Information Sciences 295 (2015) 196–220

Contents lists available at ScienceDirect

Information Sciences

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

Double-fold localized multiple matrixized learning machine

Changming Zhu [a], Zhe Wang [a,b,c,*], Daqi Gao [a,*], Xiang Feng [a]

[a] Department of Computer Science & Engineering, East China University of Science & Technology, Shanghai 200237, PR China
[b] Shanghai Key Laboratory of Intelligent Information Processing, Shanghai 200433, PR China
[c] Provincial Key Laboratory for Computer Information Processing Technology, Soochow University, Suzhou 215006, PR China

ARTICLE INFO

Article history:
Received 29 November 2012
Received in revised form 15 May 2014
Accepted 5 October 2014
Available online 22 October 2014

Keywords:
Localized learning
Matrixized learning
Nonlinear learning
Gating model
Pattern representation

ABSTRACT

In this paper, we develop an effective multiple-matrixized learning machine named Double-fold Localized Multiple Matrixized Learning Machine (DLMMLM). The characteristic of the proposed DLMMLM is that it possesses double folds of local information from data. The first fold lies in the whole representation space which consists of different matrix representations. It is known that each pattern can be represented by different matrix representations. The matrices have their respective representation information and can play different discriminant roles in the final classification. Therefore from the viewpoint of the whole representation space, each matrix has its own local information. The second fold is that in each matrix representation learning, different patterns represented with the same matrix representation can carry different information. Therefore in the pattern space with the same matrix size, local information of different patterns should be introduced into the classifier design. On the whole, the advantages of the proposed DLMMLM are: (i) establishing a pattern-depended function in the matrixized learning so as to realize different roles of patterns for the first time; (ii) adopting the double-fold local information in both the representation space and the pattern space; (iii) proposing a new nonlinear classifier that is different from the state-of-the-art kernelization one; and (iv) getting a tighter empirical generalization risk bound in terms of the Rademacher complexity and thus achieving a statistically superior classification performance than those classifiers without the introduction of the double-fold local information.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

It is necessary to choose an appropriate representation for patterns in terms of pattern classification [11]. In statistical pattern classification, a pattern is generally represented by a point in a d-dimensional space [3,11,27]. Such a representation is viewed as vector representation and can bring a convenience in mathematics. Patterns with vector representation are called vector patterns. The classifier design which is based on vector patterns is called Vector-pattern-oriented Classifier (VecC) or vectorized learning machine. When vectorized learning machines deal with patterns with matrix representation, these patterns have to be vectorized. Patterns with matrix representation are called matrix patterns and images are classical matrix patterns. However, such a procedure brings three potential problems [23,34,43]. First one is some implicit structural or local contextual information of these matrix patterns will be lost after vectorization. The lost information is useful for

* Corresponding authors at: Department of Computer Science & Engineering, East China University of Science & Technology, Shanghai, 200237, PR China (Z. Wang).

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

http://dx.doi.org/10.1016/j.ins.2014.10.024
0020-0255/© 2014 Elsevier Inc. All rights reserved.

classification. Second one is vectorizing a matrix pattern increases the dimension of pattern so that a large memory is required. Third one is the dimension of vectorized pattern is very high, at the same time the number of these patterns is small, then the corresponding classifier trained from these patterns is easily overtrained. To solve these problems, Matrix-pattern-oriented Classifier (MatC), i.e., matrixized learning machine, has been developed. The matrixized learning machine is based on matrix patterns. Moreover, a vector or matrix pattern can be reshaped to a new one by some certain reshaping ways [6]. It has been demonstrated that there are two advantages for the classifier design based on matrix patterns without a vectorization preprocessing. One is reducing the computational complexity and the other is improving the classification performance [7,23,43,44]. Based on the above advantages, some researchers have developed Matrix-pattern-oriented Ho–Kashyap classifier with regularization learning (MatMHKS) [6], New Least Squares Support Vector Classification based on matrix patterns (MatLSSVC) [36], and One-Class Support Vector Machines based on matrix patterns (OCSVM) [42].

