Transferable Discriminative Dimensionality Reduction

Wenting Tu
Department of Computer Science and Technology
East China Normal University
Shanghai, China
Email: w.tingtu@gmail.com

Shiliang Sun
Department of Computer Science and Technology
East China Normal University
Shanghai, China
Email: slsun@cs.ecnu.edu.cn

***Abstract*—In transfer learning scenarios, previous discriminative dimensionality reduction methods tend to perform poorly owing to the difference between source and target distributions. In such cases, it is unsuitable to only consider discrimination in the low-dimensional source latent space since this would generalize badly to target domains. In this paper, we propose a new dimensionality reduction method for transfer learning scenarios, which is called transferable discriminative dimensionality reduction (TDDR). By resolving an objective function that encourages the separation of the domain-merged data and penalizes the distance between source and target distributions, we can find a low-dimensional latent space which guarantees not only the discrimination of projected samples, but also the transferability to enable later classification or regression models constructed in the source domain to generalize well to the target domain. In the experiments, we firstly analyze the perspective of transfer learning in brain-computer interface (BCI) research and then test TDDR on two real datasets from BCI applications. The experimental results show that the TDDR method can learn a low-dimensional latent feature space where the source models can perform well in the target domain.**

***Keywords*-dimensionality reduction; Fisher discriminant analysis; transfer learning; brain-computer interface**

I. INTRODUCTION

During recent years, both of dimensionality reduction and transfer learning have become hot topics in machine learning and data mining. In the single domain scenario where distributions of source and target domains are identical, many popular dimensionality reduction approaches, e.g. principal components analysis (PCA) [1] and locality preserving projection (LPP) [2], and Fisher discriminant analysis (FDA) [3], that focus on finding a low-dimensional latent space obtain good results. However, in many real-world applications, this the assumption does not hold and traditional classification methods might perform badly. For example, FDA tends to give undesired results if samples in training and test sets belong to different distributions. Transfer learning [4], [5], [6], [7] aims to solve the problem when the training data from a source domain and the test data from a target domain follow different distributions or are represented in different feature spaces.

Our work aims to design a dimensionality reduction method for transfer learning by improving the previous framework of discriminative dimensionality reduction. It learns a low-dimensional latent space that not only preserves discrimination of data, but also bridges the source to target domain. There are two main motivations under our work. First, we wish to improve the effectiveness of statistical models which are constructed on the source data but used to the target domain. This can be achieving by encouraging source and target distributions in the low-dimensional latent space close to each other, while the discrimination of the space is preserved. In other words, our method takes both discrimination and transferability of the feature space into consideration. The second motivation is to preserve the low-cost merit of previous discriminative dimensionality reduction. This encourages us to design mathematical terms which are simple and can be merged into the previous framework. Under the motivations mentioned above, we propose transferable discrimination dimensionality reduction (TDDR) in this paper. First, we design mathematical terms that evaluate the discrimination of domain merged training samples and the transferability of an embedded space to bridge source to target domain. Then, we merge these terms into the previous framework of discriminative dimensionality reduction. As a result, TDDR simultaneously maximizes the discrimination of data and minimizes the distance between distributions of the data in different domains. Meanwhile, it can be solved by the generalized eigen-decomposition, which is simple, rapid and accurate.

We testify our method using real-world data. It is worth mentioning that we also analyze the perspective of transfer learning in brain-computer interface (BCI) research, which is another important work of this paper. Up to now, as we know, most of the transfer learning methods are experimented on the real-world application such as text classification, image classification and clustering, sentiment classification, wifi digit, and so on. There is very few literature discuss its perspectives for BCI applications. In this section, we exploit perspectives of transfer learning methods in addressing limitations in BCI research and perform TDDR in real and public BCI datasets to verify it.

The rest of this paper is organized as follows. In Sec. 2, we propose TDDR. Experiments on real-world data sets, together with the perspective analysis of transfer learning in the BCI research, are given in Sec. 3. Finally, Sec. 4 concludes this paper and gives possible future research topics.

II. TDDR

One important assumption of the popular dimensionality reduction FDA (can see [3] for details) is that training and test data should be drawn from a same distribution. However, in transfer learning, the source training and target test data are always from different distributions, and thus FDA may generalize badly.

In transfer learning, a common setting is that there are a lot of labeled samples in the source domain and a few or no labeled data in the target domain. The final objective is to predict labels of test samples from the target domain. Since the target training set is too small, classification tasks often mostly depend on source training samples. However, due to the difference of distributions between the source and target domains, the model trained from the source training set often performs poorly on target test samples. As a result, the low-dimensional latent space for dimensionality reduction in transfer learning should take not only the discrimination but also the transferability – the ability to bridge source to target domain, into consideration.

