Pattern Recognition 48 (2015) 3797–3809

Contents lists available at ScienceDirect

Pattern Recognition

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

Matrixized learning machine with modified pairwise constraints

Yujin Zhu, Zhe Wang*, Daqi Gao*

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

ARTICLE INFO

Article history:
Received 13 October 2014
Received in revised form
17 May 2015
Accepted 22 May 2015
Available online 4 June 2015

Keywords:
Matrixized classifier
Regularized learning
Modified pairwise constraints
Pattern recognition

ABSTRACT

Matrix-pattern-oriented Classifier Design (MatCD) has been demonstrated to be effective in terms of the classification performance since it utilizes two-sided weight vectors to constrain the matrix-based patterns. However, the existing MatCD might not be able to acquire the prior distribution knowledge, such as the relationship between two patterns. Inspired by the Pairwise Constraints (PC) method, i.e., *must-links* and *cannot-links* between the patterns, this paper introduces a new regularization term named R_p with a modified PC method into MatCD. The new classifier design strategy is expected to not only learn the structural information of each pattern itself, but also acquire the prior distribution knowledge about each constrained pair with both the discrimination metric from the traditional PC and the spatial distance measure from the heat kernel method. In practice, this paper selects one typical matrixized classifier named MatMHKS as the basic building block and introduces the term R_p into it. The newly-proposed classifier is named MLMMPC and the subsequent experiments validate the effectiveness of it. Two major contributions of this paper can be concluded as (1) improving the existing matrix-pattern-oriented classifier design techniques and (2) modifying the traditional PC method by combining the discrimination metric and the distance measure together.

1. Introduction

The matrixized method has been applied to solve the matrix-based pattern representation problems and demonstrated to be competitive and even superior to the original vector-oriented method based on the following evidence. Firstly compared to the existing vector-oriented learning method, the matrixized one is designed to operate a matrix-based pattern directly without transforming the pattern into a vector as the preprocessing [12]. To explain, we consider that if a matrix-based pattern (such as an image) $\tilde{\mathbf{A}} \in \mathbb{R}^{m \times n}$ is transformed into a vector $\tilde{\mathbf{a}} \in \mathbb{R}^{mn \times 1}$, the dimensionality of the vector might be so high that the existing vector-oriented methods using the weight vector $\tilde{\mathbf{w}} \in \mathbb{R}^{mn \times 1}$ could be confronted with the curse of dimensionality when dealing with $\tilde{\mathbf{a}}$ [11]. On the other hand, the matrixized method could utilize the two weight vectors $\tilde{\mathbf{u}} \in \mathbb{R}^{m \times 1}$ and $\tilde{\mathbf{v}} \in \mathbb{R}^{n \times 1}$ to constrain the pattern $\tilde{\mathbf{A}}$ on its row and column respectively to avoid the exorbitant size of the pattern, i.e., $\tilde{\mathbf{u}}^T\tilde{\mathbf{A}}\tilde{\mathbf{v}}$ rather than $\tilde{\mathbf{w}}^T\tilde{\mathbf{a}}$ [12]. Furthermore, the memory consumption of the weight vector is decreased from the $m \times n$ of the vectorized method to $m+n$ of the matrixized one. Secondly, the matrixized method is expected to prevent the structural information of the matrix-based pattern $\tilde{\mathbf{A}}$ from being fully destroyed since the method imposes the elements in the same row/column of $\tilde{\mathbf{A}}$ to accept the same sub-weight of $\tilde{\mathbf{u}}/\tilde{\mathbf{v}}$. In practice, each element of one matrix-based pattern usually owns less informative knowledge while shares more tied relationship with the other elements, so that the degree of freedom of the selected model should be less than $m \times n$ [18]. Hence, we could say that the matrixized method is just consistent with the specialty of the matrix-based pattern. Thirdly, the matrixized method is easy to be approached and could be conveniently deteriorated to the vectorized one. It implies that the matrixized method could deal with the vector-based patterns too. Moreover, the matrixized method could provide a new perspective for operating the original vector-based patterns, such as the multi-view learning using different matrices transformed from the same vector [48] and the interpolation mapping expanding a vector to a matrix with extra information [47]. The detailed comparison between the matrixized model and the vectorized model is demonstrated in Section 2.

In application, the matrixized method is suitable for both classifier design [12,45] and feature extraction [6,34,60]. To be convenient, we adopt the same abbreviations used in the previous work [45] to name the Matrix-pattern-oriented Classifier Design as MatCD, the Matrix-pattern-oriented Feature Extraction as MatFE, the Vector-pattern-oriented Classifier Design as VecCD, and the Vector-pattern-oriented Feature Extraction as VecFE. The early work focuses on using the new-designed MatFE to substitute the conventional VecFE, such as the Two-Dimensional Principal Components Analysis (2DPCA) [52], the Two-Dimensional Linear

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

http://dx.doi.org/10.1016/j.patcog.2015.05.023
0031-3203/

Discriminant Analysis (2DLDA) [27], the Two-Directional 2DPCA ($(2D)^2$PCA) [60], and the Two-Directional 2DLDA ($(2D)^2$LDA) [34]. In spite of the improvement in feature extraction, the subsequent classifier is still based on the traditional vectorized methods, i.e., the whole process could be concluded as MatFE + VecCD. To search for the new possibility, the previous work [45] investigates the combination forms between the matrixized and vectorized method, finding that the form MatFE + MatCD indeed performs effectively and efficiently. Besides, since both the matrixized and the vectorized method follow the minimum risk framework (1) [12,40], it is natural and rational to introduce the matrixized method into classifier design in order to make simple and effective classifiers that can manipulate matrix-based patterns directly, i.e., developing MatCD.

