CONNECTING THE DOTS WITHOUT CLUES: UNSUPERVISED DOMAIN ADAPTATION FOR CROSS-DOMAIN VISUAL CLASSIFICATION

Wei-Yu Chen[1,2], *Tzu-Ming Harry Hsu*[1,2], *Cheng-An Hou*[2], *Yi-Ren Yeh*[3], *Yu-Chiang Frank Wang*[2]

[1]Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan
[2]Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan
[3]Department of Applied Mathematics, Chinese Culture University, Taipei, Taiwan

ABSTRACT

Many real-world visual classification tasks require one to recognize test data in a particular domain of interest, while the training data can only be collected from a different domain. This can be viewed as the problem of unsupervised domain adaptation, in which the domain difference and the lack of cross-domain label/correspondence information make the recognition task very difficult. In this paper, we propose to exploit the cross-domain data correspondence using both observed data similarity and labels transferred from the source domain. This allows us to perform distribution matching for cross-domain data with recognition guarantees. Our experiments on three different cross-domain visual classification tasks would confirm the effectiveness of our method, which is shown to perform favorably against state-of-the-art unsupervised domain adaptation approaches.

Index Terms— Unsupervised domain adaptation, transfer learning, cross-domain visual classification

1. INTRODUCTION

Most pattern recognition methods assume that training and testing data exhibit the same or similar feature distributions. Unfortunately, this scenario might not be practical for real-world applications. For example, one might need to recognize images at a particular view, while the training ones are captured by cameras at distinct views or with different resolutions [1]. In such cases, training and test data are considered to be in different *domains*, and a bias (or mismatch) between their feature distributions can be observed. As a result, features/classifiers learned from the source domain cannot be expected to generalize well to the test data in the target domain.

To overcome the domain mismatch problem, researchers advance the technique of *domain adaptation* for cross-domain classification [2, 3]. Domain adaptation is to associate source and target domain data by eliminating the domain bias. Depending on the number of labeled data available in the target domain, the tasks of domain adaptation can be divided into different categories [3, 4, 5, 6, 7]. In this paper, we focus on the challenging problem of *unsupervised domain adaptation*, in which one is able to collect labeled data in the source domain data while *only* unlabeled test data can be observed in the target domain.

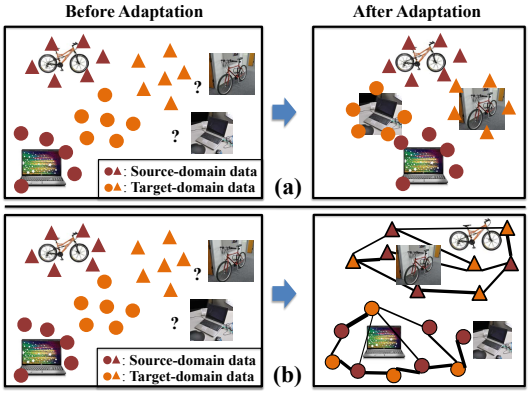

Fig. 1. Unsupervised domain adaptation using (a) feature-matching based methods (e.g., [6] or [14]) and (b) ours. Note that methods like [6] or [14] match cross-domain distributions with global means, while ours is able to exploit cross-domain data correspondences (shown in connected lines) with recovered label information.

Generally, two strategies exist for unsupervised domain adaptation [2, 7]: *instance reweighting* [8, 9] and *feature matching* [10, 11, 12, 13]. The former advocates the weighting of source domain instances for compensating the domain bias. While this strategy is applicable when both source and target domain data are of the same type of features, the correspondences between cross-domain data and their contributions are not exploited during the reweighting process.

