A DISCRIMINATIVE DOMAIN ADAPTATION MODEL FOR CROSS-DOMAIN IMAGE CLASSIFICATION

Yen-Cheng Chou, Chia-Po Wei, and Yu-Chiang Frank Wang

Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan
{chouyencheng, cpwei, ycwang}@citi.sinica.edu.tw

ABSTRACT

Techniques of domain adaptation have been applied to address cross-domain recognition problems. In particular, such techniques favor the scenarios in which labeled data can be obtained at the source domain, but only few labeled target domain data are available during the training stage. In this paper, we propose a domain adaptation approach which is able to transfer source domain labeled data to the target domain, so that one can collect a sufficient amount of training data at that domain for recognition purposes. By advancing low-rank matrix decomposition for obtaining representative cross-domain data, our proposed model aims at transferring source domain labeled data to the target domain while preserving class label information. This introduces additional discriminating ability into our model, and thus improved recognition can be expected. Empirical results on cross-domain image datasets confirm the use of our proposed model for solving cross-domain recognition problems.

Index Terms— Domain adaptation, low-rank matrix decomposition, image classification

1. INTRODUCTION

When solving pattern recognition problems, one typically needs to collect training data in advance for designing the classifier. If the amount of training data is not sufficient, one might encounter overfitting or generalization problems, and thus the recognition performance will be degraded. However, in cross-domain recognition problems like cross-domain image classification or cross-camera action recognition, there might be only few training (labeled) data available at the domain of interest. In order to recognize the test data at that domain, it is desirable to transfer the labeled data from the source domain, so that the amount of training data would be increased at the target domain for alleviating the aforementioned problems.

The above process of transferring data from the source to target domain is considered as *domain adaptation*. Performing recognition at the target domain using transferred data has been widely applied in image processing and computer vision communities [1, 2, 3]. We note that, on the other hand, some existing works propose to derive a common feature space to relate source and target domain data for cross-domain recognition [4, 5]. Once this feature space is obtained, one can project training and test data onto this space for recognition purposes. However, this type of approach requires sufficient cross-domain data pairs for deriving the common feature space, and this might not be applicable in real-world applications.

In this paper, we focus on cross-domain classification problems in which only a small amount of labeled data are available at the target domain, and thus training classifiers directly at that domain is not preferable. Since no source/target domain data pair can be obtained for deriving a common feature space, we need to transfer source domain labeled data to the target domain for training/testing. We propose to extract representative information from source/target domain data via low-rank matrix decomposition [6], followed by a novel domain adaptation model with discrimination guarantees. Our model advances label information of cross-domain data and introduces additional discriminative ability to the transferred data. As a result, improved recognition can be expected as verified later by our experiments.

2. A BRIEF REVIEW OF DOMAIN ADAPTATION

Visual domain adaptation is typically addressed in a supervised [1, 2, 7, 8, 9] or an unsupervised [10, 11, 3] way, depending on whether the label information of either or cross-domain data is utilized during the adaptation process. A popular approach is to adapt the classifier learned at the source domain for recognizing target domain data. For example, Yang *et al*. [7] adapted SVM classifiers learned in the source domain to handle target domain data. A relevant work in [8] applied multiple kernel learning (MKL) with SVM for observing data distributions across different domains. Duan *et al*. [9] further combined the ideas of [7, 8] and proposed an MKL-based approach for the task of video event recognition.

Another group of supervised domain adaptation approaches focus on modeling data (or features) across different domains. For example, Saenko *et al*. [1] applied metric learning for learning a transformation function which preserves the label information of cross-domain data. This metric function

was further extended in [2] to cope with the case in which the features observed at source and target domains are distinct and thus can be with different dimensionalities.