A. Problem Formulation

We formalize the problem of dimensionality reduction in transfer learning as follows. Suppose we have a large training set X_{tr}^S from the source domain S, and a very small training set X_{tr}^T from the target domain T. Later tasks (e.g. regression or classification) are evaluated on the target test set X_{te}^T. The difficulties lie in the difference between the source and target domains that prevents us from utilizing the rich samples from the source domain. Therefore, the quality of a low-dimensional space should be considered by its ability to bridge source domain to target domain, which we call as transferability. However, only pursuing transferability may make the low-dimensional space unsuitable for latter classification task. As a result, we should take both transferability and discriminability into consideration. In this section, we design mathematical terms to represent discrimination and transferability of a low-dimensional space in the transfer learning scenario. By integrating them together , we construct an objective function for transfer dimensionality reduction. This method is called as transferable discriminative dimensionality reduction (TDDR) since it pursues transferability and discrimination together. The behavior of TDDR will be also analyzed in this section.

B. Domain-Merged Within and Between-Class Scatter Matrices

In transfer learning, we often face two training sets from source domain and target domain, respectively. Thus, the within and between class scatter measurement should computed on the dataset merged by them. However, considering the different importance of the source and target domains, the training datasets should have different weights. We define merged training set as $X_{mtr}^{ST} = \{X_{tr}^S; X_{wtr}^T\}$, where X_{tr}^S denotes the source training set and X_{wtr}^T denotes the weighted target training set. The weights used to weight the target training samples can control the influence of target training samples in computing within and between class scatter measurement. Considering that target training samples are drawn from the same distribution from the test samples, we always treat them more important than source training samples, while their importance is constrained by the sample number. Therefore, we define weights as $W_{tr}^T = 1 + n_t/n_s$ (n_s and n_t is the training sample numbers of the source and target sessions). This weight attaches more importance to the target training samples owing to the distribution similarity between the target training and test distribution. It also considers the reliability of the distribution estimation of target training set since it is constrained by the sample number of the target training set.

Then, between and within-class scatter measurement are computed on the merged training set X_{mtr}^{ST} to simultaneously increase the discrimination of source and target training samples in a unified low-dimensional latent space:

$$\begin{aligned} S_B^{ST} &= (\mu_1 - \mu_2)(\mu_1 - \mu_2)^\top, \\ S_W^{ST} &= \sum_{i=1,2} \sum_{x \in X_{tr}^{ST}} (x - \mu_i)(x - \mu_i)^\top, \end{aligned} \tag{1}$$

where μ_1 and μ_2 are class means of the merged training set X_{mtr}^{ST}.

C. Between-Domain Scatter Matrices

Since later classification models are mostly based on the source samples, the low-dimensional space should take a role as a bridge between the source and target distributions so that models trained with source samples can adapted to the target domain. Here, between-domain scatter matrix related to the distance between the source and target distributions in the low-dimensional are introduced. They are designed with and without label information respectively, to reveal the separation between domains.

1) Supervised Between-Domain Scatter Matrix: To improve the transferability of the low-dimensional space, we wish to minimize the distance between projected data from the source and target domains. We define a between domain scatter matrix based on the source and target training samples:

$$S_L^{ST} = (\mu_1^S - \mu_1^T)(\mu_1^S - \mu_1^T)^\top + (\mu_2^S - \mu_2^T)(\mu_2^S - \mu_2^T)^\top, \tag{2}$$

where $\mu_i^S (i = 1, 2)$ and $\mu_i^T (i = 1, 2)$ are class means of source and target training sets. The term $\mu_i^S - \mu_i^T$ $(i = 1, 2)$ reveals the scatter of class i between source and target domain. By putting in to Q_{min} to get penalized, the class means between source and target domain are close to each other. By that way, we decrease the divergence of source and target domain in the embedding space.

2) Unsupervised Between-Domain Scatter Matrix: However, since the number of target samples are small or even null. We need define another between domain scatter without label information, so that it can evaluated based on large no-labeled source and target samples. It is defined as the distance between centers of whole data:

$$S_U^{ST} = (\mu^S - \mu^T)(\mu^S - \mu^T)^\top, \tag{3}$$

where μ^S and μ^T are means of source training set and target test set. Similarly, by adding it into the definition of Q_{min}, we can increase the similarity between distributions of projected source training and target test samples by penalizing the distance of sample means.