On the other hand, feature-matching based methods aim at relating cross-domain data distributions in a transformed space. Among existing approaches, *Maximum Mean Discrepancy* (MMD) [15] has been widely applied to measure the difference between the transformed cross-domain data. For example, Pan *et al.* [6] proposed the Transfer Component Analysis (TCA) to adapt the marginal distributions of cross-domain data. It is achieved by matching the global means of the low-dimensional kernel embedding of cross-domain data. Extended from TCA, Long *et al.* [7] presented the Joint Distribution Adaptation (JDA) to further associate the joint dis-

tributions of cross-domain data by matching both their global and class-wise means. While these approaches allow source and target domain data of different types of features, the direct use of global and class-wise means for distribution matching might not always be preferable.

To address the limitations of instance reweighting and feature matching based techniques, we propose a novel approach for solving domain mismatch problems, as illustrated in Figure 1. Our proposed algorithm can be viewed as a unified formulation which solves instance reweighting and feature matching tasks *simultaneously*. Inspired by the concept of pseudo labels for domain adaptation [7], our method is able to observe the similarities of candidate source and target domain data pairs by learning correspondence transformation between cross-domain data. Such correspondence information is applied for perform matching at the instance level, while the associated pairwise weights can be derived. As a result, we are able to disregard possible outlier data during the adaptation process. In addition, since the distribution matching is performed at the instance level, domain mismatch can be better eliminated.

We now summarize our contributions as follows:

• We uniquely integrate the concepts of instance reweighting and feature matching for unsupervised domain adaptation. The derived correspondence transformation is able to associate cross-domain data, so that cross-domain classification can be performed accordingly. (Section 2)

• We conduct experiments on benchmark cross-domain image classification datasets. We verify the effectiveness and robustness of our method, which is shown to perform favorably against several state-of-the-art unsupervised domain adaptation methods. (Section 3)

2. OUR PROPOSED METHOD

2.1. Problem formulation

We now define the problem to be solved, and introduce the notations which will be used in this paper. For unsupervised domain adaptation, we have source domain training data $X_S : \{\mathbf{x}_{s_i}\}_{i=1:N_s}$ with their corresponding labels $\{y_{s_i}\}_{i=1:N_s} \in (1, \ldots, C)$. As for the target domain, only unlabeled test data can be observed, i.e., $X_T : \{\mathbf{x}_{t_j}\}_{j=1:N_t}$. For the tasks of cross-domain image classification, we consider that both source and target domains contain the same C classes of interest, and the instances in both domains have the same type of features of dimension m. Based on the above settings, the goal of our work is to predict the labels of each data point in the target domain, denoted by y_{t_j}.

To eliminate the differences between domains without observing any target domain labels, we derive feature transformations f_s and f_t for mapping cross-domain data into a common space, in which the *distance* between the source and target domain data is minimized. In other words, we solve:

$$(f_s^*, f_t^*) = \arg\min_{f_s, f_t} Dist(f_s(X_S), f_t(X_T)), \qquad (1)$$

where $Dist(\cdot, \cdot)$ denotes the distance between transformed cross-domain data. Ideally, solving (1) indicates that one would match the distributions of cross-domain data in the derived common feature space. As noted in Section 1, existing *instance reweighting* approaches did not observe the correspondence between cross-domain pairs when eliminating the domain difference, while *feature matching* methods like TCA [6] or JDA [7] solve the above matching problem using only global and/or class-wise means.

Aiming at determining cross-domain data correspondence for improved distribution matching, we propose to solve the following optimization problem:

$$(f_s^*, f_t^*) = \arg\min_{f_s, f_t} \sum_{i=1}^{N_s} \sum_{j=1}^{N_t} w_{ij} \, \|f_s(\mathbf{x}_{s_i}) - f_t(\mathbf{x}_{t_j})\|^2, \quad (2)$$

where the weight w_{ij} (to be learned) determines the importance of the correspondence $(\mathbf{x}_{s_i}, \mathbf{x}_{t_j})$ (i.e., the cross-domain data pairs i and j). We view $\mathbf{W} \in \mathbb{R}^{N_s \times N_t}$ as the *similarity matrix*, in which each entry w_{ij} denotes the associated weight. In the next subsection, we will explain how we derive $\mathbf{W}$ and the feature transformations (f_s, f_t).