For unsupervised approaches, the Grassmann manifold has been applied for domain adaptation, and the geodesic flow of cross-domain data observed on this manifold is applied to model the domain shift [10]. Gong et al. [11] extended the work of [10] and introduced a kernel function for reducing the computational cost with improved performance. By utilizing low-rank matrix decomposition, a recent approach of [3] proposed to associate cross-domain data for recognition. Although the above works have been successfully applied to cross-domain visual classification problem, no label information was considered during the process of domain adaptation. Based on the above observations, we propose a *supervised* domain adaptation algorithm, which learns a transformation function for mapping the source domain data to the target domain with discrimination guarantees. We now detail our proposed method in the following section.

3. OUR PROPOSED METHOD

3.1. Discriminative Domain Adaptation

3.1.1. Data Refinement for Domain Adaptation

In real-world applications, visual data is often noisy and thus degrades the resulting recognition performance. For example, if one applies the bag-of-words (BoW) model for representing image or video data, the extracted features will be inevitably corrupted by the clutter or background presented. Since performing recognition using such features would cause generalization problems, it would be necessary to apply some preprocessing techniques for refining visual features prior to domain adaptation.

Suppose that there exist n instances in a d-dimensional space, we have the data matrix $D = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_n] \in \mathbb{R}^{d\times n}$. For refinement purposes, we aim at decomposing D into $D = A + E$, in which each column of $A \in \mathbb{R}^{d\times n}$ denotes the hidden structure of the input data D, and the columns of $E \in \mathbb{R}^{d\times n}$ represent the corresponding error/noise to be disregarded. In our proposed method, we advance low-rank matrix decomposition [12] for addressing the above feature refinement task. The use of such techniques is based on the assumption/observation that the hidden representative structure of D admits a low-rank structure, and the error part of D to be removed is sparse. To decompose D, we solve the following optimization problem:

$$\min_{A} \|A\|_* + \lambda \|E\|_1 \quad \text{s.t.} \quad D = A + E, \tag{1}$$

where $\|\cdot\|_*$ and $\|\cdot\|_1$ denote nuclear and ℓ_1 norms, respectively. The above problem is convex and is known as *Principal Component Pursuit*, which can be efficiently solved by techniques like inexact augmented Lagrange multiplier (IALM) [12].

For domain adaptation problems, data are collected at the source domain $\mathcal{S}$ and the target domain $\mathcal{T}$, and we have $D^{\mathcal{S}} \in \mathbb{R}^{d_{\mathcal{S}}\times n_{\mathcal{S}}}$ and $D^{\mathcal{T}} \in \mathbb{R}^{d_{\mathcal{T}}\times n_{\mathcal{T}}}$ as data matrices for both domains. Note that $d_{\mathcal{S}}$ and $d_{\mathcal{T}}$ are the feature dimensions, and $n_{\mathcal{S}}$ and $n_{\mathcal{T}}$ are the numbers of instances in source and target domains, respectively. In our work, we particularly consider the setting $n_{\mathcal{T}} << n_{\mathcal{S}}$ (i.e., only a small amount of labeled data is available at the target domain). This is why the mapping of source domain labeled data into the target domain is desirable.

Once the above low-rank based preprocessing is complete, the hidden structures (i.e., refined data) $A^{\mathcal{S}} \in \mathbb{R}^{d_{\mathcal{S}}\times n_{\mathcal{S}}}$ and $A^{\mathcal{T}} \in \mathbb{R}^{d_{\mathcal{T}}\times n_{\mathcal{T}}}$ for both domains can be obtained. We note that, these two data matrices can be viewed as representative visual features, and they will be utilized for learning our discriminative domain adaptation model.

3.1.2. Our Proposed Domain Adaptation Model

Given refined source and target domain data $A^{\mathcal{S}}$ and $A^{\mathcal{T}}$, our goal is to learn a discriminative linear transformation W for projecting source domain data into the target domain with discrimination guarantees. In order to introduce data separation ability into our model for improved recognition, we advance the label information during the adaptation process. For simplicity, we consider the features in source and target domains are of the same dimensionality (i.e., $d_{\mathcal{S}} = d_{\mathcal{T}}$) in this paper.

