Pattern Recognition 61 (2017) 47–65

Contents lists available at ScienceDirect

Pattern Recognition

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

Robust multi-source adaptation visual classification using supervised low-rank representation

JianWen Tao [a,*], Dawei Song [b], Shiting Wen [a], Wenjun Hu [c]

[a] School of Information Science and Engineering, Ningbo Institute of Technology, Zhejiang University, Ningbo 315100, China
[b] Department of Computing, The Open University, Milton Keynes, United Kingdom
[c] School of Information and Engineering, Huzhou Teachers College, Huzhou 313000, China

ARTICLE INFO

Article history:
Received 16 September 2014
Received in revised form
19 May 2016
Accepted 3 July 2016
Available online 5 July 2016

Keywords:
Multiple source domain adaptation
Transfer learning
Supervised low rank representation
Visual classification

ABSTRACT

How to guarantee the robustness of multi-source adaptation visual classification is an important challenge in current visual learning community. To this end, we address in this paper the problem of robust visual classification with few labeled samples from the target domain of interest by leveraging multiple prior source models. Motivated by the recent success of low rank representation, we formulate this problem as a robust multi-source adaptation visual classification (RMAVC) model with supervised low rank representation by combining the strength of discriminative information from the target domain and the prior models from multiple source domains. Specifically, we propose a joint supervised low rank representation and multi-source adaptation visual classification framework, which achieves dual goals of finding the most discriminative low rank representation and multi-source adaptation classifier parameters for the target domain. While it is showed in this paper that the proposed RMAVC framework is effective and can produce high accuracy on several tasks of multi-source adaptation visual classification, this framework fails to consider the local geometrical structure of the target data and the heterogeneousness among multiple source domains. Hence, under this framework, we further present two effective extensions or variants, i.e., RMAVCK and RMAVC_FM, by exploiting multiple kernel trick and flexible manifold regularization, respectively. The proposed framework and its variants are robust for classifying visual objects accurately and the experimental results demonstrate the effectiveness of our methods on several types of image and video datasets.

1. Introduction

With the development of computer network and storage technologies, we have witnessed explosive growth of visual data such as images and videos. These explosive amounts of emerging visual data have brought a big challenge to the task of visual classification by using those traditional methods [1–4]. On the one hand, expensive and time-consuming human labor is required for the collection of labels of new emerging data. On the other hand, there exists a large amount of outdated labeled data from previous tasks. To this end, the recent literatures have witnessed an increasing focus on transfer learning or domain adaptation learning (DAL) [7,8] problems in computer vision and machine learning [9–11,13]. Different with conventional classification methods, one (or multiple) related source/auxiliary domain(s) is(are) provided in domain adaptation classification to assist the learning process in target domain, and the distribution divergence or model disparity between source and target domains should be minimized in the process of adaptation [16]. Recently, many DAL techniques have been proposed to solve the problem of mismatch between source and target domains in the field of image or video concept detection and classification [12–15,17]. Depending on how the source information is exploited, the division of DAL methods is between model-based and representation-based adaptation. The former advocates implicit adaptation to the target distribution by adjusting a classifier from the source domain(s) (e.g., [17,18] and [21]), whereas the latter attempts to achieve alignment by adjusting the representation of the source data via learning a transformation (e.g. [13–16], and [19]). Orthogonal to this, the extant methods can also be classified into supervised and unsupervised adaptation, based on whether labels of target domain have been exploited during the adaptation. In this paper, we focus on the supervised model-based adaptation methods.

While the effectiveness and efficiency of cross-domain visual classification make it of particular use in practice, it also brings a new issue, i.e., how to handle the errors (e.g., noises or corruptions) which possibly exist in target training data. Theoretically speaking, noise and outlier samples abound in training data by

* Corresponding authors.
E-mail address: jianwen_tao@aliyun.com (J. Tao).

http://dx.doi.org/10.1016/j.patcog.2016.07.006
0031-3203/

nature [4,5], which needs more robustness on a learning model compared to traditional DAL tasks. Hence, how to guarantee the robustness of cross-domain visual recognition is an important challenge in robust visual classification tasks. With this viewpoint, we study in this paper the robust visual classification problem by exploiting multiple source models learned previously from existing related source domains [15,17,18,21,58]. Specifically, given a set of models previously learned from multiple prior sources, and a set of large unlabeled samples and few labeled samples from the target domain of interest, drawn from some different distribution with those from which the prior sources are respectively drawn, our objective is to learn jointly a robust representation for all samples from the target domain and a effective classifier relying on auxiliary multi-source models for further visual classification.

