Applied Soft Computing 30 (2015) 14–25

Contents lists available at ScienceDirect

Applied Soft Computing

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

Multiple Empirical Kernel Learning with dynamic pairwise constraints

Zhe Wang*, Qi Fan, Daqi Gao*

Department of Computer Science & Engineering, East China University of Science & Technology, Shanghai 200237, PR China

ARTICLE INFO

Article history:
Received 31 July 2013
Received in revised form 13 October 2014
Accepted 22 January 2015
Available online 30 January 2015

Keywords:
Multiple Kernel Learning
Dynamic pairwise constraints
Empirical Kernel Mapping
Prior knowledge
Pattern recognition

ABSTRACT

Unlike the traditional Multiple Kernel Learning (MKL) with the implicit kernels, Multiple Empirical Kernel Learning (MEKL) explicitly maps the original data space into multiple feature spaces via different empirical kernels. MEKL has been demonstrated to bring good classification performance and to be much easier in processing and analyzing the adaptability of kernels for the input space. In this paper, we incorporate the dynamic pairwise constraints into MEKL to propose a novel Multiple Empirical Kernel Learning with dynamic Pairwise Constraints method (MEKLPC). It is known that the pairwise constraint provides the relationship between two samples, which tells whether these samples belong to the same class or not. In the present work, we boost the original pairwise constraints and design the dynamic pairwise constraints which can pay more attention onto the boundary samples and thus to make the decision hyperplane more reasonable and accurate. Thus, the proposed MEKLPC not only inherits the advantages of the MEKL, but also owns multiple folds of prior information. Firstly, MEKLPC gets the side-information and boosts the classification performance significantly in each feature space. Here, the side-information is the dynamic pairwise constraints which are constructed by the samples near the decision boundary, i.e. the boundary samples. Secondly, in each mapped feature space, MEKLPC still measures the empirical risk and generalization risk. Lastly, different feature spaces mapped by multiple empirical kernels can agree to their outputs for the same input sample as much as possible. To the best of our knowledge, it is the first time to introduce the dynamic pairwise constraints into the MEKL framework in the present work. The experiments on a number of real-world data sets demonstrate the feasibility and effectiveness of MEKLPC.

1. Introduction

Kernel-based learning method has been successfully applied [23,26,32,33]. It maps the input space into a feature space, i.e. $\Phi(x): x \to \mathcal{F}$. There are two kinds of $\Phi(x)$ including *implicit* and *explicit* forms represented by $\Phi^i(x)$ and $\Phi^e(x)$, respectively. The implicit mapping $\Phi^i(x)$ called Implicit Kernel Mapping (IKM) [23] is achieved by a kernel function $k(x_i, x_j) = \Phi^i(x_i) \cdot \Phi^i(x_j)$ by which the explicit form of $\Phi^i(x)$ is not necessary to be given. In contrast, the $\Phi^e(x)$ called Empirical Kernel Mapping (EKM) [35] has to give the explicit form of $\Phi^e(x)$ to get the exact features of x in feature space. On the other hand, according to the number of kernels used in the learning process, kernel-based learning can be divided into Single Kernel Learning (SKL) [7] and Multiple Kernel Learning (MKL) [39]. SKL maps the input space into one feature space, while MKL maps it into multiple feature spaces through the corresponding mapping functions. Most of the existing MKL using the IKM is called the Multiple Implicit Kernel Learning (MIKL). In contrast, the MKL employing the EKM is called the Multiple Empirical Kernel Learning (MEKL).

Although MIKL has got much attention [1,18,27,28] in recent decades, it is the necessity of inner-product in IKM that restricts other methods unsatisfying this formulation to be kernelized. For instance, it is pretty difficult to formulate the Kernel Direct Discriminant Analysis [22]. Moreover, for some linear discriminant analysis algorithms, such as the Uncorrelated Linear Discriminant Analysis [38], and the Orthogonal Linear Discriminant Analysis [37], to directly kernelize them via the kernel trick is impossible, since these algorithms need to compute the singular value decomposition [34]. Fortunately, most methods can be directly implemented in EKM due to the explicit representation of the corresponding feature vectors, which results in an easy way to extend the application of the kernel-based method. Wang et al. [32] have pointed out that the EKM is exactly equal to the IKM, and the mapped spaces generated by them have the same geometrical structure. In [25,35], it is shown that the EKM is much easier in processing and analyzing the adaptability of kernels for the input space than the IKM. Moreover, the MKL is more efficient in depicting heterogeneous data sources than the SKL. To a certain extent, MKL also relaxes the model selection about kernels. Thus in this paper, we focus on the MKL with EKM, i.e. MEKL. MEKL can be viewed as the data-dependent kernel learning model since $\Phi^e(x)$ is directly generated based on the input data. The existing MEKL is treated as an alternative way of kernel learning. It mainly introduces the existing techniques into EKM, and gives the illustration on the differences between EKM and IKM. However, few researches concentrate on the inherent characteristic of EKM. This paper gives an investigation onto the structure of the empirically generated feature spaces. The traditional MEKL problem is to optimize the learning framework by minimizing the empirical risk, and the regularization risk, as well as the loss term of the multiple feature spaces [32]. We can find that it ignores some information among the training samples, which may provide great contribution to the classification performance.