2.2. Learning Cross-Domain Correspondences

2.2.1. Deriving the similarity matrix $\mathbf{W}$

When assessing the correspondence between source and target domain data in the transformed space, we consider the cross-domain data pairs with higher similarities to be assigned with larger weights. Recall that, for unsupervised domain adaptation, neither instance correspondence nor label information is available for target-domain data. Thus, we advance the pseudo labels inferred from the source domain for predicting the target domain labels [7].

In our work, we use the classifiers learned from the source domain to assign pseudo labels $\tilde{y}_{t_j}$ for target-domain data. And, we calculate the correspondence weight as follows:

$$w_{ij} = \begin{cases} exp(-\beta \|f_s(\mathbf{x}_{s_i}) - f_t(\mathbf{x}_{t_j})\|), & \text{if } y_{s_i} = \tilde{y}_{t_j} \\ 0, & \text{otherwise.} \end{cases} \qquad (3)$$

It is worth noting that, the parameter β is to control the sparsity of the similarity matrix $\mathbf{W}$ (i.e., the number of dominant correspondence pairs). When calculating $\mathbf{W}$, we apply the algorithm of *soft-Iterative Closest Point* [16] for normalizing each entry in $\mathbf{W}$. This process would make every data contribute identically and avoid possible adaptation of outlier data in the source domain.

2.2.2. Learning the correspondence transform $\mathbf{\Phi}$

With the determination of the similarity matrix $\mathbf{W}$, we now discuss how we learn the correspondence between cross-domain data in the transformed feature space with recognition guarantees. Aiming at better determining the cross-domain correspondences, we utilize the transformation $\mathbf{A} \in \mathbb{R}^{m \times k}$ derived by TCA [6] for projecting source and target data into the transformed space.

Instead of relating cross-domain data in the original feature space, our work is to associate transformed cross-domain data $f_s(\mathbf{x}) = \mathbf{A}^\top \mathbf{X_S}$ and $f_t(\mathbf{x}) = \mathbf{A}^\top \mathbf{X_T}$ by learning the correspondence transform $\mathbf{\Phi} \in \mathbb{R}^{k \times k}$. With the observed the similarity matrix $\mathbf{W}$, the introduction and learning of this transformation would allow us to identify cross-domain data pairs, while the class labels will be transferred from source to target domain for classification purposes.

To solve the above problem, we propose to solve the following optimization problem:

$$\min_{\mathbf{\Phi}} \sum_{i=1}^{N_s} \sum_{j=1}^{N_t} w_{ij} \parallel \mathbf{A}^\top \mathbf{x}_{s_i} - \mathbf{\Phi} \mathbf{A}^\top \mathbf{x}_{t_j} \parallel_F^2. \tag{4}$$

As noted above (and in Algorithm 1), since we apply the transformation of TCA for initializing $\mathbf{A}$, we do not need to apply additional constraints on $\mathbf{A}$ for avoiding trivial solution.

We see that (4) can be viewed as a robust scheme for determining distances between each cross-domain data pair. In other words, our proposed algorithm uniquely applies the concepts of instance reweighting for performing feature matching. Since we transfer the pseudo labels from source to target domains during adaptation, the resulting correspondence transformation would exhibit recognition capabilities. Later in our experiments, we will show that our proposed method would perform favorably against state-of-the-art unsupervised domain adaptation approaches.

