Pattern Recognition 44 (2011) 2395–2413

Contents lists available at ScienceDirect

Pattern Recognition

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

A novel multi-view learning developed from single-view patterns

Zhe Wang [a,b], Songcan Chen [b,*], Daqi Gao [a]

[a] Department of Computer Science and Engineering, East China University of Science and Technology, Shanghai 200237, PR China
[b] Department of Computer Science and Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, PR China

ARTICLE INFO

Article history:
Received 3 September 2009
Received in revised form
6 January 2011
Accepted 5 April 2011
Available online 14 April 2011

Keywords:
Multi-view learning
Classifier design
Rademacher complexity
Ensemble learning
Ho–Kashyap classifier
Regularization learning
Pattern recognition

ABSTRACT

The existing multi-view learning (MVL) learns how to process patterns with multiple information sources. In generalization this MVL is proven to have a significant advantage over the usual single-view learning (SVL). However, in most real-world cases we only have single source patterns to which the existing MVL is unable to be directly applied. This paper aims to develop a new MVL technique for single source patterns. To this end, we first reshape the original vector representation of single source patterns into multiple matrix representations. In doing so, we can change the original architecture of a given base classifier into different sub-ones. Each newly generated sub-classifier can classify the patterns represented with the matrix. Here each sub-classifier is taken as one view of the original base classifier. As a result, a set of sub-classifiers with different views are come into being. Then, one joint rather than separated learning process for the multi-view sub-classifiers is developed. In practice, the original base classifier employs the vector-pattern-oriented Ho–Kashyap classifier with regularization learning (called MHKS) as a paradigm which is not limited to MHKS. Thus, the proposed joint multi-view learning is named as MultiV-MHKS. Finally, the feasibility and effectiveness of the proposed MultiV-MHKS is demonstrated by the experimental results on benchmark data sets. More importantly, we have demonstrated that the proposed multi-view approach generally has a tighter generalization risk bound than its single-view one in terms of the Rademacher complexity analysis.

1. Introduction

It is well-known that it is important to integrate the prior knowledge of dealt patterns in designing classifiers [8]. In practice, patterns can generally be obtained from single or multiple information sources. If each information source is taken as one view, accordingly there are two kinds of patterns, i.e. single-view patterns and multi-view patterns.[1] Correspondingly, the learning based on single-view and multi-view patterns can be called as single-view learning (SVL) and multi-view learning (MVL), respectively. It has been proven that co-training as one typical MVL approach has a superior generalization ability to SVL [9]. Co-training learns on both labeled and unlabeled pattern sets. Both labeled and unlabeled patterns are composed of two naturally split attribute sets. Each attribute set is called one view of the patterns. In implementation, co-training algorithm requires that the two views given the class labels are conditionally independent. The independence assumption is guaranteed by the patterns composed of two naturally split attribute sets.

In this paper, we expand the existing MVL to single-view patterns and thus develop a novel MVL framework, whose underlying motivations are:

- It is known that patterns can be sorted into single-view patterns and multi-view patterns according to the number M of information sources [9–11]. However, in most real-world applications there are usually only single-view patterns available since the M has to be one. In that case, the existing MVL framework cannot effectively work since there is not any natural way to partition the attribute space [8,10–12]. Therefore, this fact motivates us to develop a new MVL framework. The new MVL is expected to create multiple different views from single-view patterns and then to learn on the generated views simultaneously.
- In the existing MVL framework, multi-view patterns are represented by multiple independent sets of attributes. Its base algorithms have the same architecture in each view so as to iteratively bootstrap each other. Here, we expect to utilize the multi-view technique due to its superior generalization to the SVL. However, different from the exist MVL on multi-view patterns, we give a new multi-view viewpoint for a given base classifier on single-view patterns. Concretely, we change the original architecture of the given base classifier and thus obtain a set of sub-classifiers with different architectures from

* Corresponding author.
E-mail address: s.chen@nuaa.edu.cn (S. Chen).

[1] Each information source can induce one attribute set for patterns. Thus, single-view patterns are generally composed of one single attribute set and multi-view patterns are generally composed of multiple attribute sets.

0031-3203/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2011.04.002

each other. Each derived sub-classifier can be taken as one view of the original base classifier, which forms a set of sub-classifiers with multiple views. For all the derived sub-classifiers, we further adopt a joint rather than separated learning process. Therefore, one new learning algorithm is developed for these multi-view sub-classifiers. It is minimized for the disagreement between the outputs of each derived classifier on the same patterns.

In practice, we select the vector-pattern-oriented linear classifier as the so-discussed base classifier. Before being classified, any pattern whatever form it originally is, should be transformed into a vector representation in the vectorial case [33]. However, it is not always efficient to construct a vector-pattern-oriented classifier since the vectorization for patterns such as images might lead to a high computation and a loss of spatial information [21,23,26,34,40]. For overcoming the disadvantage, we proposed a matrix-pattern-oriented Ho–Kashyap classifier named MatMHKS [21,40] in the previous work. MatMHKS is a matrixized version of the vectorial Ho–Kashyap classifier with regularization learning (namely MHKS) [20]. The literature [21,23,34,40] has demonstrated the significant advantages of the matrixized classifier design in terms of both classification and computational performance.