To take the label information into the design of our domain adaptation model, we propose to maximize the correlation between the refined data $WA^{\mathcal{S}}$ transformed from the source domain and that at the target domain $A^{\mathcal{T}}$, if such cross domain data belong to the same class. Therefore, we define a label consistency matrix $L \in \mathbb{R}^{n_{\mathcal{S}}\times n_{\mathcal{T}}}$ as

$$L_{i,j} = \begin{cases} 1, & \text{if label}(\mathbf{d}_i^{\mathcal{S}}) = \text{label}(\mathbf{d}_j^{\mathcal{T}}) \\ 0, & \text{otherwise,} \end{cases} \tag{2}$$

where $\mathbf{d}_i^{\mathcal{S}}$ is the ith instance in the source domain, and $\mathbf{d}_j^{\mathcal{T}}$ represents the jth instance in the target domain. Next, we derive the linear transformation W for discriminative domain adaptation by solving the following optimization problem:

$$\begin{aligned} &\max_{W} \sum_{i,j} \left[\left((WA^{\mathcal{S}})^{\mathsf{T}} A^{\mathcal{T}} \right) \odot L \right]_{i,j} \\ &\text{s.t.} \quad WW^{\mathsf{T}} = I, \end{aligned} \tag{3}$$

where $\odot$ denotes the element-wise matrix product. In (3), the equality constraint implies that the columns of W are orthonormal, which ensures that each column of $A^{\mathcal{S}}$ and $WA^{\mathcal{S}}$ has the same l_2-norm. Without this normalization constraint, (3) might approach infinity and produce trivial solutions. From (3), it can be seen that only the correlation between $WA^{\mathcal{S}}$ and $A^{\mathcal{T}}$ of the same category will be maximized.

To derive the closed-form solution of (3), we rewrite the objective function of (3) as

$$\text{Tr}\left(\left((WA^{\mathcal{S}})^{\mathsf{T}} A^{\mathcal{T}} \right)^{\mathsf{T}} L \right) = \text{Tr}\left(W \cdot A^{\mathcal{S}} L A^{\mathcal{T}^{\mathsf{T}}} \right). \tag{4}$$

According to [13], the maximum of (4) can be achieved by having $W = VU^{\mathsf{T}}$, if we have the singular value decomposition of $A^{\mathcal{S}}LA^{\mathcal{T}^{\mathsf{T}}}$ as $U\Sigma V^{\mathsf{T}}$ (see [13] for detailed derivations). Once W is learned, the design of our domain adaptation model is complete. We now summarize our proposed method in Algorithm 1.

Algorithm 1 Discriminative Domain Adaptation

Input: Source data matrix $D^{\mathcal{S}}$, target data matrix $D^{\mathcal{T}}$, label consistency matrix L, and parameters $\lambda_{\mathcal{S}}$ and $\lambda_{\mathcal{T}}$

1. Obtain the hidden structures $A^{\mathcal{S}}$ and $A^{\mathcal{T}}$ via
   $\min_{A^{\mathcal{S}}} \left\|A^{\mathcal{S}}\right\|_* + \lambda_{\mathcal{S}} \left\|E^{\mathcal{S}}\right\|_1$
   s.t. $D^{\mathcal{S}} = A^{\mathcal{S}} + E^{\mathcal{S}}$
   $\min_{A^{\mathcal{T}}} \left\|A^{\mathcal{T}}\right\|_* + \lambda_{\mathcal{T}} \left\|E^{\mathcal{T}}\right\|_1$
   s.t. $D^{\mathcal{T}} = A^{\mathcal{T}} + E^{\mathcal{T}}$
2. Solve (3) by computing the SVD of $A^{\mathcal{S}}LA^{\mathcal{T}^{\mathsf{T}}} = U\Sigma V^{\mathsf{T}}$. Obtain the solution of (3) by $W = VU^{\mathsf{T}}$

Output: Linear transformation W for domain adaptation

3.1.3. Performing Recognition

Given source domain labeled data $D^{\mathcal{S}}$ and a small amount of training data $D^{\mathcal{T}}$ at the target domain, we first perform the above learning process for deriving W. Once W is observed, we train classifiers at the target domain using projected source domain labeled data $WD^{\mathcal{S}}$ and the target domain training data $D^{\mathcal{T}}$. The learned classifier will then be applied to recognize test inputs at the target domain.