2.3. Optimization

To jointly optimize $\mathbf{W}$ and $\mathbf{\Phi}$, we first rewrite (4) as follows:

$$\begin{aligned} \min_{\mathbf{\Phi}} tr(\mathbf{\Phi}(\sum_{i=1}^{N_s} \sum_{j=1}^{N_t} w_{ij}\ \mathbf{z}_{t_j} \mathbf{z}_{t_j}{}^\top)\mathbf{\Phi}^\top \\ - 2(\sum_{i=1}^{N_s} \sum_{j=1}^{N_t} w_{ij}\ \mathbf{z}_{s_i} \mathbf{z}_{t_j}{}^\top))\mathbf{\Phi}^\top + R, \end{aligned} \tag{5}$$

where $\mathbf{z}_{s_i} = \mathbf{A}^\top \mathbf{x}_{s_i}$ and $\mathbf{z}_{t_j} = \mathbf{A}^\top \mathbf{x}_{t_j}$. R denotes the term unrelated to $\mathbf{\Phi}$. For the sake of simplicity, we define

$$\tilde{\Sigma}_{st} = \sum_{i=1}^{N_s} \sum_{j=1}^{N_t} w_{ij}\ \mathbf{z}_{s_i} \mathbf{z}_{t_j}^\top \text{ and } \tilde{\Sigma}_{tt} = \sum_{i=1}^{N_s} \sum_{j=1}^{N_t} w_{ij}\ \mathbf{z}_{t_j} \mathbf{z}_{t_j}^\top,$$

and thus Φ can be derived by taking the partial derivatives:

$$\mathbf{\Phi} = (\tilde{\Sigma}_{st})(\tilde{\Sigma}_{tt})^{-1}. \tag{6}$$

Algorithm 1 Direct Distribution Matching

Input: X_s, X_t, y_s, dimension k, parameter λ.
Initialization: Projection matrix $\mathbf{A}$ derived by TCA, $\mathbf{\Phi} = \mathbf{I}$
  while Not Converge **do**
    Determine pseudo labels $\tilde{y_t}$ by source classifiers.
    Update $\mathbf{W}$ by (3) and Φ by solving (4)
  end while
  Classify transformed test data by nearest neighbor classifier
Output: Classified label y_t

Based on the above derivations, we update the pseudo labels $\tilde{y}_t$, the similarity matrix $\mathbf{W}$ and the correspondence transform $\mathbf{\Phi}$ iteratively for solving (4). Once both $\mathbf{W}$ and $\mathbf{\Phi}$ are obtained, recognition can be simply achieved by projecting test data $\mathbf{X_T}$ into the transformed space, followed by determining its correspondence/similarity to source-domain data. Our algorithm is summarized in Algorithm 1.

3. EXPERIMENTS

3.1. Cross-View Object Recognition

For solving this task, we consider the COIL-20 dataset [17] which consists of 20 objects with 1,440 images. Each object category contains 72 images which are taken on a turntable for 5 degrees apart. Each image is of size 32×32 with gray-scale pixels. Since the images are with plain background, we directly describe each image by a 1024 dimension vector. Following the setting of [7], the dataset are partitioned into two subsets, COIL1 and COIL2. COIL1 contains all images with objects rotated by $[0°, 85°] \cup [180°, 265°]$ and COIL2 contains objects rotated by $[90°, 175°] \cup [270°, 355°]$, so each subset has 720 images of 20 classes. As a results we have two cross-view pairs to be considered (i.e., COIL1$\Rightarrow$ COIL2 and COIL2$\Rightarrow$ COIL1). Obviously, images at different views would exhibit significant variations and thus make the recognition problem difficult.

For comparisons, several state-of-the-art unsupervised domain adaptation methods including TCA [6], JDA [7], TJM [14] are considered. We also perform direct classification (i.e., no domain adaptation) using source-domain classifiers (we use nearest neighbors (NN)). In our wok, we let $\lambda = 0.1$ and dimension $k = 30$ for all methods, and fix $\beta = 8$. The results are listed in the first two rows in Table 1, which show that our method clearly achieved improved cross-view recognition performance than other approaches.