3) Semi-supervised Between-Domain Scatter Matrix: If small target training samples and large target test samples are both offered, we can combine the supervised and unsupervised between domain scatter matrices to generate a semi-supervised one. Similar with weight definition in the merged training set X_{mtr}^{ST}, we attach more importance to the target samples with label information. Therefore, the semi-supervised between domain scatter matrix is formulated as follows:

$$S^{ST} = S_U^{ST} + (1 + n_{tr}^T / n_{te}^T) S_L^{ST}, \tag{4}$$

where n_{tr}^T and n_{te}^T are sample numbers of target training and test set.

D. Transferable Discriminative Dimensionality Reduction

By integrating $S_B^{ST}, S_W^{ST}, S^{ST}$ defined above in to the framework introduced in Sect. 2, we obtain transferable discriminative dimensionality reduction by maximizing:

$$J(\phi) = \frac{\phi^\top S_B^{ST} \phi}{\phi^\top (S_W^{ST} + \alpha S^{ST}) \phi}, \tag{5}$$

where α is a parameter to control the balance between the desired levels of discrimination and transferability.

In real applications of TDDR, α can be defined by users or determined by the K-fold cross-validation technology. However, since training samples are mostly from the source domain, the cross-validation technology used to determine parameters in transfer learning should have some reasonable modification. Here, we propose a "target-priority" strategy to modify the parameter selection step in the previous cross-validation technology when it is used to perform parameter selection in transfer learning scenarios. Speaking briefly, we firstly choose a set of α-values corresponding to the best performance on target samples (owning to the small number of target samples, there are always a lot of α-values corresponding to the best performance on target samples). Then, in that set, we choose the α-value corresponding to the best performance on source samples as the final α to be used in later tasks. This strategy gives a prior consideration to the performance on target samples because these samples are drawn from the same distribution as test samples. As a result, this parameter selection method adapts to transfer learning problems and is used to determine α in our later experiments with K is 10 and the candidate α-value set is $[0.1, 0.15, 0.2, 0.25, \cdots, 1]$.

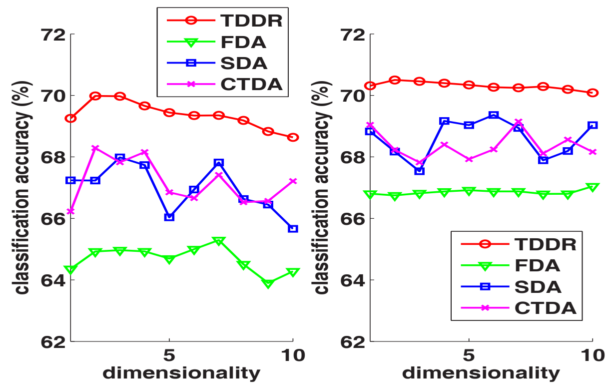

Figure 1. Comparison of performances of TDDR, FDA, SDA, and CTDA spaces in the subject transfer experiment (left: kNN; right: SVM).

III. Real-World Application: Transfer Learning for Practical BCIs

Long time-consuming training sessions bring huge difficulties to the practical wide use of BCI devices [8]. For solving this limitation, we can use training sets of other users whose characteristics are similar (e.g. same age or sex) with the current user who is unable or unwilling to perform any training session to train the BCI devices. We term this strategy as subject transfer. Moreover, for the users who performed training sessions before, we can use the training set that collected during the latest training session to help the current task. This is called session transfer. However, due to the large inter-subject variabilities among different users and time-variances during different time sessions, subject transfer and session transfer face the problem that training and test sets are drawn from different distributions. As a result, transfer learning methods are of a lot of potentialities to achieve subject transfer and session transfer to help bring BCI system to people's real lives. Here we employ the TDDR algorithm on two real datasets in the BCI research to simulate performing subject transfer and session transfer, respectively.

We employ the subject transfer and session transfer experiments on real-data [9] and [10]. Except the FDA method, we compare our method with two other state-of-arts: Semi-supervised Discriminant Analysis (SDA) and Cluster based Transferred Discriminant Analysis (CTDA), which are also extended versions of discriminative dimensionality reduction (For details, can be see [11] and [12]). In the subject transfer

Table I
THE CLASSIFICATION ACCURACIES (%) OF kNN ($k = 5$) CLASSIFIERS IN TDDR, FDA, SDA, AND CTDA SPACES.

