Neurocomputing 175 (2016) 610–634

Contents lists available at ScienceDirect

Neurocomputing

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

Improved multi-kernel classification machine with Nyström approximation technique and Universum data

Changming Zhu

College of Information Engineering, Shanghai Maritime University, Shanghai 201306, PR China

A R T I C L E I N F O

Article history:
Received 15 June 2015
Received in revised form
8 September 2015
Accepted 28 October 2015
Communicated by Haiqin Yang
Available online 9 November 2015

Keywords:
Multiple kernel learning
Nyström approximation
Reduced computation complexity
Generalization risk analysis
Pattern classification
Universum learning

A B S T R A C T

Universum learning can reflect priori knowledge about application domain and improve classification performances. The kernelized modification of Ho–Kashyap algorithm with squared approximation of the misclassification errors (KMHKS) is an effective learning machine for nonlinearly separable classification problems. While KMHKS only adopts one kernel function, so a multi-kernel classification machine with reduced complexity named Nyström approximation matrix with multiple KMHKSs (NMKMHKS) has been developed. But NMKMHKS has to initialize many parameters and it has not an ability to process noises well. To this end, some scholars propose an improved multi-kernel classification machine with Nyström approximation technique (INMKMHKS). INMKMHKS is based on a new way of generating kernel functions and a new Nyström approximation technique. Related experiments have validated that INMKMHKS possesses five advantages: (1) avoiding the problem of setting too many parameters; (2) keeping comparable space and computational complexities after comparing with NMKMHKS; (3) having a tighter generalization risk bound in terms of Rademacher complexity analysis; (4) possessing an ability to process noises and practical images; (5) a superior recognition can be gotten in a strong correlation between multiple used kernels, which can give a guide advice for choosing kernels. But for traditional multiple kernel learning (MKL), many classification machines including INMKMHKS focus on MKL optimizations, for example, the optimization of model for a MKL. It is difficult for us to find or create a new optimization way now. Indeed, if one pays more attention to data themselves, performance of MKL will also have an improvement. This paper adopts INMKMHKS as a basic learning machine and focuses on data themselves, then it introduces Universum learning into the procedure of INMKMHKS and proposes Universum-based INMKMHKS (Uni-INMKMHKS). The motivation of Uni-INMKMHKS is that it can design a learning machine from data perspective and avoid paying too much attention to MKL optimization which is difficult for scholars to some extent. The contribution of Uni-INMKMHKS is that it has a better recognition than INMKMHKS in average with Universum learning used and inherits the advantages of INMKMHKS simultaneously.

1. Introduction

1.1. Background

Pattern recognition is a branch of machine learning that focuses on recognition of patterns and regularities in data [1]. For recognition of patterns, many approaches have been proposed and some fields are developed as well. One of them is support vector machine (SVM) [2] which is also a state-of-the-art learning machine. The objective of SVM is to map the original input patterns into a high-dimensional kernel space using the 'kernel trick' and then to construct a linear decision function in this space so that patterns become separated with a maximum margin. Suppose there is a data set $X=\{x_i,\varphi_i\}_{i=1}^{N}$ and pattern is $x_i\in\mathcal{R}^d$, class label is $\varphi_i\in\{1,-1\}$. The standard SVM should solve the following problem:

$$\min_{\omega,b,\xi}\ \frac{1}{2}\omega^T\omega+C\sum_{i=1}^{N}\xi_i$$
$$\text{s.t.}\quad \varphi_i(\omega^T\Phi(x_i)+b)\geq 1-\xi_i$$
$$\xi_i\geq 0,\quad i=1,2,\ldots,N \tag{1}$$

where Φ is the mapping function. The solutions of ω and b form the linear decision function. ξ is the slack variable. Parameter C is a trade-off coefficient between the classification performance and the Vapnik–Chervonenkis (VC) dimension [2] and the choice of this parameter may have a strong effect on the performance of a

E-mail addresses: zcm19880301@163.com, cmzhu@shmtu.edu.cn

http://dx.doi.org/10.1016/j.neucom.2015.10.102
0925-2312/

learning machine. Generally, SVM always solves its dual problem as follows:

$$\min_{\alpha} \quad \tfrac{1}{2}\alpha^T Q \alpha - \alpha^T \alpha$$
$$\text{s.t.} \quad 0 \le \alpha_i \le C$$
$$\varphi^T \alpha = 0 \tag{2}$$

where the i-th row and j-th column datum of $Q \in \mathcal{R}^{N \times N}$, i.e., Q_{ij} is equal to $\varphi_i \varphi_y \Phi(x_i)^T \Phi(x_j)$. Then the solution of Eq. (2) is used to compute ω and b in Eq. (1). To avoid computing $\Phi(x_i)^T \Phi(x_j)$, the SVM uses a kernel function $k(x_i, x_j)$ as a replacement. Some commonly used kernels are linear kernel, polynomial kernel, or Radial Basis Functions (RBF) kernel.

Besides SVM, there are other two hot spots of present research which are also validated to be effective for the design of learning machines. One is kernel learning and the other is Universum learning.