3.2. Relation to Existing Work

We now discuss how our proposed method differs from existing domain adaptation approaches of Canonical Correlation Analysis (CCA) [14], Symmetric Regularized Cross-domain transformation (SRC-t) [1], Robust Domain Adaptation with Low-rank Reconstruction (RDALR) [3] and Geodesic Flow Kernel (GFK) [11]. When applying CCA [14] for domain adaptation, it aims at projecting $\mathcal{S}$ and $\mathcal{T}$ onto a subspace in which the correlation between cross-domain data is maximized. Since the label information is not considered, the CCA subspace does not distinguish between different classes and thus might not be preferable for classification [5].

Inspired by metric learning, SRC-t [1] derives a metric function W for associating cross domain data. Although both SRC-t and our method exploit label information during the domain adaptation process, SRC-t does not extract representative features from cross-domain visual data (as we do). As verified later in our experiments, its recognition performance will be sensitive to the noise and thus might not achieve satisfactory results.

On the other hand, a recent work of RDALR [3] addresses similar problems with the ability of handling noisy cross domain data. Although promising results are reported in [3], it does not consider the label information and thus their domain adaptation model lacks the discriminating ability as ours does. Moreover, RDALR requires the assumption that the transformed source data will be linearly reconstructed by target domain data. This assumption might not be practical when only a small amount of target domain data is available (which is the scenario of interest in our work).

Finally, GFK [11] advances the Grassmann manifold for domain adaptation, in which the geodesic flow derived from $D^{\mathcal{S}}$ and $D^{\mathcal{T}}$ is applied to model the domain shift. While GFK produces a closed-form solution and reports impressive results for cross-domain classification, it is an unsupervised method and does not consider label information during domain adaptation. It does not explicitly handle noisy or outlier data as ours or RDALR does either. Later we will verify that our proposed method outperforms these popular domain adaptation methods.

4. EXPERIMENTS

4.1. Datasets and Settings

To evaluate our proposed method for cross-domain image classification problems, we consider the *Office* [1] and *Caltech-256* [15] datasets for experiments. The *Office* dataset contains image of 31 object categories, which are collected from three different domains: *amazon*, *dslr* and *webcam*. Images of *amazon* are collected from the Internet (mostly with plain background). While *dslr* images are with high/sufficient resolution, *webcam* images are often overexposed or blurred with lower resolution. The *Caltech-256* dataset consists of object images of 256 categories (at least 80 images per category). These two datasets have been applied for solving cross-domain image classification tasks [1, 2, 10, 11, 3].

To represent each object image, we extract SURF features from images from *amazon* and construct a codebook with 800 visual words by k-means clustering (as [1] did). All object images are thus converted into a 800-dimensional normalized Bag-of-Words (BoW) model, which will be applied for both domain adaptation and recognition. When designing classifiers at the target domain, we consider SVM classifiers with RBF kernels. The SVM parameters C and γ are selected via a three-fold cross validation over the range of $\{2^{-5}, 2^{-3}, \ldots, 2^{3}\}$. As noted in Algorithm 1, our proposed model needs to determine two parameters $\lambda^{\mathcal{S}}$ and $\lambda^{\mathcal{T}}$ when performing data refinement via low-rank matrix decomposition. For simplicity, we set $\lambda^{\mathcal{S}}$ and $\lambda^{\mathcal{T}}$ to be the same value, which is determined for the best recognition performance over the range of $\{0.01, 0.02, \ldots, 0.15\}$.

4.2. Recognition Results for *Office*

We first consider all 31 classes of three image domains in the *Office* dataset for cross-domain recognition. We apply the

Table 1. Comparisons of recognition performance (%) on the *Office* dataset. Note that $\mathcal{S}$ represents the source domain, and $\mathcal{T}$ is the target domain. We have W, D, and A as image domains of *webcam*, *dslr*, and *amazon*, respectively.