Notice that in visual recognition tasks, visual representation is a crucial procedure for robust visual data processing and understanding. Liu et al. [20] recently proposed the low-rank representation (LRR) and use it to recover robust subspace structures of data. LRR can jointly obtain the representation of all the data under a global low rank constraint, and thus it is better at capturing the global data structures of target data. While being potentially powerful, LRR depends heavily on the configuration of its key parameter [52]. In order to figure out under which conditions LRR can be effective, Liu et al. [52] further established a theoretical analysis for LRR and then derived an estimate to the model parameter of LRR. Recent researches have demonstrated that LRR is a powerful visual data representation method and is receiving increasing interest in image segmentation and classification and face recognition [5,22–27]. These algorithms first define the characterized low rank structure and then implement classification steps, as a result, the performance of classification steps is largely determined by the recovered low rank representation structure. The reason is that once the low rank representation of data is obtained, it is fixed in the following classification steps. If the low rank representation matrices and the optimal classification vectors can be jointly learned at the same time, the optimal low rank representation matrices can be affected by the classification vectors, therefore, the performance of the algorithm might be improved. Besides, to the best of our knowledge, few works presently have yet jointly considered supervised low-rank representation and multiple sources adaptation to solve robust visual classification problems. In other words, how to harness the model information from source domain(s) to effectively boost visual classification performance of target datum, which is reconstructed with LRR, is yet an unresolved issue.

Based on the above motivations, we propose in this paper a joint supervised low rank representation and multi-source adaptation visual classification framework, which achieves the dual goal of finding the most discriminative low rank representation and optimal visual classifier parameters for target visual recognition. The multi-source adaptation ensures that our framework is effective for exploiting model information learned from multiple sources, and the lowest rankness ensures that it can better capture the global cluster or subspace structures of target data and is more robust to noise and outliers. We therefore name our method Robust Multi-source Adaptation Visual Classification using supervised LRR or termed RMAVC for short. While the idea of model-based multi-source adaptation learning has been explored recently for computer vision primarily for video or image classification [17,18,21], our method is intrinsically different from these previously proposed algorithms for visual classification. In our approach, label information from training data is incorporated into low rank representation learning process by adding a ridged regression regularization term to the objective function of LRR. The dictionary learned by our method has good reconstruction and classification capabilities. With this high-quality dictionary, we are able to learn jointly a low rank representation and a robust classification model for target domain. To address this joint optimization objective of RMAVC, We also propose an effective algorithm, which alternates between solving for subsets of parameters of the classification whilst preserving the low-rank representation.

Specifically, this paper makes the following main contributions:

(1) We present a Robust Multi-source Adaptation Visual Classification (RMAVC) framework by using supervised low rank feature representation. The framework jointly optimize the supervised low rank representation with a recovered overcomplete dictionary and find an optimal target classifier model adapted from other multi-source models, in which the optimal low rank representation matrix can adaptively change with the optimal low rank classification vector. To the best of our knowledge, our work is the first to introduce the supervised low-rank representation for robust multi-source adaptation visual learning.
(2) RMAVC can be readily introduced to many popular kernel methods such as SVM, support vector regression (SVR), least-squares SVM (LS-SVM) [32] and so on, and extend these algorithms to the corresponding domain adaptation methods.
(3) We formulate an optimization problem that combines the objective function of the supervised low rank representation with the classification by leveraging multiple source models, and further give an algorithm that alternates between solving for optimal variables of the function, whilst preserving the low-rank representation.
(4) Under the RMAVC framework, we further propose two effective extensions, i.e., RMAVC with multiple kernel trick (RMAVCK) and RMAVC with flexible manifold regularization (RMAVC_FM). Based on the tool of transductive Rademacher complexity analysis [49], a generalization error bound is also derived for RMAVC_FM.
(5) Comprehensive experiments on real visual classification tasks verify the robustness and effectiveness of the proposed framework and its extensions.

The rest of the paper is organized as follows. In Section 2, the previous related works about LRR are briefly reviewed. We presented our robust multi-source adaptation visual classification framework and its corresponding solutions in Section 3, in which we further gives two effective extensions under this framework. Discussions about the generalization error bound of RMAVC_FM, selection of adaptation weights, and connections with other works are conducted in Section 4. The experimental results on visual data classification are discussed in Section 5. Finally, we draw a conclusion and discuss the future work in Section 6.

2. Briefly review on low rank representation methods

2.1. Notations and definitions

In the following, we denote with small and capital bold letters respectively column vectors and matrices, e.g., $a = [a_1, a_2, \ldots, a_d]^T \in \mathbb{R}^d$, and $\mathbf{A} \in \mathbb{R}^{N\times d}$. The (j, i) entry of $\mathbf{A}$ is represented by the symbol $[\mathbf{A}]_{ji}$, and therefore its jth row and ith column by $[\mathbf{A}]_{j,:}$ and $[\mathbf{A}]_{:,i}$, respectively. Moreover we indicate with $\|a\|_p := \left(\sum_{i=1}^{d} |a_i|^p\right)^{1/p}$ the p-norm of a vector a, and with $\|\mathbf{A}\|_F^2 = \sum_{i=1}^{n}\sum_{j=1}^{d}[\mathbf{A}]_{ij}^2$ the Frobenius norm of the matrix $\mathbf{A}$. We also denote by $\|\mathbf{A}\|_{2,1}$ the $l_{2,1}$ norm for sample-specific noise [20]: $\|\mathbf{A}\|_{2,1} = \sum_{j=1}^{n}\sqrt{\sum_{i=1}^{d}[\mathbf{A}]_{ji}^2}$. The transpose of a vector or matrix is