Method	Dimensionality					Average
	1	2	3	4	5	
TDDR	$\mathbf{75.81 \pm 3.3}$	$\mathbf{75.55 \pm 3.2}$	$\mathbf{74.94 \pm 3.1}$	$\mathbf{73.74 \pm 4.0}$	$\mathbf{72.08 \pm 4.6}$	$\mathbf{74.43 \pm 3.6}$
FDA	71.50 ± 3.2	68.71 ± 4.3	66.65 ± 5.5	64.58 ± 5.4	63.37 ± 5.6	66.96 ± 3.6
SDA	73.68 ± 3.5	71.43 ± 4.1	72.57 ± 4.3	64.80 ± 4.9	66.90 ± 5.0	69.87 ± 4.4
CTDA	72.76 ± 3.8	71.08 ± 3.6	67.28 ± 4.8	71.16 ± 5.0	68.64 ± 5.6	70.25 ± 4.5

experiment, two classifiers (kNN with k = 5, SVM with C = 1 and polynomial kernel) are employed to perform classification tasks. With the dimensionality of features spaces varies form 2 to 10, the average result over all situations (subject x help y) is presented. In the subject transfer experiment, we do not use any training sample of the target subject since no target training sample is more similar with the real subject transfer setting. Therefore, in this experiment, $S^{ST} = S_U^{ST}$ in the TDDR approach. Otherwise, we vary the number of training samples of target user from 2 to 8 and the average result is presented. Table I and Fig.1 show the performances in the subject transfer and session transfer experiments, respectively. According to these results, we find TDDR space is much better than the spaces obtained with other approaches in according to the higher accuracy of the classification models. Compared with SDA and CTDA, our method has less parameters and lower computational burden. This may be the reason, we speculate, why our method can more robust and better performances.

IV. CONCLUSION AND FUTURE WORK

In this paper, we propose a dimensionality reduction method in transfer leaning that called Transferable Discriminative Dimensionality Reduction (TDDR). Its objective is to guarantee discrimination and transferability of the low-dimensional latent space simultaneously. The theory analysis and experimental results verify its effectiveness to find a feature space against the difference between training and test distributions. Moreover, the perspective analysis of transfer learning methods in BCI research is another contribution in this paper. We show that two strategies termed subject transfer and session transfer which are a lot of potential for practical application of BCI systems can be achieved with the help of transfer learning methods, such as our TDDR algorithm.

One remaining important direction worth researching is to extend TDDR to more versions to deal with different transfer learning scenarios. This is motivated by the fact that, in real-world application, the condition of different transfer learning scenarios may be different. Here, we offer some directions and suggestions of future TDDR extensions to encourage people to investigate: kernel TDDR, zero target training (no target training data), inductive transfer learning (target test samples are unseen), multiple sources (the number of source domains are more than one), online transfer learning (for online and real-time tasks) and so on.

ACKNOWLEDGMENT

This work is supported in part by the National Natural Science Foundation of China under Project 61075005, and the Fundamental Research Funds for the Central Universities.

REFERENCES

[1] Jolliffe, I.: Principal component analysis. New York: Springer. (2002)

[2] He, X. and Niyogi, P.: Locality preserving projections. Advances in Neural Information Processing Systems, 16, 153–160 (2003)

[3] Fisher, R.A.: The use of multiple measurements in taxonomic problems. Annals of Eugenics, 7, 179–188 (1936)

[4] Caruana, R.: Multitask learning. Machine Learning, 28(1), 41–75.(1997)

[5] Pan, S.J. and Yang, Q.: A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 1345–1359 (2009)

[6] Dai, W. and Yang, Q. and Xue, G.R. and Yu, Y.: Boosting for transfer learning. Proceedings of the 24th international conference on Machine learning. 193–200 (2007)

[7] Raina, R. and Battle, A. and Lee, H. and Packer, B. and Ng, A.Y.: Self-taught learning: transfer learning from unlabeled data. Proceedings of the 24th International Conference on Machine learning, 759–766 (2007)

[8] Pfurtscheller, G. and Neuper, C. and Guger, C. and et.al: Current trends in Graz brain-computer interface (BCI) research. IEEE Transactions on Rehabilitation Engineering, 8(2), 216–219 (2000)

[9] Millan, J.: On the need for on-line learning in brain-computer interfaces. Proceedings of the 2004 IEEE International Joint Conference on Neural Networks, 4, 2877–2882 (2004)

[10] Lal, T.N. and Hinterberger, T. and Widman, G. et al: Methods towards invasive human brain computer interfaces. Advances in Neural Information Processing Systems, 17, 737–744 (2005)

[11] Cai, D. and He, X. and Han, J.: MSemi-supervised discriminant analysis. Proceedings of the IEEE 11th International Conference on Computer Vision, 1–7 (2007)

[12] Wang, Z. and Song, Y. and Zhang, C.: Transferred dimensionality reduction. Machine Learning and Knowledge Discovery in Databases, 550–565 (2008)