J. Vis. Commun. Image R. 33 (2015) 265–272

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Projected Transfer Sparse Coding for cross domain image representation ☆

Xiao Li, Min Fang *, Ju-Jie Zhang

School of Computer Science and Technology, Xidian University, 710071, China

ARTICLE INFO

Article history:
Received 30 July 2015
Accepted 28 September 2015
Available online 8 October 2015

Keywords:
Image representation
Domain adaptation
Sparse coding
Projection matrix
$L_{2,1}$ norm
Shared dictionary
Maximum Mean Discrepancy
KSVD

ABSTRACT

Sparse coding has been used for image representation successfully. However, when there is considerable variation between source and target domain, sparse coding cannot achieve satisfactory results. In this paper, we proposed a Projected Transfer Sparse Coding algorithm. In order to reduce their distribution difference, we project source and target data into a shared low dimensional space. Meanwhile, we learn a projection matrix and a shared dictionary and the sparse coding of source and target data in the low dimensional space. Unlike existing methods, the sparse representations are learnt using the projected data which are invariant to the distribution difference and the irrelevant samples. Thus, the sparse representations are robust and can improve the classification performance. We do not need to know any explicit correspondence across domains. We learn the projection matrix, the discriminative sparse representations, and the dictionary in a unified objective function. Our image representation method yields state-of-the-art results.

1. Introduction

Image representation is one of the major topics in computer vision, in which sparse coding has been successfully used and achieves outstanding performance. Sparse coding yields sparse representation so that the data is represented by a linear combination of a few dictionary elements. Olshausen et al. in [1] have pointed out that natural images can be well represented by sparse coding. Recently, sparse coding has been successfully used in computer vision problems, e.g., image classification [2,3], object recognition [4] and face recognition [5–7]. Sparse coding has been used for computer vision tasks in the following three aspects. Firstly, sparse coding has been used in the image feature extraction process of Bag-of-Word (BoW) model, boosting the performance of image classification [2]. Secondly, sparse coding can be used for image representations [8,9]. Sparse representations mean using a few dictionary items to represent images, making them easy to understand and interpret. Thirdly, for face recognition, training images are often chosen as a dictionary and test images are then represented by the dictionary and classified by assigning the class with the lowest reconstruction error. Thus, sparse coding can be used as a sparse representation-based classifier [5–7].

The merits of sparse coding are that samples can be well interpreted by the linear combination of a few dictionary elements. Besides, the sparse representations can capture the main information of images, thus resisting the noise to a certain degree.

For example, we aim to classify images obtained by a web camera. However, the labeled images are very limited to train a robust image classifier. It is expansive and complex to annotate the unlabeled images. We regard this dataset as target domain. Thankfully, we have a large amount of labeled images obtained by a digital SLR camera, namely source domain. So our goal is to leverage source domain images to help our target classification tasks. When the feature distribution of source domain differs greatly from that of target domain, directly applying the classifier or object models trained from source domain on target domain is impossible. Visual domains differ in image distributions greatly even if they contain images of the same categories. There are many factors for the phenomenon, such as scene, location, pose, viewing angle, background clutter. Transfer learning [8,10,11] has been widely studied to solve this problem, producing excellent results. In [12], a survey on transfer learning is presented.

Sparse coding methods may be not suitable when training and test data have different distributions. In recent years, many researchers have focused on the use of sparse coding in transfer learning. Source and target samples are combined to learn a shared dictionary and sparse representations [8,13]. They seek the sparse representations of all source and target samples. However, when

☆ This paper has been recommended for acceptance by M.T. Sun.

* Corresponding author.
E-mail address: mfang@mail.xidian.edu.cn (M. Fang).

http://dx.doi.org/10.1016/j.jvcir.2015.09.018
1047-3203/

source domain is very different from target domain, some source samples are not relevant to the target samples in the original space even using Minimum Mean Discrepancy (MMD) to reduce the distribution distance. In [14], Shekhar et al. proposed a method which jointly learns the projections of source and target data and the dictionary in the projected low dimensional space using the few labeled data in the target domain. Qiu et al. [15] used regression to adapt dictionaries. However, in practical applications, we possibly cannot get labels of the target data.

In this paper, we focus on the unsupervised domain adaptation problem where the target labels are unavailable. We project source and target data onto a low dimensional space. The projection matrix is restricted by Minimum Mean Discrepancy to reduce the distribution difference between source and target domain, which can help learn a compact shared dictionary. Some samples in the source domain may be irrelevant to the target domain even in the low dimensional shared space, so we restraint the source projection matrix according to their relevance to the target data using $L_{2,1}$ norm which can make the rows of matrix sparse, selecting the source samples. As a consequence, the irrelevant samples are discarded. We learn the projection matrix, the discriminative sparse representations of all the samples, and the dictionary in a unified objective function. Thus, the learned low dimensional representation can well represent the common structure of the data. Moreover, the dictionary is representative of the projected data from both domains and the sparse representations are very effective for representing them. We perform experiments on USPS–MNIST, MSRC–VOC2007, Office–Caltech256 dataset pairs which are under different distributions. The results demonstrate that our method performs better than other state-of-the-art methods.

