Knowledge-Based Systems 83 (2015) 13–31

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Multiple Matrix Learning Machine with Five Aspects of Pattern Information

Changming Zhu [a,b,*], Daqi Gao [b,*]

[a] College of Information Engineering, Shanghai Maritime University, Shanghai 201306, PR China
[b] Department of Computer Science & Engineering, East China University of Science & Technology, Shanghai 200237, PR China

ARTICLE INFO

Article history:
Received 19 September 2014
Received in revised form 1 March 2015
Accepted 2 March 2015
Available online 11 March 2015

Keywords:
Cluster structure
Gating model
Multiple matrix learning machine
Pattern information
Matrix representation

ABSTRACT

This paper proposes an effective Multiple Matrix Learning Machine with Five Aspects of Pattern Information (MMFI). First aspect lies in the class label of each training or validation pattern. Second aspect is the values of components for each pattern. Third aspect is the relationship between patterns in the local regions of input space. Fourth aspect is the representation information and discriminant roles of different matrix representations for patterns. Fifth aspect is the information of patterns in each matrix representation learning. The innovations of the proposed MMFI are: (1) establishing a pattern-dependent function in the matrix learning so as to realize different roles of patterns for the first time; (2) adopting five aspects of pattern information so that a more feasible learning machine can be trained. The advantages of MMFI are: (1) proposing a new nonlinear learning machine which is different from the state-of-the-art kernelization one; (2) achieving a statistically superior classification performance than those learning machines without the introduction of five aspects of pattern information; (3) possessing a lower or comparable computational-complexity than other compared multiple matrix learning machines.

1. Introduction

1.1. Background of the problem

It is necessary to choose an appropriate representation for patterns in terms of pattern classification [1]. In statistical pattern classification, a pattern is represented by a point in a d-dimensional space. Such a representation is viewed as vector representation which brings a convenience in mathematics. Some learning machines including Ho–Kashyap (HK) algorithm [1] and the modification of Ho–Kashyap algorithm with squared approximation of the misclassification errors (MHKS) [2] have validated that. Patterns with vector representation are named vector patterns. A learning machine which is designed on the base of vector patterns is called vector-pattern-oriented classifier (VecC) or vector learning machine. When vector learning machines process patterns with matrix representation, i.e., matrix patterns, these patterns have to be vectorized. However, such a process brings three potential problems. One is the loss of implicit structural or contextual information of these matrix patterns. Another is the need of a large memory. The third is a high risk of over-training for a learning machine [3–5]. To solve these problems, matrix-pattern-oriented classifier (MatC), i.e., matrix learning machine, has been developed. A matrix 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. It has been demonstrated that a matrix learning machine has two advantages. One is reducing the computational-complexity and the other is improving the classification performance [4–7]. Based on the above advantages, matrix-pattern-oriented Ho–Kashyap classifier with regularization learning (MatMHKS) [8], new least squares support vector classification based on matrix patterns (MatLSSVC) [9], and one-class support vector machines based on matrix patterns (OCSVM) [10] have been proposed.

On the other hand, it can be found that classification performances of matrix learning machines are matrix-dependent, i.e., their performances heavily rely on the reshaping ways that reshape the original vector or matrix pattern into a new matrix pattern [8]. It is difficult to determine the best reshaping way. To overcome this issue, some solutions are developed. For example, both multi-view learning developed from single-view patterns with Ho–Kashyap linear classification method (MVMHKS) [11]

* Corresponding authors at: College of Information Engineering, Shanghai Maritime University, Shanghai 201306, PR China (Changming Zhu); Department of Computer Science & Engineering, East China University of Science & Technology, Shanghai 200237, PR China (Daqi Gao).

E-mail addresses: zcm19880301@163.com (C. Zhu), gaodaqihaha@163.com (D. Gao).

http://dx.doi.org/10.1016/j.knosys.2015.03.004
0950-7051/

and regularized multi-view machine based on response surface technique (RMVMHKS) [12] fuse multiple matrix representations into a joint learning machine. In the joint learning machine, the learning machine derived from a corresponding matrix representation is named a matrix sub-learning machine. The generation process of a matrix sub-learning machine is also named a matrix representation learning. Both MVMHKS and RMVMHKS are multiple matrix learning machines since they can combine multiple matrix sub-learning machines into a joint matrix learning machine. But MatMHKS is a single matrix learning machine which is based on one matrix representation. It has been validated that a multiple matrix learning machine has a superior performance than a single one [11,12].

But all previous solutions have a serious shortcoming that five aspects of pattern information are not fully used. Actually, most real-world data sets possess five aspects of pattern information which can bring a better performance to a learning machine.

1.2. Five aspects of pattern information

Since most real-world data sets possess five aspects of pattern information, we give the details of them below.

