Deep Transfer Metric Learning

Xuewen Yang

June 8 2018

Abstract

Conventional metric learning methods usually assume that the training and test samples are captured in similar scenarios so that their distributions are assumed to be the same. This assumption doesnt hold in many real visual recognition applications, especially when samples are captured across different datasets. In this paper, we propose a new deep transfer metric learning (DTML) method to learn a set of hierarchical nonlinear transformations for cross-domain visual recognition by transferring discriminative knowledge from the labeled source domain to the unlabeled target domain. Specifically, our DTML learns a deep metric network by maximizing the inter-class variations and minimizing the intra-class variations, and minimizing the distribution divergence between the source domain and the target domain at the top layer of the network. To better exploit the discriminative information from the source domain, we further develop a deeply supervised transfer metric learning (DSTML) method by including an additional objective on DTML where the output of both the hidden layers and the top layer are optimized jointly. Experimental results on cross-dataset face verification and person re-identification validate the effectiveness of the proposed methods.

1. Introduction

How to design a good similarity function plays an important role in many computer vision and pattern recognition tasks. Generally, the optimal similarity function for a given vision problem is task-specific because the underlying data distributions for different tasks are usually different. Recent advances in machine learning have shown that learning a distance metric directly from a set of training examples can usually achieve proposing performance than hand-crafted distance metrics [1] [6]. In recent years, a variety of metric learning algorithms have been proposed in the literature [3] [6], and some of them have successfully applied in visual analysis applications such as face recognition [3], image classification [6], human activity recognition [2], person re-identification [4] and visual search.

To this end, in this work, we propose a new deep transfer metric learning (DTML) method for cross-dataset visual recognition. Figure 1 illustrates the basic idea of the proposed method. Our method learns a set of hierarchical nonlinear transformations by transferring discriminative knowledge from the labeled source domain to the unlabeled target domain, under which the inter-class variations are maximized and the intra-class variations are minimized, and the distribution divergence between the source domain and the target domain at the top layer of the network is minimized, simultaneously. To better exploit the discriminative information from the source domain, we further develop a deeply supervised transfer metric learning (DSTML) method by including an additional objective on DTML where the output of both the hidden layers and the top layer are optimized jointly. Experimental results on cross-dataset face verification and person re-identification demonstrate the effectiveness of the proposed methods.

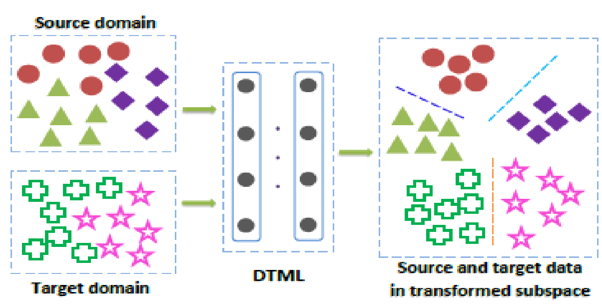

Figure 1. The basic idea of the proposed DTML method. For each sample in the training sets from the source domain and the target domain, we pass it to the developed deep neural network. We enforce two constraints on the outputs of all training samples at the top of the network: 1) the inter-class variations are maximized and the intra-class variations are minimized, and 2) the distribution divergence between the source domain and the target domain at the top layer of the network is minimized.

2. Related Work

2.1. Deep Learning

In recent years, deep learning has attracted much attention in computer vision and machine learning due to its superb performance in various tasks. Generally, deep learning aims to learn hierarchical feature representations directly from raw data. Recent advances have shown that deep learning have been successfully applied to many visual tasks such as image classification, object detection, action recognition [5], and face recognition. Many deep learning models have been proposed in recent years, and representative methods include deep convolutional neural networks, deep neural networks, deep stacked auto-encoder [5], deep belief networks, and deeply-supervised nets. However, most of them aim to learn feature representations via deep model rather than similarity measure. More recently, deep learning has also been used in metric learning, and several metric learning methods have been proposed. For example, Cai *et al.* introduced a nonlinear metric learning method using the stacked independent subspace analysis. Hu *et al.* proposed a discriminative deep metric learning method which employs a conventional neural network by enforcing a large margin criterion at the top layer of the network. While these methods have achieved reasonably good performance, they assume that the training and test samples are captured in the same environments, which is not always satisfied in many real applications. In this work, we proposed a deep transfer metric learning approach by learning a deep metric network and considering the distribution difference between the source domain and the target domain.

2.2. Transfer Learning

Transfer learning aims to address the problem when the distribution of the training data from the source domain is different from that of the target domain. Over the past decades, a variety of transfer learning algorithms have been proposed and they can be mainly categorized into two classes: instance-based and feature-based. For the first class, different weights are learned to rank the training samples in the source domain for better learning in the target domain. For the second class, a common feature space is usually learned which can transfer the information learned from the source domain to the target domain. In recent years, several transfer learning techniques have been presented and representative methods include domain transfer support vector machine, transfer dimensionality reduction, and transfer metric learning [7]. While some proposing results can be obtained by these transfer learning methods, most of them only consider minimizing the distribution difference between the source domain and the target domain by using linear mappings or the kernel trick, which are not effective enough to transfer the knowledge if the distribution difference is large and the transfer functions are usually not explicitly obtained. In this work, we borrow the idea of deep learning and propose a deep transfer metric learning method by learning a discriminative distance network with some information transferred from the source domain.

References

[1] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon. Information-theoretic metric learning. In *ICML*, 2007. 1

[2] T. Du and A. Sorokin. Human activity recognition with metric learning. In *ECCV*, 2008. 1

[3] J. Hu, J. Lu, and Y. P. Tan. Discriminative deep metric learning for face verification in the wild. In *CVPR*, 2014. 1

[4] M. Kostinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof. Large scale metric learning from equivalence constraints. In *CVPR*, 2012. 1

[5] Q. V. Le, W. Y. Zou, S. Y. Yeung, and A. Y. Ng. Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis. In *CVPR*, 2011. 2

[6] K. Q. Weinberger and L. K. Saul. Distance metric learning for large margin nearest neighbor classification. *JMLR*, 2009. 1

[7] Y. Zhang and D. Y. Yeung. Transfer metric learning with semi-supervised extension. *TIST*, 2012. 2