Kernel learning aims to adopt kernel functions to map the original input patterns into the kernel space. In such a kernel space, these original input patterns maybe linear separable even though they are nonlinear separable in the input space. Then in this kernel space, we can classify these patterns with some classical linear learning machines used. In the recent three decades, some kernel learning machines have been proposed. For example, single kernel learning (SKL) [2], multiple kernel learning (MKL) [3], kernelized modification of Ho–Kashyap algorithm with squared approximation of the misclassification errors (KMHKS) [4], multi-views KMHKS (MultiV-KMHKS) [5], multi-kernels discriminant analysis (MKDA) with semi-definite program (SDP) [6], ℓ_p-MKDA with semi-infinite program (SIP) [7], multi-view KMHKS based on Nyström approximation (MVNA-KMHKS) [8], multi-kernel classification machine with Nyström approximation technique (NMKMHKS) [9], and improved NMKMHKS (INMKMHKS) [10]. Here, KMHKS is a SKL while MultiV-KMHKS, MKDA (SDP), ℓ_p-MKDA (SIP), MVNA-KMHKS, NMKMHKS, and INMKMHKS are MKLs. Related experiments have validated that (1) MKL always has a better performance than a SKL; (2) these kernel learning machines always have better performances than SVM; (3) INMKMHKS has a better average performance than NMKMHKS, MVNA-KMHKS, MultiV-KMHKS, MKDA (SDP), MKL, and ℓ_p-MKDA (SIP) [10].

Universum learning [11] is proposed to incorporate priori knowledge about application domain into the learning process. This knowledge is derived from additional unlabeled patterns which are also called virtual examples or Universum patterns and these Universum patterns are not real training patterns. Many learning machines with Universum leaning are proposed, for example, Cherkassky and Dai have proposed Universum support vector machine (U-SVM) [12] and Liu et al. have proposed self-Universum support vector machine (SUSVM) [13]. In U-SVM, with the comparison between U-SVM and traditional SVM, it is found that quality of Universum patterns can affect the performance of a learning machine, especially on some small-scale data sets. Furthermore, Chen and Zhang [14] have validated that Universum patterns distributing between object classes are more useful for the generation of classification hyperplane. Such useful Universum patterns are called In-Between Universum patterns (IBU) and IBU can improve the performances of learning machines on some semi-supervised learning problems. Due to the advantages of Universum learning, the applications of it have been gradually spread into document clustering [15], body pose recognition [16], Boosting strategy [17], dimensionality reduction technique [18], and multi-view learning [19]. But in terms of the creation of Universum patterns, experiments about U-SVM, SUSVM, and other related learning machines have validated that different creation ways bring different performances. Moreover, all creation ways they used have high computation complexities.

1.2. Motivation

As what we said before, MKL always has a better performance than a SKL, and many scholars pay more attention to MKL optimizations, for example, the optimization of model for a MKL. They always want to create a new MKL model, improve an original MKL one, or propose some new optimization methods for the model. They think that MKL optimization brings a greater improvement for a learning machine. But we should notice that MKL optimization is not an easy task. First, creating a new and simple model is not very easy. It seems that scholars are hard to create a simple and new model which can be effective than SVM or neural network (NN). Second, improving the original model is also difficult. If we add more regularization terms for an original model, we will always encounter the complexity problem. Suppose that the improved model brings higher time and space complexities, we have to say that the improved model is not effective. Third, the proposal of some new optimization methods for the model of a MKL is also difficult. Although some learning machines including MKDA (SDP) and ℓ_p-MKDA (SIP) adopt effective optimization methods for MKL, developing new optimization methods is also a challenge due to this always demands the scholars and readers have advanced mathematics knowledge.

According to the above contents, we know MKL optimization is important for the design of a kernel learning machine. But this is not very easy. So in this paper, we want to design a learning machine from another perspective. This perspective is the data. We know data themselves have some characteristics, for example, mean, variance, feature information. Making full use of these characteristics and paying attention on data structures will help us to find some new methods to improve the performance of a learning machine. Here, since Universum learning has been validated to be effective and a learning machine with Universum learning has a better classification performance, thus this paper introduces Universum learning and makes use of Universum patterns so as to design a learning machine from the data perspective. Furthermore, since INMKMHKS has a better performance than other MKL-related learning machines which are mentioned in this paper, so here we adopt INMKMHKS as a basic machine. With the combination of INMKMHKS and Universum learning, this paper proposes an Universum-based INMKMHKS (Uni-INMKMHKS). Moreover, from the data perspective, in order to create Universum patterns better and more feasible, we design a new strategy called Creating In-Between Universum data (CIBU) to create Universum patterns. Different from existing creation methods, CIBU can reduce the computation complexity and make full use of the domain knowledge of whole data distribution.

As a summary, we adopt INMKMHKS as a basic learning machine, then we apply Universum leaning and CIBU simultaneously so as to propose Uni-INMKMHK. The motivation of Uni-INMKMHKS is that it can design a learning machine from data perspective and avoid paying too much attention to MKL optimization which is difficult for scholars and readers to some extent.

1.3. Contribution

Our proposed Uni-INMKMHKS has two contributions. (1) Since we adopt INMKMHKS as a basic learning machine, so Uni-INMKMHKS inherits the advantages of INMKMHKS: I.e., Avoiding the problem of setting too many parameters. Keeping comparable space and computational complexities after comparing with NMKMHKS, i.e., Uni-INMKMHKS also reduces the computational complexity of finding the solution scale from $O(Mn^3)$ to $O(Mnm^2)$,