J. Vis. Commun. Image R. 38 (2016) 307–315

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

A weakly supervised large margin domain adaptation method for isolated handwritten digit recognition ☆

Hamidreza Hosseinzadeh [a,*], Farbod Razzazi [a], Ehsanollah Kabir [b]

[a] Department of Electrical and Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran
[b] Department of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran

ARTICLE INFO

Article history:
Received 14 December 2014
Accepted 23 February 2016
Available online 2 March 2016

Keywords:
Writer adaptation
Domain adaptation
Handwriting recognition
Transformation learning
Feature learning
Large margin semi-supervised learning

ABSTRACT

Learning handwriting categories fail to perform well when trained and tested on data from different databases. In this paper, we propose a novel large margin domain adaptation algorithm which is able to learn a transformation between training and test datasets in addition to adapting the parameters of classifier using a few or even no training labeled samples from target handwriting dataset. Additionally, we developed a framework of ensemble projection feature learning for datasets representation as a front end for our algorithm to utilize the abundant unlabeled samples in target domain. Experiments on different handwritten digit datasets adaptations demonstrate that the proposed large margin domain adaptation algorithm achieves superior classification accuracy comparing with the state of the art methods. Quantitative evaluation of the proposed algorithm shows that semi-supervised adaptation utilizing one sample per class of target domain set reduces the error rates by 64.72% comparing with a corresponding SVM classifier.

1. Introduction

Supervised machine learning has already been widely studied and achieved significant success, along with the advances in machine learning. When applying the results of supervised machine learned methods to classification or regression problems, it is typically assumed that the labeled training data (source) and test data (target) are drawn from the same distribution which is far away from 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 When an existing training data is outdated, and the new labeled dataset is very little, and practically it is difficult to recollect the training data, the classifiers should be learned on the training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 minimum number of labeled samples from the target domain. In such cases, adaptation algorithms from training dataset to test dataset would be desirable [1–3].

Handwriting style variations across writers make handwriting recognition a challenging problem. To deal with this variation, 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*. As another application, the style of written text is highly dependent to the design of the data gathering form, the specifications and behavior of writers' community (e.g. age, educational skills, the time period permitted for writing, environmental conditions and the importance of the written data for the writer). Therefore, in many cases, the training source dataset is far from the test target dataset. This is known as *corpus adaptation* which is categorized as one of domain adaptation approaches. Our study is focused on this category of adaptation. Although, we did not investigate our algorithm for writer adaptation in this paper, the method can be easily tuned for it.

Domain adaptation has been studied in two main scenarios: one is the semi-supervised scenario, where the target domain has few labeled data. The other is the unsupervised scenario that considers only unlabeled data for the target domain to adapt the classifier. In both scenarios, the source is generally rich in labeled samples. The proposed algorithm in this paper was successfully deployed 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

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

* Corresponding author.
E-mail addresses: hr.hosseinzadeh@srbiau.ac.ir (H. Hosseinzadeh), razzazi@srbiau.ac.ir (F. Razzazi), kabir@modares.ac.ir (E. Kabir).

http://dx.doi.org/10.1016/j.jvcir.2016.02.018
1047-3203/

methods are Gaussian-fields and Harmonic-Function [4], Local-Global Consistency [5], and Manifold Regularization [6] for this task. Despite of leaning on a strong theory, these methods, unfortunately, cannot label unseen data well [7], because the whole graph should be reconstructed again every time new samples come. In our problem, this approach cannot be successfully employed; because we cannot find a suitable chain of samples to label the unlabeled samples. Therefore, we should use other approaches for handwriting corpus adaptation.

Semi-supervised writer adaptation has attracted much attention in recent years. Frinken and Bunke have used self-training strategy for adapting a neural network classifier for handwritten words recognition [8]. In [9], 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. [10] have proposed to combine the supervised and self-supervised approaches for semi-supervised writer adaptation. Ball and Srihari [11] have used a self-training strategy for HMM model re-training for English and Arabic handwriting recognition. Vajda et al. have proposed a semi-supervised ensemble learning method for reducing the human effort in character labeling [12]. However, writer adaptation approaches cannot be directly used in corpus adaptation problem.

