Engineering Applications of Artificial Intelligence 52 (2016) 119–125

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

LMDT: A weakly-supervised large-margin-domain-transfer for handwritten digit recognition

Hamidreza Hosseinzadeh *, Farbod Razzazi

Department of Electrical and Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran

ARTICLE INFO

Article history:
Received 20 June 2015
Received in revised form
10 January 2016
Accepted 26 February 2016
Available online 22 March 2016

Keywords:
Writer adaptation
Domain adaptation
Handwriting recognition
Transfer learning
Discriminant learning

ABSTRACT

Performance of handwritten character recognition systems degrades significantly when they are trained and tested on different databases. In this paper, we propose a novel large margin domain transfer algorithm, which is able to jointly reduce the data distribution mismatch of training (source) and test (target) datasets, as well as learning a target classifier by relying on a set of pre-learned classifiers with the labeled source data in addition to a few available target labels. The proposed method optimizes the combination coefficients of pre-learned classifiers to obtain the minimum mismatch between results on the source and target datasets. Our method is applicable both in semi-supervised and unsupervised domain adaptation scenarios, while most of the previous competing domain adaptation methods work only in semi-supervised scenario. Experiments on adaptation to different handwritten digit datasets demonstrate that this method achieves superior classification accuracy on target sets, comparing to the state of the art methods. Quantitative evaluation shows that an unsupervised adaptation reduces the error rates by 40.2% comparing with the SVM classifier trained by the labeled samples from the source domain.

1. Introduction

Supervised machine learning has already been widely studied and achieved significant success. When applying supervised machine learning methods to the classification or regression problems, it is typically assumed that the labeled training data (source) and the test data (target) are drawn from the same distribution which is far away from the realistic conditions. However, many real world applications, especially in handwriting recognition, challenge this assumption. When an existing training data is outdated, and the new labeled dataset is very small, or practically it is difficult to recollect a new training data, the classifiers should be learned on the old training data to infer primary models in the first stage. Then they should be adapted well to the new distribution of the test data in the test phase using a small number of the labeled samples from the target domain. In such cases, the trained model should be adapted to the test samples (Pan and Yang, 2010; Patel et al., 2015; Shao et al., 2015).

Writer adaptation and corpus adaptation are two main applications of domain adaptation in optical character recognition. To deal with writer adaptation, that is the handwriting style variations across different writers, a general (writer independent) classifier should be learned with large training data from many writers. The classifier can adapt toward a new handwriting style with the help of some writer dependent data (either labeled or unlabeled). This is known as writer adaptation (Zhang and Liu, 2013). As another application, the style of written text is highly dependent to the design of the data gathering form, the writers' community (age, educational skills, the time permitted for writing, environmental conditions and the importance of the written data for the writer). Therefore, in some cases, the training dataset characteristics are different from the test dataset. Our study is focused on the corpus adaptation. Although, we did not investigate our algorithm for writer adaptation in this paper, the method can be easily tuned for it.

Domain adaptation problem has been studied in two main scenarios: one is the semi-supervised domain adaptation scenario, where the target domain has few labeled data. The other is the unsupervised domain adaptation scenario that considers only unlabeled data for the target dataset to adapt the classifier. In both scenarios, the source is generally rich in labeled samples. Therefore, they are designed scenarios for evaluating the domain adaptation problem, not training on the source set which is extremely supervised. The approach of this paper can be employed in both unsupervised and semi-supervised domain adaptation scenarios.

A subset of common semi-supervised classifiers is based on label propagation over a graph, where nodes represent data points and edge weights measure their pairwise similarities. Well known methods are Gaussian-fields and Harmonic-Function (Zhu et al.,

* Corresponding author.
E-mail address: hr.hosseinzadeh@srbiau.ac.ir (H. Hosseinzadeh).

http://dx.doi.org/10.1016/j.engappai.2016.02.014
0952-1976/

2003), Local–Global Consistency (Zhou et al., 2004), and Manifold Regularization (Belkin et al., 2006) for this task. Despite of leaning on a strong theory, these methods, unfortunately, cannot label unseen data well (Zhu et al., 2005), because the whole graph should be reconstructed when each new samples is presented to the learning classifier.

Another group of methods have focused on adaptation of learned classifier parameters, using a few numbers of labeled data from the target. Tommasi and Caputo have presented a Naive Bayes Nearest Neighbor (NBNN)-based adaptation algorithm that learns a Mahalanobis metric for each class iteratively, while inducing a large margin to separate classes (Tommasi and Caputo, 2013). In another study, Cross-domain SVM (CD-SVM) measures a distance of source samples from the target domain to define a weight for each source training sample, and then retrains the SVM with the combination of target and source reweighted samples (Jiang et al., 2008). Several multiple kernel learning (MKL) methods have been proposed for solving domain adaptation problems. In (Duan et al., 2012a, 2012b), adaptive MKL has been used to learn a kernel function based on multiple types of kernels as well as a target classifier, in which the combination of kernels affect the target classifier. In (Guo and Wang, 2013), a domain adaptive input-output kernel learning (DA-IOKL) algorithm has been introduced, which simultaneously learns both the input and output kernels with a discriminative vector-valued decision function.