This paper is organized as follows. Section 2 provides a brief review of related work. In Section 3, we develop a novel Projected Transfer Sparse Coding (PTSC) algorithm to deal with the issue of cross domain image representation problems. In Section 4, extensive experiments are conducted. In Sections 5, the conclusions and future work are presented.

2. Related work

Researchers have developed lots of algorithms to get sparse and efficient representations of images. It has received growing attention because of its advantage and promising performance for many computer vision applications [5,8,13]. Given input samples $Y = [y_1, \dots, y_n] \in R^{d\times n}$, the sparse coding $X \in R^{m\times n}$ and dictionary $D = [d_1, \dots, d_m] \in R^{d\times m}$ can be trained by solving the following problem

$$\begin{aligned}(D,X) = \min_{D,X} \|Y - DX\|_F^2 \\ s.t.\ \|x_i\|_0 \leqslant T\end{aligned} \tag{1}$$

where x_i is a column of X, and T the sparsity factor which means that every sparse coding item has fewer than T nonzero elements. The Frobenius norm is defined as $\|P\|_F = \sqrt{\sum_i^r \sum_j^m P_{ij}^2}$. There are many optimization methods for solving the sparse coding problem such as matching pursuit [16], orthogonal matching pursuit [17] and basic pursuit [18]. We can get the dictionary from the data by some optimizing algorithms such as KSVD [19] and MOD [20]. The KSVD algorithm is efficient and requires fewer iterations to converge, therefore, we use the KSVD algorithm in our experiment.

However, traditional sparse coding methods [5,19] did not consider the distribution mismatch, known as distribution difference, among the training and testing domains. Transfer learning aims to deal with the different distributions between source and target domains. A classical strategy aims to learn a new domain invariant feature representation by minimizing the distribution divergence. It is measured by the empirical Maximum Mean Discrepancy (MMD) [8,13,21–23]. Domain Transfer Multiple Kernel Learning [22] simultaneously minimizes the structural risk function of SVM and Maximum Mean Discrepancy (MMD) in kernel space. Duan et al. [23] proposed an Adaptive Multiple Kernel Learning (AMKL) to cope with the distribution mismatch between web and consumer video domain. Transfer Component Analysis (TCA), proposed in [10], focused on reducing the distance between the two marginal distributions while maximizing the data variance. SA [24] learns a linear transformation between the subspaces of the domains inducted by eigenvectors. Long et al. in [25] proposed a method to incorporate feature matching and instance reweighting by minimizing the MMD and $L_{2,1}$ norm with Principal Component Analysis to construct new representations that is robust to the distribution difference. In this paper, we also use $L_{2,1}$ norm to restrict source projection matrix to eliminate the irrelevant source samples.

Transfer Sparse Coding (TSC) algorithm [8] aims to construct sparse representations for the images by incorporating sparse coding, Maximum Mean Discrepancy (MMD) and graph Laplacian into a unified objective. Al-Shedivat et al. [13] presented a method unifying sparse representations, domain transfer and classification into an objective. They all learn the sparse representations in the original space which is different from our method that source and target data are projected onto a low dimensional space. Zhu and Shao [26] proposed a method which measures the cross domain divergence by constructing virtual correspondences across both domains through a transformation matrix. The method learns two dictionaries for source and target domain. Shekhar et al. [14] proposed a method to learn the projections of source and target data and a shared dictionary in the projected low dimensional space in a supervised way. Huang in [28] proposed a method to solve the cross domain image synthesis and recognition problems by jointly solving the coupled dictionary and the projection matrix. However, we focus on the unsupervised domain adaptation problem where the labels of target data are unavailable. Partially Shared Dictionary Learning [27] proposed by Ranjan jointly learns different dictionaries for source and target domains. The dictionary elements which are common across domains are then used to obtain the sparse representations which are used to train a classifier. PSDL learns different dictionaries for source and target domains in original space while we learn the dictionary and sparse coding in a common space and reweight source samples, simultaneously.

3. Projected Transfer Sparse Coding (PTSC) algorithm

Define source domain set $D_s = \{Y_s, H_s\}$, target domain set $D_t = \{Y_t\}$, in which $Y_s \in R^{d\times n_s}$ is source data, $H_s \in R^{n_s\times 1}$ is the label of source data and $Y_t \in R^{d\times n_t}$ is target data. n_s is the number of source samples, n_t target samples. Suppose $Y = [Y_s\ Y_t] \in R^{d\times n}$, $n = n_s + n_t$. Source and target samples come from different probability distributions. In this paper, we introduce a unified objective function to solve the cross domain image representation problems, which we call Projected Transfer Sparse Coding algorithm.

3.1. Sparse coding

We use the matrix $P \in R^{k\times d}$ to project source and target data onto the shared subspace, k being the dimension of the subspace. In order to get the dictionary $D \in R^{k\times p}$ and sparse coding $X \in R^{p\times n}$, the projected source and target data are reconstructed by minimizing the following reconstruction error.