First aspect lies in the class label of each training or validation pattern. As we know, a data set consists of label-known patterns and label-unknown patterns. Label-known patterns are used to train a learning machine and label-unknown patterns are used to test the effectiveness of a learning machine. In terms of training a learning machine, class label of each training or validation pattern is a kind of important information that reflects the belonging of a pattern.

Second aspect lies in the values of components for each pattern. Each pattern consists of multiple components. For example, there is a pattern $x_i \in \mathbb{R}^d$ and $x_i = \{x_{i1}, x_{i2}, \ldots, x_{id}\}$ where x_{ij} represents j-th component of x_i, x_i is i-th pattern of a data set, and $j = 1, 2, \ldots, d$. Then x_{ij} gives information of x_i in j-th field and different components give information of x_i in different fields. This kind of information is also important to design a learning machine.

Third aspect derives from the relationship between patterns in the local regions of input space. It has been validated that in an input space, a data set can be divided into multiple subsets with a clustering approach used. Each subset is named a cluster or local region. These clusters have high intra-cluster similarities and low inter-cluster similarities. The relationship between patterns in a cluster is treated as the cluster structure. Some experiments have validated that information from cluster structures has a great influence on the design of a learning machine [13–15].

Fourth aspect is the representation information and discriminant roles of matrix representations for patterns. Each pattern has several different matrix representations. These matrix representations possess respective representation information and discriminant roles. Then, the corresponding matrix sub-learning machines play different roles on the final classification and their weights are also different. If a matrix representation has more useful representation information and discriminant role, the weight of its corresponding matrix sub-learning machine is larger.

Fifth aspect is the information of patterns in each matrix representation learning. In a matrix representation learning, although patterns are represented with a same matrix representation, they still carry different information and play different roles on the design of the corresponding matrix sub-learning machine. So their information is important to the design of a learning machine.

1.3. Raise of the proposed learning machine

According to the above five aspects of pattern information, it is found that in the process of RMVMHKS, one can assign feasible weights to matrix sub-learning machines according to the different representation information and discriminant roles of matrix representations. So that the fourth aspect of pattern information has been reflected. But the weight cannot reflect information of different patterns in each matrix representation learning. Therefore, RMVMHKS does not possess the fifth aspect of pattern information. As another multiple matrix learning machine, MVMHKS assigns a same weight to each matrix sub-learning machine so that representation information and discriminant role of each matrix representation are same. This weight cannot reflect information of patterns in each matrix representation learning either. Therefore, MVMHKS possesses neither the fourth nor the fifth aspects of pattern information. Furthermore, it is also found that HK, MHKS, MatMHKS, MVMHKS, and RMVMHKS do not take relationship between patterns in a same cluster into account and they just make use of the first and second aspects of pattern information.

So in order to make full use of these five aspects of pattern information, we use RMVMHKS as the basic learning machine. After that, we adopt a clustering approach and a pattern-dependent function which is a matrix gating model. By adopting the pattern-dependent function, the weight of a matrix sub-learning machine is dependent on not only the corresponding matrix representation but also the patterns themselves. Doing that, fourth and fifth aspects of pattern information can be reflected simultaneously. By adopting the clustering approach, the whole data set can be divided into multiple clusters and third aspect of pattern information can be reflected with used cluster structures. The proposed new learning machine is named Multiple Matrix Learning Machine with Five Aspects of Pattern Information (MMFI). We highlight the innovations and advantages of the proposed MMFI below.

1.3.1. Innovations of MMFI

(1) The proposed learning machine establishes a matrix gating model which is a pattern-dependent function in the matrix learning for the first time so as to realize different roles of patterns. Optimizing this matrix gating model is a new algorithm to assign appropriate weights to matrix sub-learning machines. These weights in the final classification are different, and they are dependent on the patterns. It means that these weights reflect two kinds of useful information. One is the different discriminant roles and representation information of matrix representations. The other is the information of patterns in each matrix representation learning. (2) The proposed learning machine adopts five aspects of pattern information. Although the proposed learning machine has only two more aspects of pattern information compared with RMVMHKS, it is the first time to introduce five aspects of pattern information into a matrix learning simultaneously.

1.3.2. Advantages of MMFI

(1) The proposed learning machine is a new nonlinear learning machine which introduces five aspects of pattern information. It is different from the state-of-the-art kernelization learning machine, i.e., support vector machine (SVM) [16]. In MMFI, we need not a kernel matrix to fuse pattern information as SVM does. What we just need are patterns themselves and their class labels. Furthermore, MMFI is designed on the base of patterns in the input space whereas the main notion of SVM is the mapping of kernel functions. For the differences of the design frameworks, the effective applicable scopes are not exactly same for these two learning machines. (2) The proposed learning machine MMFI gets a statistically superior classification performance than those other learning machines without the introduction of five aspects of pattern information. (3) MMFI possesses a lower or comparable computational-complexity than other compared multiple matrix learning machines.