$\mathcal{S}$	$\rightarrow$	$\mathcal{T}$	TDO	NC	SRC-t [1]	RDALR [3]	Our Method
W	$\rightarrow$	D	27.40 ± 0.12	23.66 ± 0.03	25.3	32.89 ± 1.2	32.84 ± 0.11
D	$\rightarrow$	W	31.21 ± 0.13	36.52 ± 0.05	36.1	36.85 ± 1.9	35.00 ± 0.11
A	$\rightarrow$	W	52.94 ± 0.07	41.74 ± 0.05	50.4	50.71 ± 0.8	56.15 ± 0.05
Average			37.18	33.97	37.26	40.15	41.33

Table 2. Comparisons of recognition performance (%) on ten object categories of the *Office* and *Caltech256* datasets. Note that we have W, D, A, and C as image domains of *webcam*, *dslr*, *amazon*, and *Caltech-256*, respectively.

$\mathcal{S}$	$\rightarrow$	$\mathcal{T}$	TDO	NC	SRC-t [1]	GFK [11]	Our Method
A	$\rightarrow$	W	56.79 ± 0.22	50.77 ± 0.15	36.0 ± 1.0	53.7 ± 0.8	61.30 ± 0.27
A	$\rightarrow$	D	53.98 ± 0.37	47.48 ± 0.19	33.7 ± 0.9	47.0 ± 1.2	57.36 ± 0.10
A	$\rightarrow$	C	27.18 ± 0.24	40.26 ± 0.07	27.3 ± 0.7	37.8 ± 0.4	39.43 ± 0.23
W	$\rightarrow$	A	40.85 ± 0.08	39.93 ± 0.05	32.3 ± 0.8	42.8 ± 0.7	44.71 ± 0.14
W	$\rightarrow$	D	54.06 ± 0.14	66.77 ± 0.22	51.3 ± 0.9	75.0 ± 0.7	73.39 ± 0.02
W	$\rightarrow$	C	28.49 ± 0.21	32.22 ± 0.04	21.7 ± 0.5	32.8 ± 0.7	31.66 ± 0.09
D	$\rightarrow$	A	33.16 ± 0.02	39.51 ± 0.06	30.3 ± 0.8	45.0 ± 0.7	$46.02 \pm 6e^{-3}$
D	$\rightarrow$	W	59.15 ± 0.18	74.26 ± 0.15	55.6 ± 0.7	78.7 ± 0.5	76.23 ± 0.00
D	$\rightarrow$	C	27.04 ± 0.17	32.93 ± 0.04	22.5 ± 0.6	32.7 ± 0.4	33.74 ± 0.01
C	$\rightarrow$	A	39.39 ± 0.09	47.86 ± 0.07	33.7 ± 0.8	42.0 ± 0.5	47.32 ± 0.08
C	$\rightarrow$	W	58.58 ± 0.28	51.13 ± 0.31	34.7 ± 1.0	54.2 ± 0.9	63.51 ± 0.10
C	$\rightarrow$	D	53.82 ± 0.12	50.19 ± 0.14	35.0 ± 1.1	49.5 ± 0.9	57.36 ± 0.06
Average			44.37	47.78	34.51	49.27	52.67

settings of [1] for collecting source domain labeled data: 8 images randomly selected per class from *dslr/webcam*, and 20 images per class from *amazon*. For the target domain, 3 images per class are randomly selected for training, and the rest are for testing. As suggested in [2], particular object images for testing in each category are held out during training, which makes the recognition problem more practical yet challenging. We repeat the above experiments 20 times, and report the average recognition performance.

To compare our proposed model with other methods, we consider (1) training and testing using target data only (noted as TDO), (2) naive combination (NC) of source and target domain labeled data for training, (3) SRC-t [1] and (4) RDALR [3]. Table 1 lists the average recognition performance of different approaches, and it can be seen that our method achieved the best or comparable results among all.

