Globalized and localized matrix-pattern-oriented classification machine

Zhe Wang[*], Yujin Zhu, Daqi Gao[**], Weibin Guo

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

ARTICLE INFO

ABSTRACT

Inspired by the matrix-based methods used in feature extraction and selection, one matrix-pattern-oriented classification framework has been designed in our previous work and demonstrated to utilize one matrix pattern itself more effectively to improve the classification performance in practice. However, this matrix-based framework neglects the prior structural information of the whole input space that is made up of all the matrix patterns. This paper aims to overcome such flaw through taking advantage of one structure learning method named Alternative Robust Local Embedding (ARLE). As a result, a new regularization term R_{gl} is designed, expected to simultaneously represent the globality and the locality of the whole data domain, further boosting the existing matrix-based classification method. To our knowledge, it is the first trial to introduce both the globality and the locality of the whole data space into the matrixized classifier design. In order to validate the proposed approach, the designed R_{gl} is applied into the previous work matrix-pattern-oriented Ho-Kashyap classifier (MatMHKS) to construct a new globalized and localized MatMHKS named GLMatMHKS. The experimental results on a broad range of data validate that GLMatMHKS not only inherits the advantages of the matrixized learning, but also uses the prior structural information more reasonably to guide the classification machine design.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

This paper aims to boost the classification performance of one existing matrix-pattern-oriented learning framework through utilizing both the globality and the locality of the whole matrix pattern space. The problem aroused from the traditional vector-oriented methods while transforming the input matrix-based pattern into a vector [3,6] might find a solution under the help of the matrixized method, which manipulates a matrix-based pattern directly, avoiding such vectorizing so as to overcome the loss of the structural information between the elements of images [23] and the higher computational complexity [5]. The matrixized method is firstly proposed in feature extraction [16,32], e.g., the Two-Dimensional Principal Component Analysis (2DPCA) [32] extracting features directly from matrix patterns and shown to be better than the classical Principal Component Analysis (PCA) [12] and Two-Dimensional Linear Discriminant Analysis (2DLDA) [16] shown to not only increase the classification performance, but also

overcome the singularity problem implicitly in classical Linear Discriminant Analysis (LDA) [8]. Further, (2D)²PCA [34] and (2D)²LDA [18] are independently developed. Both of them can simultaneously consider the row and column directions of matrix patterns, achieving the same or even higher recognition accuracy than the 2DPCA and 2DLDA which only consider the row direction of matrix patterns.

However, the classifiers used after those matrixized feature extractors still resort to the traditional vector-based techniques and ask for vector-based patterns. As a solution, our previous work has proposed a new matrix-pattern-oriented learning machine named MatMHKS for short [6,27] and successfully applied this new method to some matrix-based SVMs including MatLSSVM [24] and KMatLSSVM [28]. MatMHKS follows a classical regularization framework for boosting generalization [22,17] as follows:

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

In this framework, R_{emp} is the empirical risk term considering an input pattern with its original matrix representation, and the regularization term R_{reg} tries to penalize the roughness or smoothness of the whole framework. The coefficient $C \geq 0$ controls the trade-off between R_{emp} and R_{reg}. MatMHKS is usually effective and efficient except for one problem: R_{emp} only considers the relationship of the elements in the same row or column of each matrix pattern while

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

R_{reg} is not able to preserve any local information, either. In a word, MatMHKS just considers the structural information of each individual pattern but fails to fully explore the structure information of the whole space. Therefore, the first motivation of this paper is to find a way to capture both global and local knowledge from the original domain and boost MatMHKS.