This paper explores the relationship between samples in each feature space through the pairwise constraints. It is known that pairwise constraint provides the

* Corresponding authors. Tel.: +86 15000779526.
E-mail addresses: wangzhe@ecust.edu.cn (Z. Wang), gaodaqi@ecust.edu.cn (D. Gao).

http://dx.doi.org/10.1016/j.asoc.2015.01.040
1568-4946/

relationship between two samples, which tells whether they belong to the same class or not. In this paper, we boost the original pairwise constraints, which are considered to be static during the training process, by designing the dynamic pairwise constraints, which are dynamically constructed by the boundary samples, to result in the decision hyperplane more reasonable and accurate. Furthermore, we introduce the dynamic pairwise constraints into MEKL framework to propose a Multiple Empirical Kernel Learning with dynamic Pairwise Constraints method (MEKLPC). The proposed MEKLPC not only inherits the advantages of MEKL, but also owns multiple folds of prior information. Firstly, in each mapped feature space, MEKLPC gets the side-information to promote the classification performance. Here, the side-information is the dynamic pairwise constraints which are constructed by the samples near the decision boundary, i.e. the boundary samples. Secondly, in each mapped feature space, MEKLPC still measures the empirical risk and generalization risk. Lastly, different feature spaces can agree to their outputs for the same input sample as much as possible through a loss term of these multiple feature spaces. To the best of our knowledge, the present work is the first time to introduce the dynamic pairwise constraints into MEKL framework.

In order to generate an instance of the proposed MEKLPC, we adopt our previous MEKL work named MultiK-MHKS [32] as the incorporated paradigm. In practice, MEKLPC firstly maps the input data into multiple feature spaces by the corresponding EKMs. Then, it introduces the dynamic pairwise constraints into each feature space to obtain the corresponding decision function. Finally, the final decision function is obtained by combining the decision functions of all feature spaces. To validate the feasibility and effectiveness of MEKLPC, the experiments on a number of real-world data sets are implemented, and demonstrate that MEKLPC provides a superior performance.

The rest of this paper is organized as follows. Section 2 presents a brief review on the related work of the existing pairwise constraints. Section 3 demonstrates the designed dynamic pairwise constraints. Section 4 gives the detailed illustration on the proposed MEKLPC. The experimental results of MEKLPC on real-world data sets are reported in Section 5. Finally, the conclusions are represented in Section 6.

2. Related work

In recent years, the pairwise constraint is attracting more and more interests since they have been proven to be an effective way in expressing the prior knowledge of the relationship of each pair samples. Thus, they are introduced into some basic algorithms, such as semi-supervised learning, feature selection, and ensemble learning, to make use of the pairwise information of the samples.

In semi-supervised learning, lots of literatures report several different semi-supervised clustering algorithms by introducing the pairwise constraints, such as semi-supervised hard clustering (PCKmeans [3], COPKmeans [31]), the semi-supervised fuzzy clustering (AFCC) [14], and the semi-supervised spectral clustering. But, the penalty cost function of the pairwise constraints could not be appropriate for cooperating with the clustering objective function by competitive agglomeration. To address this problem, Gao et al. [13] propose a new semi-supervised fuzzy clustering algorithm (SCAPC) by redefining the objective function. Zeng et al. [40] propose a discriminative learning approach to incorporate pairwise constraints into a two-class maximum margin clustering framework. Moreover, lots of researches concentrate on evaluating their effects of noise imposing on semi-supervised clustering [40,43], due to that the pairwise constraints provided by distinct domain experts may conflict to each other. To avoid the conflictions, Jiang et al. [16] propose the elite pairwise constraints, including elite *must-link* (EML) and elite *cannot-link* (ECL) constraints [16]. The EML and ECL have no conflictions since both EML and ECL are required to be satisfied in every optimal partition.