On the other hand, it can be found that the classification performances of the matrixized learning machines are matrixization-dependent. The corresponding experiments have shown their performances heavily rely on the reshaping ways which reshape the original vector or matrix pattern into a new matrix pattern. It is difficult to determine which reshaping way is the best one [36]. To overcome this problem, some researchers have provided some solutions. For example, AdaMatLSSVC [35] is an ensemble-based strategy constructed by multiple MatLSSVCs with different reshaping ways. In addition, both the multi-view learning developed from single-view patterns with Ho–Kashyap linear classification strategy (MultiV-MHKS) [37] and the regularized multi-view machine based on the response surface technique (RMultiV-MHKS) [38] fuse multiple matrix representations into a joint learning machine, respectively. In the joint learning machine, the classifier derived from a corresponding matrix representation is named as a matrixized sub-classifier. The generation process of a matrixized sub-classifier is also named as a matrix representation learning. Both MultiV-MHKS and RMultiV-MHKS are multiple-matrixized learning machines since they can combine multiple matrixized sub-classifiers into a joint matrixized learning machine. But MatMHKS is a single-matrixized learning machine which is based on one matrix representation. It has been validated that the multiple-matrixized learning machine has a superior performance than the corresponding single one [37,38].

While all previous solutions have a serious shortcoming that the local information from data is not fully used. Actually, most real-world data sets possess local structures which have local information that can be brought into account to achieve better performance for a learning machine. Recently, some scholars have made full use of the local information and proposed the solutions with local or localized properties. These solutions are named as local learning machines. For example, Tang et al. [32] have developed a Local Learning-based classifier (LL) which is effective in single-image super-resolution problem and Wang et al. [40] have developed a threefold structured classifier design based on matrix patterns (TSMHKS). Both of them extract local information from data by dividing the whole data set into multiple meaningful clusters. Here, each cluster has a high intra-cluster similarity and low inter-cluster similarity. Structures of patterns in these clusters are named as local structures and these local structures play important roles in the classification performances. Then Normalized Cut-based Graph Partitioning (NCGP) [31] resorts to the image segmentation technique to partition an image so as to get some local structures of image pixels. Moreover, with the usage of localized generalization errors for patterns located within neighborhoods of testing patterns, Multiple Classifier Systems (MCSs) [46] and a Radial Basis Function Network learning (RBFN) [45] have been developed to improve the performances of classifiers. The structures of the neighborhoods of a testing pattern can be treated as local structures. Further, Wang et al. [39] have proposed Multiple Localized Empirical Kernel Learning (MLEKL) which adopts multiple Empirical Kernel Mappings (EKMs) to map the data set into multiple feature spaces. Then one can use a linearly separable algorithm in each feature space to generate a sub-classifier and combine multiple generated sub-classifiers in together so that a high-performance joint learning machine is formed. In such a learning machine, MLEKL assigns the weight to each sub-classifier according to the structures of patterns in each corresponding feature space. The structure is also treated as the local structure. Here, it has been found that these local learning machines have two disadvantages. First disadvantage is ineffective local structures because of unfeasible clusters and image segmentation techniques. Second disadvantage is that EKM needs to be explicitly given the form of mapping. Doing so will cause higher time and space complexities. Therefore, we focus on local information from data especially in the process of multiple-matrixized learning machine and propose a new local learning machine. This local learning machine should avoid the previous two disadvantages.

Indeed, there are double folds of local information for a data set. It is known that in a joint learning machine, a matrixized sub-classifier is derived from a corresponding matrix representation. Different matrixized sub-classifiers play different discriminant roles in the final classification. It is found that if a matrix representation offers more useful representation information and discriminant role, the corresponding matrixized sub-classifier will show greater influence on the final classification. Further, we should assign a larger weight to this matrixized sub-classifier. As each pattern can be represented by different matrix representations, we treat these matrices as a whole representation space. From the viewpoint of the whole representation space, representation information and discriminant role of each matrix are named as local information and viewed as the first fold of local information from data. In terms of the second fold, we know in each matrix representation learning, different patterns represented with the same matrix representation can carry different information and they play different roles in the corresponding matrixized sub-classifier design. Moreover, a pattern has different information and roles in the different matrix representation learning processes. Therefore in the pattern space with the same matrix size, information of different patterns is named as local information. In other words, in each matrix representation learning, information of these different patterns is viewed as the second fold of local information from data.

Download English Version:

https://daneshyari.com/en/article/392168

Download Persian Version:

https://daneshyari.com/article/392168

Daneshyari.com