Similarly, the study on globality and locality in structural learning experiences the shift of focus from vector-based ideas to matrix-based ones. Firstly, the structural learning pays main attention to the vector-based methods of feature extraction or selection in early days, e.g., the Locally Discriminating Projection (LDP) constructing both the local and class information to solve the ignorance of the label information caused by the original Locality Preserving Projection (LPP) [37] and the Local-Global Information Combination (LGIC) scheme boosting the traditional Finger Knuckle Print (FKP) through adopting the local feature coded from the orientation information extracted by the Gabor filters and the global feature obtained from the Fourier transform coefficients of the image [35]. Moreover, a new method tries to capture two directional projections corresponding to extraction of vertical and horizontal local face image features and optimizes such information through the extreme learning machine [19]. Furthermore, the Local and Global Structure Preserving (LGFS) and its extended form E-LGFS are both proposed to overcome the ignorance of the global structure of patterns and the high redundancy of the selected features caused by Laplacian Score [20]. The LGFS uses both nearest neighborhood graph and farthest neighborhood graph to describe the underlying local and global structural information while the E-LGFS takes advantage of normalized mutual information to measure the dependency between a pair of features [20]. Secondly, the matrix-oriented method using structural information in feature extraction has been widely developed in recent years. To improve the traditional Two-Dimensional LPP (2DLPP), the Two-Dimensional Supervised Local Similarity and Diversity Projection (2DSLSDP) is proposed through defining two novel weighted adjacency graphs named similarity graph and diversity graph, respectively [9]. An extension of graph-based image feature extraction method named Sparse Two-Dimensional Locality Discriminant Projections (S2DLDP) is proposed to reduce the high computational cost of the traditional 2DLGEDA and 2DDLPP [15]. Thirdly, both the vector- and matrix-based structural learning for classifier design gain their popularity in recent decade. After discussing both the superiority and inferiority between the Single View Learning (SVL) and the Multiple View Learning (MVL), Wang et al. [26] find the difficulty in getting multi-view information from single view pattern in reality and thus propose a novel MVL based on MHKS named MultiV-MHKS. On the other side, Mehmet et al. [10] try to introduce the localized information into multi-kernel learning based on the Support Vector Machine (SVM). Furthermore, Wang et al. [25] reconstruct the regularization term R_{reg} in Eq. (1) to make the whole framework spatially smooth, which in a sense considers the problem of globality and locality. In spite of the comprehensive studies in both feature extraction/selection or classifier design, it seems that introducing both global and local information of the original domain into a matrix-based learning machine has never been discussed before. Inspired by the previous researches, the second motivation of this paper is to develop such machine which not only inherits the advantages of the original matrix-based one that considers the individual matrix structure, but also incorporates the globality and the locality of the whole space consisting of all the matrix patterns.

Consequently, this paper proposes a new regularization term R_{gl} taking advantages of a technique named Alternative Robust Local Embedding (ARLE) [30], not only to solve the inherent flaw of MatMHKS mentioned above, but also to make a trial to combine the global and local measure into a matrix-based learning machine.

The proposed R_{gl} might introduce a natural measure to characterize the whole data distribution in the matrix case and expected to capture the global and local structural information simultaneously. Subsequently, we boost the original framework (1) by adding a new regularization term R_{gl} and thus get the following function:

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

In practice, the proposed learning framework (2) is implemented through adopting our previous work MatMHKS as the basic building block, and denoted as GLMatMHKS for short. The experimental results on a broad range of data validate the feasibility and effectiveness of the proposed GLMatMHKS.

The main contributions of this paper can be highlighted as follows:

- To our knowledge, this paper *first* introduces both the globality and the locality of the whole data space into the matrixized classifier design through designing a new regularization term R_{gl}, which is expected to represent the globality and the locality of the whole data domain.
- This paper extends the existing matrix-pattern-oriented learning techniques so as to make them as a special example of the proposed framework.
- This paper demonstrates that the proposed GLMatMHKS has a tighter generalization risk bound than MatMHKS in terms of the Rademacher complexity due to the introduction of the designed term R_{gl}.

The rest of this paper is organized as follows. Section 2 presents a brief introduction on the preliminary knowledge of MatMHKS and ARLE. Section 3 demonstrates the architecture of the proposed globalized and localized matrix-pattern-oriented classification machine. Section 4 reports on all the experimental results. Section 5 provides a theoretical and experimental discussion between the proposed method and the other two methods belonging to the Ho-Kashyap paradigm in terms of the Rademacher complexity. Finally, the conclusion and future work are given in Section 6.

2. Related work

In this section, a concise introduction of the two preliminary method MatMHKS and ARLE used in the proposed algorithm are demonstrated.

2.1. MatMHKS

Since the proposed algorithm is improved from our previous work MatMHKS, in this subsection some typical classifiers belonging to the paradigm of the Ho-Kashyap algorithm (HK) [11] including the vectorized Modified Ho-Kashyap algorithm (MHKS) [17] and the matrixized MatMHKS [6] are given a brief introduction.

The original binary-class classification problem is given with N samples ($\boldsymbol{x}_p$, φ_p), $p = 1, \ldots, N$, where $\boldsymbol{x}_p \in \boldsymbol{R}^d$, and the class label $\varphi_p \in \{+1, -1\}$. The objective criterion derived from the original HK should satisfy the following discrimination function when the problem is linearly separable:

$$g(\boldsymbol{x}_p) = \tilde{v}^T \boldsymbol{x}_p + v_0 \begin{cases} > 0, & \text{if} \quad \varphi_p = +1 \\ < 0, & \text{if} \quad \varphi_p = -1 \end{cases}, \tag{3}$$

In this equation, $\tilde{v} \in \boldsymbol{R}^d$ is a weight vector and $v_0 \in \boldsymbol{R}$ is a bias value. Further, we can rewrite this equation into an equation:

$$\boldsymbol{Yv} > \boldsymbol{0}_{N \times 1}, \tag{4}$$