3.2. Cross Domain Handwritten Digit Recognition

Next, we use USPS and MNIST datasets for evaluating our performance on cross-domain handwritten digit recognition. The former contains 7,291 training images and 2007 test images of size 16×16 pixels (of 10 categories), while the latter consists of 60,000 and 10,000 images of size 28×28 pixels for training and testing, respectively. Again, we follow the

Table 1. Performance comparisons for cross-domain visual classification. Note that we have $S \Rightarrow T$ indicate adaptation of data from source S to target domains T.

Methods	NN	TCA [6]	JDA [7]	TJM [14]	Ours
COIL1 ⇒ COIL2	83.61	88.47	93.75	89.86	**98.80**
COIL2 ⇒ COIL1	82.78	86.11	91.67	87.92	**96.70**
USPS ⇒ MNIST	44.70	53.05	60.00	51.35	**61.70**
MNIST ⇒ USPS	65.94	58.78	**72.11**	61.11	63.00

setting of [7], and randomly sample 1800 and 2000 images from USPS and MNIST respectively. Each image is represented by a 256 (e.g., 16×16) dimensional vector (in terms of their grayscale pixel values).

By utilizing the same recent/baseline unsupervised domain adaptation approaches (and settings) for comparisons, we list the recognition results in the bottom two rows of Table 1. From this table, we see that our approach achieved comparable or improved results over state-of-the-art methods.

3.3. Cross-Domain Object Recognition

Finally, we address the challenging task of cross-domain object recognition using Office [11] and Caltech-256 [18] datasets. The Office dataset consists of image from 31 object classes, which are collected from three different domains: Amazon, DSLR, and Webcam. The images of Amazon are collected from the Internet, those of DSLR are taken with high resolution cameras, while the webcam images are typically taken with low-resolution, over-exposed or blurred sensors. The *Caltech-256* dataset contains object images of 256 categories. As did in [5], we combine the above datasets and select the 10 shared categories to construct image data in four different domains: Amazon (A), DSLR (D), Webcam (W), and Caltech (C). As a result, a total of 12 different cross-domain pairs will be available for evaluation. Detailed settings such as the number of training images per category can be found in [5].

To describe each object image, we advance the $DeCAF_6$ features [19] to extract a 4,096-dimensional feature vector. As shown in [19], the DeCAF feature is able to achieve remarkable performance in generic image classification tasks. As for the parameters, we set $\lambda = 0.1$, dimension $k = 30$ (via PCA), and $\beta = 4$. The recognition performance and comparisons are presented in Table 2. As shown in this table, our proposed method performed favorably against state-of-the-art methods. Based on the above experiments, the effectiveness of our unsupervised domain adaptation approach for cross-domain visual classification can be successfully verified.

3.4. Remarks on Convergence

Lastly, we discuss the issue of convergence as mentioned in Section 2.3. Figure 2 shows the classification rates on two selected cross-domain pairs with increasing iteration numbers.

Table 2. Recognition results of cross-domain object recognition.

Methods	NN	TCA [6]	JDA [7]	TJM [14]	Ours
A ⇒ W	71.19	76.61	**84.07**	77.97	83.05
A ⇒ D	80.89	81.53	**84.71**	**84.71**	84.08
A ⇒ C	82.19	82.72	84.42	81.48	**87.36**
W ⇒ A	77.35	79.44	89.98	86.12	**92.07**
W ⇒ D	100.00	100.00	100.00	100.00	100.00
W ⇒ C	73.46	74.44	81.66	78.81	**85.40**
D ⇒ A	84.55	85.91	**92.17**	88.62	91.75
D ⇒ W	98.98	99.32	**100.00**	98.64	**100.00**
D ⇒ C	77.92	76.76	84.06	80.32	**85.84**
C ⇒ A	90.19	89.98	90.40	90.92	**93.01**
C ⇒ W	78.31	81.02	86.10	82.71	**95.93**
C ⇒ D	87.90	87.26	88.54	88.54	**91.72**
Average	83.58	84.58	88.84	86.57	**90.85**

