in $F$ such that

$$v=\sum_{i=1}^{n}\alpha_i\varphi(x_i) \tag{3}$$

The problem is then reduced to obtaining the coefficients $\alpha_i$. This can be formulated according to the following eigenvalue problem when Eq. (3) is substituted into Eq. (2)

$$\tilde{K}\alpha=n\lambda\alpha \tag{4}$$

where $\tilde{K}$ is a standardized $K$ matrix, and $\alpha$ is the corresponding eigenvector, the matrix $K$ is an $n\times n$ kernel matrix of the training samples, in which $k_{ij}=k(x_i,x_j)=\langle\varphi(x_i),\varphi(x_j)\rangle$. The relationship between $\tilde{K}$ and $K$ is governed by

$$\tilde{K}=K-LK-KL+LKL \tag{5}$$

where $L$ is an $n\times n$ matrix in which all the elements are equal to $1/n$ and $n$ is the sample number. Therefore, the eigenvalues are ordered as $\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_n$, and the corresponding eigenvectors set is given by $\alpha_1,\alpha_2,\ldots,\alpha_n$. In fact, the dimensionality of the data can be reduced by retaining only the first $m$ eigenvectors in the feature space, typically $m\ll n$.

The samples are extracted through projecting onto the $m$ eigenvectors. An input sample $x$ is projected onto the $k$th principal component $v$ in the feature space $\varphi(x)$ according to

$$\beta_k=\langle v_k,\varphi(x)\rangle=\sum_{i=1}^{m}\alpha_i^{(k)}K(x_i,x) \tag{6}$$

Therefore, $x$ is expressed as a vector in $F$ after $\varphi(x)$ is normalized, that is

$$(\beta_1,\beta_2,\ldots,\beta_m)^T=\left[\frac{\sum_{i=1}^{n}\alpha_i^{(1)}[K(x_i,x)-\sum_{i=1}^{n}K(x_i,x)/n]}{\sqrt{n\lambda_1^{\alpha}}},\ldots,\right.$$
$$\left.\times\frac{\sum_{i=1}^{n}\alpha_i^{(m)}[K(x_i,x)-\sum_{i=1}^{n}K(x_i,x)/n]}{\sqrt{n\lambda_m^{\alpha}}}\right]^T \tag{7}$$

where $\alpha^{(1)}$, $\alpha^{(2)}$, … ,$\alpha^{(m)}$ are the $m$ eigenvectors associated with the first $m$ largest eigenvalues $\lambda_1^{\alpha},\lambda_2^{\alpha},\ldots,\lambda_m^{\alpha}$, respectively, $\alpha_j^{(i)}$ denotes the $j$th component of the vector $\alpha^{(i)}$, and $n$ is the number of samples.