Information Sciences 257 (2014) 70–80

Contents lists available at ScienceDirect

Information Sciences

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

Kernel self-optimization learning for kernel-based feature extraction and recognition

Jun-Bao Li [a,b,*], Yun-Heng Wang [a], Shu-Chuan Chu [c], John F. Roddick [c]

[a] Innovative Information Industry Research Center, Shenzhen Graduate School, Harbin Institute of Technology, China
[b] Department of Automatic Test and Control, Harbin Institute of Technology, China
[c] School of Computer Science, Engineering and Mathematics, Flinders Univ. of South Aus., Australia

ARTICLE INFO

Article history:
Received 28 December 2011
Received in revised form 14 July 2013
Accepted 5 August 2013
Available online 14 August 2013

Keywords:
Kernel method
Data-dependent kernel
Kernel self-optimization
Kernel principal component analysis
Kernel discriminant analysis
Kernel locality preserving projection

ABSTRACT

Kernel learning is becoming an important research topic in the area of machine learning, and it has wide applications in pattern recognition, computer vision, image and signal processing. Kernel learning provides a promising solution to nonlinear problems, including nonlinear feature extraction, classification and clustering. However, in kernel-based systems, the problem of the kernel function and its parameters remains to be solved. Methods of choosing parameters from a discrete set of values have been presented in previous studies, but these methods do not change the data distribution structure in the kernel-based mapping space. Accordingly, performance is not improved because the current kernel optimization does not change the data distribution. Based on this problem, this paper presents a uniform framework for kernel self-optimization with the ability to adjust the data structure. The data-dependent kernel is extended and applied to kernel learning, and optimization equations with two criteria for measuring data discrimination are used to solve the optimal parameter values. Some experiments are performed to evaluate the performance in popular kernel learning methods, including kernel principal components analysis (KPCA), kernel discriminant analysis (KDA) and kernel locality-preserving projection (KLPP). These evaluations show that the framework of kernel self-optimization is feasible for enhancing kernel-based learning methods.

1. Introduction

Dimensionality reduction-based feature extraction is an important and essential process in many data analysis systems and has many applications in facial recognition, handwriting recognition, human facial expression analysis and speech recognition. Machine learning systems have applications in texture analysis [11], smart homes [12], fault detection [13], SOM [5], biometrics [28], data streams [31], and credibility measures [9], multimodal biometrics [21] and fuzzy system [24]. Dimensionality reduction (DR) is an important feature extraction method, and many DR methods have been proposed, such as linear discriminant analysis (LDA) and principal components analysis (PCA) [1,25]. These DR methods are linear and fail to capture the nonlinear relationships among the data during the feature extraction process. To improve the performance of these DR methods, a Kernel trick was introduced to DR, and this method performed well on a nonlinear feature extraction task [4]. Accordingly, kernel versions of linear DR methods were developed, such as kernel principal components analysis (KPCA), kernel discriminant analysis (KDA) [30] and improved methods [2,18,32,6,19]. Moreover, the kernel locality preserving projection (KLPP) is widely used in the facial recognition and radar target recognition fields [16,8]. Improved KLPP meth-

* Corresponding author at: Department of Automatic Test and Control, Harbin Institute of Technology, China. Tel.: +86 452 86402998.
E-mail addresses: junbaolihit@gmail.com (J.-B. Li), scchu@bit.kuas.edu.tw (S.-C. Chu).

0020-0255/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ins.2013.08.011

ods have been proposed in previous studies [30,7,29,10]. Kernel learning methods improve the performance of many linear feature extraction methods owing to the self-adaptive data distributions for classification. The geometrical structure of the data in the kernel mapping space is determined entirely by the kernel function, and the kernel function has a significant influence on kernel-based systems. The discriminative ability of the data in the feature space could be problematic with an inappropriate kernel. In the previous studies [10,17,14], researchers applied a method of selecting the kernel parameters from a set of discrete values, but this method does not change the geometrical structure of the data for classification. Xiong proposed a data-dependent kernel for kernel optimization [27], and Amari presented a support vector machine classifier that modifies the kernel function [1]. In our previous work [23,15], we presented a data-dependent kernel-based KDA algorithm for a facial recognition application.

Some kernel-based learning methods and application systems were developed for pattern recognition, data mining, computer vision, image and signal processing, and other applications. The recognition and prediction performance of learning systems are improved by the kernel trick. The current kernel-based system still maintains the kernel function and its parameters, and the kernel function directly determines the data distribution. Accordingly, the performance of the kernel learning system is affected by the kernel function. This paper presents a framework for kernel optimization for a kernel-based learning system. To verify the effectiveness of the kernel optimization scheme, the proposed kernel optimization framework is applied to popular kernel learning methods, including KPCA, KDA and KLPP.

The rest of this paper is organized as follows. In Section 2, the kernel optimization framework is presented. A comprehensive experimental comparison and analysis are implemented to test the performance of kernel self-optimization on simulated data, the UCI, ORL and YALE databases, in Section 3.

2. Data-dependent kernel self-optimization

2.1. Motivation and framework

Computation of the kernel matrix is a key step of kernel-based learning methods for classification and clustering. This procedure has two problems: computational burden and kernel selection. This paper presents a framework for kernel optimization to improve the kernel-based learning system. We apply a data-dependent kernel function in the kernel-optimization framework. With a data-dependent kernel, the geometric structure of the data is modified adaptively. Some kernel parameter-optimized learning methods have been proposed in previous work [10,17,14], but a general framework of data-dependent kernel optimization has not been proposed. As shown in Fig. 1, the framework includes three procedures of kernel optimization, training and testing for general kernel-based applications.

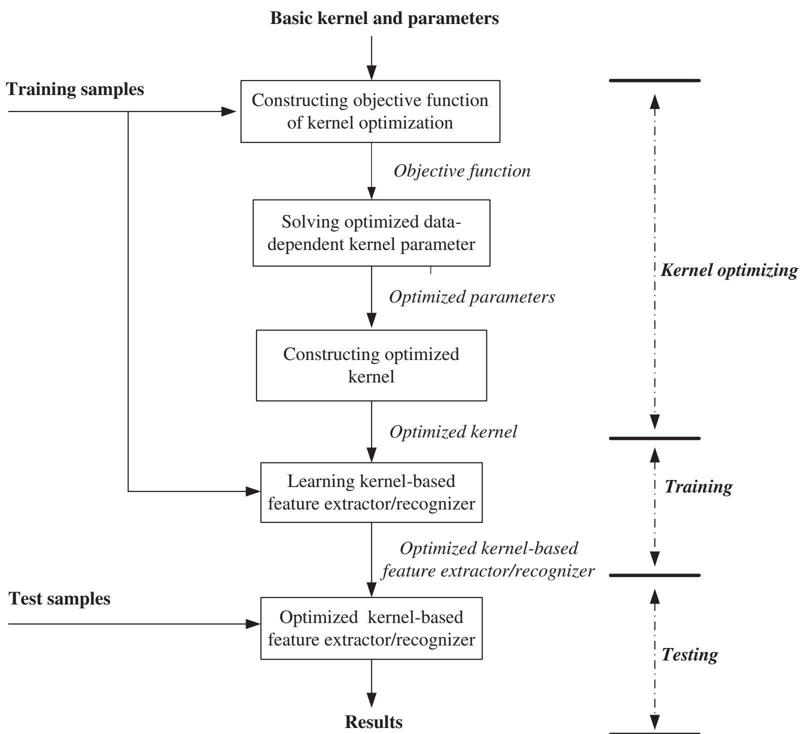

Fig. 1. Framework of kernel optimization and its application.