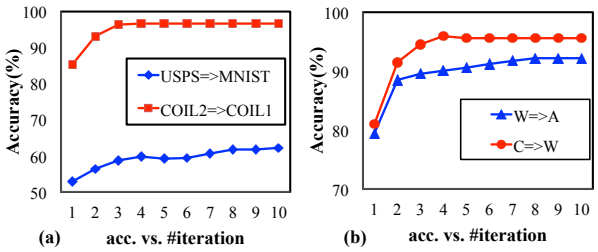

Fig. 2. Convergence analysis (i.e., accuracy vs. iteration number) for (a) cross-view object recognition and cross-domain handwritten digit recognition, and (b) cross-domain object recognition. Note that only one cross-domain pair for each is shown due to space limitation.

It can be seen that our proposed method achieved improved and converging performances within 5-10 iterations. This observation also applies to other cross-domain pairs in our experiments. Therefore, we successfully verify the convergence of the proposed method.

4. CONCLUSION

We proposed a novel unsupervised domain adaptation approach which jointly exploits the correspondence between cross-domain data and recovers the label information in the target domain. Inspired by instance reweighting and feature matching techniques for domain adaptation, our method is able to match cross-domain feature distributions at the instance level, while the contribution of each cross-domain pair can be properly and automatically identified. In our experiments, we successfully achieved improved results on the tasks of cross-view object recognition, cross-domain handwritten digit recognition, and cross-domain object recognition.

Acknowledgement

This work is supported in part by the Ministry of Science and Technology of Taiwan via MOST103-2221-E-001-021-MY2.

5. REFERENCES

[1] A. Torralba and A. A. Efros, "Unbiased look at dataset bias," in *IEEE CVPR*, 2011.

[2] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE TKDE*, 2010.

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

[4] 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.

[5] B. Gong, K. Grauman, and F. Sha, "Connecting the dots with landmarks: Discriminatively learning domain-invariant features for unsupervised domain adaptation," in *ICML*, 2013.

[6] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Trans. Neural Networks*, 2011.

[7] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *IEEE ICCV*, 2013.

[8] M. Sugiyama, S. Nakajima, H. Kashima, P. V. Buenau, and M. Kawanabe, "Direct importance estimation with model selection and its application to covariate shift adaptation," in *NIPS*, 2008.

[9] J. Blitzer, R. McDonald, and F. Pereira, "Domain adaptation with structural correspondence learning," in *EMNLP*, 2006.

[10] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars, "Unsupervised visual domain adaptation using subspace alignment," in *IEEE ICCV*, 2013.

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

[12] C. Zhang, Y. Zhang, S. Wang, J. Pang, C. Liang, Q. Huang, and Q. Tian, "Undo the codebook bias by linear transformation for visual applications," in *ACM MM*, 2013.

[13] Q. Qiu, V. M. Patel, P. Turaga, and R. Chellappa, "Domain adaptive dictionary learning," in *ECCV*, 2012.

[14] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer joint matching for unsupervised domain adaptation," in *IEEE CVPR*, 2014.

[15] A. Gretton, K. M. Borgwardt, M. Rasch, B. Schölkopf, and A. J. Smola, "A kernel method for the two sample problem," in *NIPS*, 2007.

[16] S. Gold, A. Rangarajan, C.-P. Lu, S. Pappu, and E. Mjolsness, "New algorithms for 2d and 3d point matching:: pose estimation and correspondence," *Pattern Recognition*, 1998.

[17] S. A. Nene, S. K. Nayar, H. Murase, et al., "Columbia object image library (coil-20)," Tech. Rep., Technical Report CUCS-005-96, 1996.

[18] G. Griffin, A. Holub, and P. Perona, "Caltech-256 object category dataset," 2007.

[19] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, "Decaf: A deep convolutional activation feature for generic visual recognition," *ICML*, 2014.