Some researchers introduce the pairwise constraints into the feature selection methods to improve the feature selection performance. The feature selection is traditionally divided into two groups, i.e. the supervised and unsupervised feature selection, based on whether the class labels are used or not. Zhang et al. [42] propose a supervised feature selection method, called constraints score by using pairwise constraints as the supervised information. Their experimental results present that constraint score achieves impressive results on real world data sets. But, one major disadvantage is that the performance is dependent on the good selection of the composition and cardinality of constraint set [29]. To offset this disadvantage, Sun et al. [29] propose a Bagging Constraint Score method (BCS) by importing bagging into constraint score. Yang et al. [36] propose a novel hypothesis-margin based approach for feature selection with side pairwise constraints, named Simba-sc which only adopts the *cannot-link* constraints as the supervised information because that *cannot-link* constraints have more contribution to hypothesis-margin or margin [36].

Since the ensemble method can obtain stronger classifier by combining multiple weaker ones, some literatures adopt the ensemble method to promote the flexibility of the pairwise constraints. Zhang et al. [41] propose a novel framework to extend AdaBoost with pairwise constraints based on the gradient descent view of boosting. Their proposed framework is almost as simple and flexible as AdaBoost. Although many algorithms are developed to classify multi-label data, they usually do not consider the pairwise relations between the sample labels, which play important roles in classification [41]. Moreover, to extend traditional pairwise constraints to multi-label scenario, Li et al. [20] present a novel multi-label classification framework named Variable Pairwise Constraint projection for Multi-label Ensemble (VPCME) by adopting a boosting-like strategy.

In this paper, we extend the pairwise constraints to the supervised classification methods by incorporating pairwise constraints into MEKL methods. Unlike the traditional methods which only consider pairwise constraints static during the training stage, we design a novel method which can dynamically determine the pairwise constraints after each training iteration. With dynamic pairwise constraints, the proposed MEKLPC results in the superior performance on most of the adopted real-world data sets.

3. Dynamic pairwise constraints

In this section, we present the demonstration on the designed dynamic pairwise constraints by which we dynamically determine which samples should be considered to construct the pairwise constraints during the training processes.

For the training samples $\{(x_i, \varphi_i)\}_{i=1}^N$, $\varphi_i \in \{+1, -1\}$, a pairwise constraint set $C = \{x_{j1}, x_{j2}, l_j\}_{j=1}^n$, where $l_j = 1$ indicates that the pair of samples (x_{j1}, x_{j2}) must link, $l_j = -1$ denotes that they cannot link, i.e. x_{j1} and x_{j2} in the same class if $l_j = 1$, or in different classes if $l_j = -1$. Suppose that the decision function $f(x) = \widehat{w}^T \cdot x + w_0 = w^T \cdot x'$, where $w = [\widehat{w}^T, w_0]^T$ and $x' = [x^T, 1]^T$. A preliminary work on pairwise constraints adopts $L(x_i, x_j, l) = |f(x_i) - lf(x_j)|$ as the pairwise constraint penalty function. But, we find that it may not work robustly in practice when considering each pair samples as the pairwise constraint. Fig. 1 gives the illustration on this phenomenon. The solid red line represents the decision hyperplane $f(x) = -0.48x^1 + 0.203x^2 + 0.06$ which is learned by our proposal by considering each pair points as the pairwise constraint in the input space. The pairwise constraints of samples (supposing $x_1 = [0.104, 0.420]^T$ and $x_2 = [0.285, 0.509]^T$) in red diamonds are *must-link*. Although x_1, x_2 are divided to the same side of the hyperplane, the pairwise penalty value $L(x_1, x_2, 1) = |f(x_1) - f(x_2)| = 0.188$. However, as to the points (supposing $x_3 = [0.306, 0.520]^T$ and $x_4 = [0.422, 0.677]^T$ belonging to the same class) in the red squares, we find that x_3, x_4 are divided into the different sides of the hyperplane, but the pairwise loss value $L(x_3, x_4, 1) = |f(x_3) - f(x_4)| = 0.143$ is smaller than $L(x_1, x_2, 1)$. It is obvious that we should pay more attention to the pair samples (x_3, x_4) to achieve a robust hyperplane. But, the traditional pairwise constraint penalty strategy provides more concentration on (x_1, x_2) rather than (x_3, x_4) when considering each pair of samples as the pairwise constraint.

To overcome this problem, we design a dynamic pairwise constraint selection method which dynamically determines the points as the pairwise constraint samples. As shown in Fig. 2, we only