In the unsupervised adaptation scenario, most of the adaptation approaches have used existing classifiers, but define new transformed features to capture the correspondence between the training and test data distributions. Ben-David et al. have tried to directly learn a new representation which minimizes a bound on the test data generalization error [13]. Gong et al. [14] have considered an infinite set of intermediate subspaces through learning 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 [15] have proposed a model by assuming a Gaussian field class conditional distribution for field classification. Tenenbaum and Freeman [16] have employed a bilinear model to separate the style and class knowledge in a group of patterns. Zhang et al. have proposed to train a style normalized transformation for each field [17]. Our method may be categorized as a transformed based domain adaptation.

1.1. Related work

Domain adaptation can be formulated on the basis of classifier parameter adaptation. Yang et al. [18] have proposed adaptive support vector machine (A-SVM) in which a target classifier is adapted from the existing source classifiers which has been previously trained with the labeled samples from the source domain. 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 $\sum_{k=1}^{K} \gamma_k = 1$ where K is the total number of source domains. In the experiments of [18], weights of all source classifiers are considered as assumed to be equal. Moreover, the authors have assumed that the target classifier is learned with only one kernel.

In a similar approach, Schweikert et al. [19] have presented a strategy for domain adaptation, which consists of a linear combination of the source classifiers together with the target classifier. 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 $\gamma \in [0,1]$ is the weight parameter to balance the two terms and is determined via grid search by optimizing multi-class error on the labeled target samples. Unfortunately, both these methods [18,19] do not employ the unlabeled samples in the target domain. In [20], adaptive multiple kernel learning (A-MKL) has been proposed to simultaneously learn a kernel function based on multiple types of kernels as well as a target classifier by minimizing both the structural risk function and the distribution mismatch between source and target domain.

Some other approaches are based on reweighting or selecting samples for the source domain data in order to minimize the difference between distributions of domains [21,22]. The key ideas behind these methods are that not all samples are considered equally for adaptation. Lastly, methods proposed in [23,24] are based on identifying which source samples are relevant for the target task.

Other approaches lie on the basis of transform based domain adaptation. They try to link between the train and test data distribution by feature space transformation. In [25], the adaptation has been performed by augmenting the feature space of both train and test datasets using feature replication (FR). The augmented features are then used for SVM training. Li et al. [26] have proposed a heterogeneous feature augmentation (HFA) method, in which the labeled training data from both the source and target domains are transformed into a common subspace by using two different projection matrices, and simultaneously learn the classifier with standard SVM in both linear and nonlinear cases. Saenko et al. [27] have proposed a method for semi-supervised domain adaptation based on metric learning to adapt features into a domain invariant space by learning a symmetric transformation.

Our proposed algorithm is focused on these three approaches, transform based domain adaptation, sample based, and classifier parameter adaptation.

1.2. Contribution of the paper

In this paper, we propose a new domain adaptation framework named as "large margin domain adaptation" (LMDA) to jointly learn a transformation from target to source dataset to map target features into the source domain as well as adapt the classifier parameters. It will be shown that regularizing the classifying functions alone would be inefficient. Therefore, LMDA method is developed for joint transformation based domain adaptation and classifier parameter adaptation. LMDA is shown to be a good choice for isolated characters recognition application. The tests were conducted on classification of the new handwriting style that comes from a different distribution as that of the training data, both in unsupervised and semi-supervised domain adaptation scenarios. Under this framework, we extend an ensemble projection feature representation as a front end of our algorithm to utilize the abundant unlabeled samples in the target domain. The main contributions of our paper include:

- To deal with the significant difference between feature distributions of source and target domains, we proposed LMDA procedure to select proper samples of source domain with the best match to the distribution of the target domain. These samples are used for extraction of the adaptation transformation.