$$\min J = R_{emp} + CR_{reg}, \tag{1}$$

There are two terms on the right of the framework (1). The first term is defined as the empirical risk term R_{emp}. As for the vectorized method, the core mathematical form of R_{emp} is $\tilde{\mathbf{w}}^T\tilde{\mathbf{a}}$ where $\tilde{\mathbf{w}}$ is the weight vector and $\tilde{\mathbf{a}}$ is the pattern. In the matrixized method, the form becomes $\tilde{\mathbf{u}}^T\tilde{\mathbf{A}}\tilde{\mathbf{v}}$ where the weight vectors are $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{v}}$. Further, the second term is the regularization term R_{reg}, which makes the objective function smooth and C is the penalty coefficient. It could be seen that the matrixized method is factually a natural extension of the vectorized method and could provide an alternative feasible pattern representation in classifier design [45]. One typical example of the successfully-designed matrix-oriented classifiers is the Matrix-pattern-oriented Modified Ho–Kashyap classifier (MatMHKS) [12].

Even though the matrixized method utilizes two weight vectors as the side constraints to impose a classifier to learning more structural information of elements from the matrix, there might be some useful knowledge neglected. Reasons include (1) The term R_{emp} treats the input patterns separately and compares the predicted result of each pattern to its true result in turn, without analyzing the relationships among different patterns. (2) The term R_{reg} only considers the global smoothness. Overall, neither R_{emp} nor R_{reg} is available to acquire any distribution knowledge of the dataset.

To solve the problem, this paper designs a new framework for MatCD to boost the original framework (1) by introducing a novel regularization term R_p:

$$\min J = R_{emp} + CR_{reg} + \lambda R_p. \tag{2}$$

The term R_p is expected to learn more spatial information, especially the relationship between patterns, as well as maintaining the superiority of the traditional matrixized learning machine. Moreover, λ is the penalty coefficient. Inspired by a classic effective method named the Pairwise Constraints (PC) [19,20,24,57], this paper tries to utilize the PC strategy to formulate R_p. The introduction of PC is considered to be theoretically feasible based on two considerations: firstly, there are studies successfully utilizing PC to solve or boost the vectorized classification problems [16,22,28,33,55,58] and it could be proved that the matrixized method is the vectorized model imposed with the extra structural constraints [12]. Therefore, it seems natural to introduce PC into the matrixized method. Secondly, PC mainly focuses on the distribution information while the original constraints of the matrixized method consider the elements relationship of one pattern, which seems complementary rather than contradictory. Specially, the traditional PC might decide that the pair of two spatially remote patterns is equally significant to the pair of two spatially near patterns if the discrimination metric values of both these two pairs are similar, i.e., the traditional PC pays more attention to the prediction difference than the spacial distance between pairs. This inclination might be of less benefit for some linear classification problems. As a solution, this paper modifies the traditional PC strategy by introducing the heat-kernel-based distance measure method [5,32,**?**] to strengthen the relationship between patterns. More detailed description on both the traditional and the modified PC is presented in Section 3. With the help of the modified PC method, the designed regularization term R_p is expected to explicitly learn the spatial relationship of the patterns in the whole data domain thus improving the classification generalization. To our best knowledge, it is the first time that PC is introduced into the matrixized classifier design framework. The newly-proposed method is named MLMMPC for short. In practice, the previous work MatMHKS is used as the basic building block of MLMMPC, since both the two methods are originated from the same framework.

The major contribution of this paper lies in the following aspects:

- *Significance*: This paper extends the existing matrix-pattern-oriented classifier design techniques so as to make the proposed learning machine as a special example of the framework (2). Further, the relationship between the matrixized method and the vectorized method is analyzed.
- *Novelty*: (1) This paper focuses on the pattern distribution in whole input space by introducing a new regularization term R_p into the traditional matrixized framework; it is expected to explicitly acquire the relationship between the original patterns with the help of the modified PC method. As the result, a novel algorithm named MLMMPC is proposed. (2) This paper develops the traditional PC with a well-known distance-measure strategy and introduces the modified PC into the supervised matrix-oriented classifier design tasks to make the learning process more adaptable and flexible.
- *Experiments*: (1) The feasibility and effectiveness of the proposed algorithm is validated in subsequent experiments, compared to some other classical algorithms on both matrix- and vector-based datasets. (2) Influences of special parameters of MLMMPC are investigated.

The rest of this paper is organized as follows. Section 2 discusses the relationship between the matrixized method and the vectorized method and gives a brief introduction on the matrixized classifiers, especially MatMHKS. Section 3 reviews the related PC method and proposes the new modified PC strategy. Section 4 presents the architecture of the proposed method MLMMPC. Section 5 reports on all the experimental results. Finally, conclusions are given in Section 6.

2. Matrixized learning machine (MLM)

In this part, the relationship between the matrixized method and the vectorized method is discussed. Both theoretical demonstrations and related application instances are presented. Then, the related work of MatCD is introduced. Specifically, the derivation from the original Ho–Kashyap algorithm (HK) [17] to the typical matrix-based algorithm MatMHKS [12] is described.

2.1. Relationship between the matrixized method and the vectorized method

As the special case of the nth-order-tensor-oriented methods based on the rank-1 projection while $n=2$ [4,37], the matrixized learning method can directly manipulate the matrix-based patterns without reshaping them to vectors. To further explain the matrixized method, its relationship with the vectorized method is demonstrated in this part. Before we start, important notations