Domain adaptation has been considered by (Yang et al., 2007; Schweikert et al., 2009) when more than one source set is available. Yang et al. have proposed adaptive SVM (A-SVM) in which a target classifier is adapted from the some existing source classifiers which are trained with the labeled samples from the source domains (Yang et al., 2007). To achieve this purpose, a delta function $\Delta f(x)$ that has been learned by using the labeled samples from the target domain, has been added into the source classifier. Therefore, target decision function may be formulated as follows:

$$f^T(\boldsymbol{x}) = \sum_{k=1}^{K} \gamma_k f_k^{\,s}(\boldsymbol{x}) + \Delta f(\boldsymbol{x}) \tag{1}$$

where $\gamma_k \in [0, 1]$ is the weight of each source classifier f^s and $\sum_{k=1}^{K} \gamma_k = 1$ and K is the total number of source domains. In the experiments of (Yang et al., 2007), weights of all source classifier are considered as equal. Moreover, the authors have assumed that the target classifier is learned with only one kernel.

Schweikert et al. have presented a strategy for domain adaptation by a linear combination of source classifiers together with the target classifier (Schweikert et al., 2009). Similar to A-SVM, source classifiers and target classifier have been learned independently by using SVM with labeled training data from source dataset and labeled samples from the target, respectively. Then, the final classifier has the following form:

$$f(\boldsymbol{x}) = \gamma f^T(\boldsymbol{x}) + \frac{1-\gamma}{K} \sum_{k=1}^{K} f_k^{\,s}(\boldsymbol{x}) \tag{2}$$

where $\gamma \epsilon [0, 1]$ is a weight parameter to balance the two terms. This weight has been determined via grid search by optimizing the multiclass error on the target labeled samples.

Recently, in (Hoffman et al., 2014) a maximum margin domain transformation (MMDT) method was proposed to learn the transformation of features. Their major purpose is to jointly project the target data onto the correct side of source learned hyperplane as well as determining the classifier parameters.

Many techniques for the semi-supervised adaptation problem have been developed especially for text classification application. A common method is to treat the labeled training data of the source as the prior information. They estimate the target data model parameters under such prior distribution (Li and Bilmes, 2007). Other methods have aimed at linking between the source and target data distribution by feature space transformation (Daumé III, 2007; Li et al., 2014). In (Daumé III, 2007), Feature Replication (FR) has been used to map the feature space of both source and target datasets onto an augmented space for SVM training. Recently, Li et al., (2014) proposed a heterogeneous feature augmentation (HFA) method for heterogeneous domains, in which the features from both the source and target domains are transformed into an augmented homogeneous common feature space. Wang and Gao have proposed a multiple-domain data representation method based on nonnegative matrix factorization (NMF) to map all the samples from multiple-domains into a common space (Wang and Gao, 2014).

Semi-supervised writer adaptation has attracted much attention in recent years. Frinken and Bunke (2009) have used a self-training strategy for adapting a neural network classifier for handwritten words recognition. In Goldman and Zhou (2000), a co-training strategy has been used to combine the neural network with HMM for handwriting recognition. The co-training strategy consists of two classifiers that teach each other on the unlabeled data. Oudot et al. (2005) have proposed to combine the supervised and self-supervised approaches for semi-supervised writer adaptation. Ball and Srihari have proposed to use the self-training strategy for HMM model retraining for English and Arabic handwriting recognition (Ball and Srihari, 2009). Vajda et al. have proposed semi-supervised ensemble learning for reducing the human effort in character labeling (Vajda et al., 2011).

In the unsupervised scenario, most of the adaptation approaches use the source trained classifiers. Then they define a new transformed feature to capture the relation between the training and test data distributions. Ben-David et al. (2007) have tried to learn directly a new representation which minimizes a bound on the test data generalization error. Gong et al. (2012) have considered an infinite set of intermediate subspaces through learn a symmetric kernel between source and target datasets by computing the geodesic flow along a latent manifold.

There have not been many works in unsupervised writer adaptation. Veeramachaneni and Nagy (2005) have proposed a model by assuming Gaussian field class conditional distribution for field classification. Tenenbaum and Freeman (2000) have employed a bilinear model to separate the style and the class knowledge in a group of patterns. Zhang et al. (2011) have proposed to train a style normalized transformation for each field.

In this paper, we propose a new domain adaptation framework for isolated characters recognition named as "large margin domain transfer" (LMDT). This approach jointly minimizes the difference between the data distribution of training and a test datasets as well as estimates a robust classifier. The tests were conducted on classification of the new handwriting styles that comes from a different distribution as that of the training data, both in the unsupervised and semi-supervised scenarios.

The rest of the paper is organized as follows: in Section 2, we introduce our proposed framework LMDT. Section 3 presents a description of employed test bench and discusses on simulation results. Finally, Section 4 concludes the paper and comments on how this algorithm can be further extended.

2. Large margin domain transfer (LMDT)

The main motivation of the proposed algorithm in this study is to maximize the classification rate in the target domain via weighting the classifier results using different kernel parameters. Due to very good generalization performance of large margin idea in classification applications, we propose a large margin domain transfer method for character recognition applications. In this section, the proposed mathematical framework for LMDT is discussed for domain adaptation.