The discriminant function of the vectorial MHKS is given as

$$g(x)=\tilde{\omega}^T x+\omega_0, \tag{1}$$

where $x\in\mathbb{R}^d$ is a vector pattern, $\tilde{\omega}\in\mathbb{R}^d$ is a weight vector, and $\omega_0\in\mathbb{R}$ is a bias. Correspondingly, the discriminant function of MatMHKS is given as

$$g(A)=u^T A\tilde{v}+v_0, \tag{2}$$

where $A\in\mathbb{R}^{m\times n}$ is a matrix pattern, $u\in\mathbb{R}^m$ and $\tilde{v}\in\mathbb{R}^n$ are the two weight vectors, and $v_0\in\mathbb{R}$ is a bias. It is found that for a given pattern, there can be one vector-form representation in the formulation (1) but multiple matrix-form representations with different dimensional sizes for the m and n in the formulation (33). In other words, there are multiple ways for reshaping the vector to the matrix. For instance, a vector $x=[1,2,3,4,5,6,7,8]^T$ could be assembled into two different matrices:

$$\begin{bmatrix}1&3&5&7\\2&4&6&8\end{bmatrix} \quad \text{and} \quad \begin{bmatrix}1&2&3&4\\5&6&7&8\end{bmatrix}^T.$$

Consequently, only one MHKS can be created for classifying the given pattern x. In contrast, multiple MatMHKSs can be created for the same task due to multiple reshaping ways from a vector to a matrix. Therefore, for the same classification problem, the solution set $\{\tilde{\omega},\omega_0\}$ of single MHKS corresponds to the solution sets $\{u^p,\tilde{v}^p,v_0^p\}_{p=1}^M$ of multiple MatMHKSs, where the weight vector sets $\{u^p,\tilde{v}^p\}_{p=1}^M$ are different from each other in terms of the dimensional size but can share a common discriminant function form $g(A)=u^T A\tilde{v}+v_0$. Here, MHKS is viewed as the base classifier. Each MatMHKS is taken as one view of the base MHKS. Our previous work [21] has validated that each MatMHKS provides one hypothesis and exhibits one representation of the original pattern. Thus multiple MatMHKSs can provide a complementarity for each other in classification due to their different representations for patterns. In order to achieve the complementarity, we syncretize the learning processing of multiple MatMHKSs into one single processing. In this case, each MatMHKS is expected to correctly classify one given pattern with the same attributes. Meanwhile it should be minimized for the disagreement between the outputs of all MatMHKSs. As a result, the single learning process is produced and one multi-view-combined classifier named MultiV-MHKS is proposed. Through the Rademacher complexity analysis, we demonstrate that the proposed multi-view MultiV-MHKS has a tighter generalization risk bound compared with the single-view MHKS.

- The proposed MultiV-MHKS algorithm is a nice way to solve the view selection problem of MatMHKS [21]. In MatMHKS, it is always a problem to select the best right matrix-form reshaped from a given vector pattern. This paper suggests one way to bypass it through choosing all the relevant ones and optimizing over them jointly. It is known that from a vector pattern as the input of MHKS to a matrix as the input of MatMHKS, the classification performance of MatMHKS relies on the different reshaping or matrixization ways [21,40]. In the processing of matrixizing a vector, different reshaping ways can induce multiple matrix patterns with different dimensional sizes of the row and column. Consequently, different reshaping ways result in different classification performances of MatMHKSs on the same vector patterns. Then for the best performance, we have to make a choice in multiple reshaping ways with the cross-validation technique at the cost of high computation [21]. Since the proposed MultiV-MHKS here simultaneously considers multiple MatMHKSs with multiple matrices, the choice of matrixizing ways could be avoided to great extent.
- The proposed MultiV-MHKS algorithm adopts the data representation in multiple views different from the other main strategies for creating good ensembles of classifiers: sampling either pattern sets or attribute (interchangeably feature) sets [13,14,48]. Compared with sampling pattern sets or feature sets, the proposed multi-view classifier design provides an alternative novel approach of producing multiple data sets for base learners, i.e. reshaping a vector pattern to different matrix ones with the same full features. In this case, the proposed multi-view classifier has the advantages in terms of the actual number of the unique samples, the size of the feature set and the representations, which brings up the superior performance of the proposed MVL here. In addition, different from the strategy of sampling pattern sets or feature sets, the proposed MVL employs a joint optimization rather than a separate learning in the training processing. To the best of our knowledge, it is novel for the proposed strategy of generating multiple training data sets on the base classifier. The implemented experimental results here have also shown that the proposed classifier MultiV-MHKS algorithm has a superior classification performance to the other strategies of ensembles.

We highlight the contributions of this paper as follows:

- Significance: This paper introduces the creation of multiple views from a single view for multi-view learning. It is known that though the existing MVL has been shown effective in the literature [8,10–12], it still relies heavily on the naturally separating the feature set into two independent components. In many settings, there might not be any natural way to partition the feature space, and thus the existing MVL framework might not be applicable. In such a scenario, the proposed approach suggested in this paper can potentially create multiple independent or at least weaker correlated views from a single view and then learn from the generated multiple views simultaneously.
- Novelty in the two aspects: In the first aspect, the learning approach proposed in this paper is different from the existing multi-view learning approach. Instead of the classifiers trained on two views iteratively boot-strapping each other, this paper proposes a joint learning approach that minimizes the