4.3. Recognition Results for *Office* and *Caltech-256*

Since ten object categories are present in both *Office* and *Caltech-256* datasets, images of *Caltech-256* can be treated as those in another domain for recognition. We apply the same settings as [11] for training and testing: for source domain data, 8 labeled images are randomly selected per class in *dslr/webcam* and 20 labeled images per category in *amazon/Caltech-256*. For target domain data, 3 labeled images per class are randomly selected for training, and the remaining ones are for testing. We also perform 20 runs for our experiments.

Besides TDO, NC, and SRC-t, we also consider GFK (with PCA for dimension reduction in both domains) [11] for comparisons. Table 2 lists the average recognition performance across different image domains. From this table, it can be seen that our proposed method obtained improved and comparable results than other domain adaptation approaches, and achieved the highest average recognition rate. Therefore, the use of our proposed model for addressing cross-domain image classification problems can be successfully verified.

5. CONCLUSION

We proposed a discriminative domain adaptation model for addressing cross-domain image classification problems. We particularly consider the scenario in which source domain labeled data are available, but only few labeled data can be obtained at the target domain. With a low-rank based technique for extracting representative cross-domain visual features, our proposed model aims at transferring source domain labeled data to the target domain while preserving label information. This allows one to collect a sufficient amount of labeled data at that domain with additional discriminating ability, and thus improved recognition can be achieved. Experiments on two benchmark image datasets for cross-domain image classification confirmed the effectiveness of our proposed method, which was shown to outperform state-of-the-art domain adaptation approaches.

Acknowledgement This work is supported in part by National Science Council of Taiwan via NSC100-2221-E-001-018-MY2.

References

[1] K. Saenko, B. Kulis, M. Fritz, and T. Darrell, "Adapting visual category models to new domains," in *ECCV*, 2010.

[2] B. Kulis, K. Saenko, and T. Darrell, "What you saw is not what you get: Domain adaptation using asymmetric kernel transforms," in *IEEE CVPR*, 2011.

[3] I-H. Jhuo, D. Liu, D. T. Lee, and S.-F. Chang, "Robust visual domain adaptation with low-rank reconstruction," in *IEEE CVPR*, 2012.

[4] J. Liu, M. Shah, B. Kuipers, and S. Savarese, "Cross-view action recognition via view knowledge transfer," in *IEEE CVPR*, 2011.

[5] C.-H. Huang, Y.-R. Yeh, and Y.-C. F. Wang, "Recognizing actions across cameras by exploring the correlated subspace," *ECCV Workshop on Video Event Categorization, Tagging and Retrieval*, 2012.

[6] J. Wright, A. Ganesh, S. Rao, and Y. Ma, "Robust principal component analysis: exact recovery of corrupted low-rank matrices by convex optimization," in *NIPS*, 2009.

[7] J. Yang, R. Yan, and A. G. Hauptmann, "Cross-domain video concept detection using adaptive SVMs," in *ACM Multimedia*, 2007.

[8] L. Duan, I. W. Tsang, D. Xu, and S. J. Maybank, "Domain transfer SVM for video concept detection," in *IEEE CVPR*, 2009.

[9] L. Duan, D. Xu, I. W. Tsang, and J. Luo, "Visual event recognition in videos by learning from web data," in *IEEE CVPR*, 2010.

[10] R. Gopalan, R. Li, and R. Chellappa, "Domain adaptation for object recognition: An unsupervised approach," in *IEEE ICCV*, 2011.

[11] B. Gong, Y. Shi, F. Sha, and K. Grauman, "Geodesic flow kernel for unsupervised domain adaptation," in *IEEE CVPR*, 2012.

[12] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *Journal of ACM*, 2011.

[13] R. A. Horn and C. R. Johnson, *Matrix Analysis*, chapter 7.4, p. 432, Cambridge University Press, 1990.

[14] H. Hotelling, "Relations between two sets of variates," *Biometrika*, 1936.

[15] G. Griffin, A. Holub, and P. Perona, "Caltech-256 object category dataset," Tech. Rep., California Institute of Technology, 2007.