HETEROGENEOUS DOMAIN ADAPTATION WITH LABEL AND STRUCTURE CONSISTENCY

Yao-Hung Hubert Tsai[1] *Yi-Ren Yeh*[2] *Yu-Chiang Frank Wang*[1]

[1]Research Center for IT Innovation, Academia Sinica, Taipei, Taiwan
[2]Department of Mathematics, National Kaohsiung Normal University, Kaohsiung, Taiwan
y.h.huberttsai@gmail.com, yryeh@nknu.edu.tw, ycwang@citi.sinica.edu.tw

ABSTRACT

Domain adaptation is a challenging task, since it associates data collected from different domains or exhibiting distinct distributions. In this paper, we particularly focus on adapting cross-domain data with distinct feature dimensions or representations. Thus, this is referred to as the task of *heterogeneous domain adaptation* (HDA). To solve HDA, we propose *Label and Structure-consistent Unilateral Projection (LS-UP)* that transforms source-domain data to the target domain, with the goal of matching cross-domain data distribution and preserving data structure after projection. The main contribution of our work is its ability in relating cross-domain data with different feature representations. We evaluate our LS-UP for HDA on two different cross-domain classification problems, and we show that our method would perform favorably against state-of-the-art approaches.

Index Terms— Domain adaptation, object recognition, text categorization

1. INTORDUCTION

In many real-world classification applications, one might not be able to collect a sufficient amount of labeled data for training due to the cost of data collection or limited data availability. Recently, *domain adaptation (DA)* addresses this task by transferring the knowledge learned from one or multiple source domains, with the goal to solve the learning task in the target domain of interest.

Most existing DA approaches assume that data across domains lie in a homogeneous feature space (i.e., with the same type of feature) [1, 2, 3]. However, this assumption might not be held when source and target domain data are collected by different sensors or processed by different feature extraction techniques. Thus, in this paper, we focus on the task of *heterogeneous domain adaptation* (HDA), in which we associate cross-domain data with distinct feature dimensions or representations. We note that, for the HDA setting, labeled data can be collected in the source domain, but only few labeled ones can be observed in the target domain. Thus, how to associate such cross-domain data with limited and biased label information becomes a very challenging problem [4, 5, 6, 7, 8, 9].

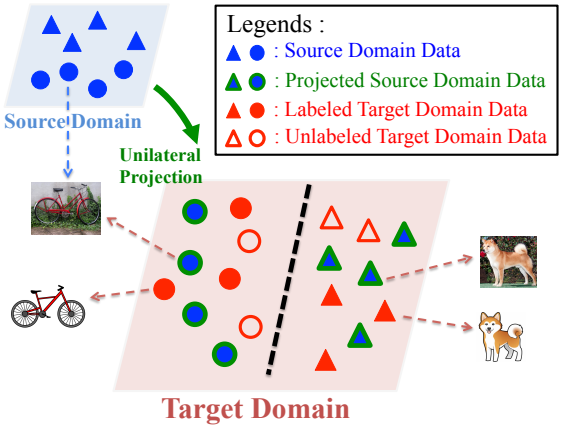

Fig. 1. Illustration of unilateral projection for heterogeneous domain adaptation (HDA).

In this paper, we propose *Label and Structure-consistent Unilateral Projection (LS-UP)* for HDA. The concept of unilateral projection is illustrated in Figure 1. As detailed later in Section 3, our LS-UP transforms the source-domain data to the target domain, and we associate and recognize cross-domain heterogeneous data in that domain accordingly. More specifically, our LS-UP associates cross-domain data by matching their class-conditional distributions, with an additional constraint enforcing label and data structure consistency. This is why improved adaptation and classification performance can be achieved. In Section 4, we will conduct experiments on two different cross-domain datasets to evaluate the performance of our method.

2. RELATED WORK

In this section, we provide a brief review of previous approaches for solving HDA problems. Generally, existing HDA approaches can be divided into two categories: *common space learning* and *domain transformation* based methods. For *common space learning* [4, 5, 6, 10, 11, 12, 13],

the goal is to find a pair of feature mappings that project source and target domain data onto a latent feature space, so that cross-domain data distributions could be matched for adaptation purposes. For example, Shi *et al.* [4] proposed a heterogeneous spectral mapping (HeMap) method that observed source and target projection matrices via spectral transformation (with no label information considered). Wang and Mahadevan [5] proposed a manifold alignment method (DAMA) by maximizing intra-class similarity and minimizing inter-class similarity. Duan *et al.* [6] proposed heterogeneous feature augmentation (HFA) that simultaneously derived a common feature space and a classification model based on maximum-margin techniques. Li *et al.* [10] extended HFA to a semi-supervised version (SHFA) that took unlabeled target domain data into consideration. Wu *et al.* [11] proposed a transfer discriminant analysis of canonical correlations (HTDCC) method to address HDA problems for cross-view action recognition. Xiao and Guo [12] presented a semi-supervised kernel matching method for heterogeneous domain adaptation (SSKMDA). Similarly, Xiao and Guo [13] considered a semi-supervised subspace co-projection (SCP) method that projected cross-domain instances onto a co-located subspace with prediction models jointly learned by labeled data.

On the other hand, *domain transformation* [8, 9, 14] methods directly learn a feature transformation from one domain to the other. For example, Kulis *et al.* [8] proposed an asymmetric regularized cross-domain transformation (ACR-t) method that transferred label information between source and target domain through a kernelized matrix learning. Hoffman *et al.* [9] presented an max-margin domain transform (MMDT) method that transformed target instances to source domain as well as learned a classification model. Zhou *et al.* [14] proposed a sparse heterogeneous feature representation (SHFR) algorithm that learned a sparse feature transformation matrix to map the weight vector of classifiers from source to target domain. Inspired by [8, 9, 14], our work aims at learning a transformation matrix that project source domain data to the target domain, with additional capabilities in enforcing label and structure consistency. In the following section, we will detail our proposed method.

3. OUR PROPOSED METHOD

3.1. Problem Settings

We first define the notation which will be used in this paper. Given source and target domain data as $D_S = \{\mathbf{x}_s^i, y_s^i\}_{i=1}^{n_s} = \{\mathbf{X}_S, \mathbf{y}_s\}$ and $D_T = \{\mathbf{x}_t^i, y_t^i\}_{i=1}^{n_t} = \{\mathbf{X}_T, \mathbf{y}_t\}$, where $\mathbf{X}_S \in \mathbb{R}^{d_s \times n_s}$ and $\mathbf{X}_T \in \mathbb{R}^{d_t \times n_t}$ represent n_s d_s-dimensional source-domain instances and n_t d_t-dimensional target-domain instances, respectively, and entries in $\mathbf{y}_s \in \mathbb{R}^{n_s \times 1}$ and $\mathbf{y}_t \in \mathbb{R}^{n_t \times 1}$ indicate their corresponding labels (from 1 up to C). For HDA, only few labeled data are available in the target domain, while a sufficient number of labeled data can be observed in the source domain.

For experiment purposes, the target domain data D_T can be further partitioned into labeled and unlabeled subset $\{D_L, D_U\}$ where $D_L = \{\mathbf{x}_l^i, y_l^i\}_{i=1}^{n_l} = \{\mathbf{X}_L, \mathbf{y}_l\}$ and $D_U = \{\mathbf{x}_u^i, y_u^i\}_{i=1}^{n_u} = \{\mathbf{X}_U, \mathbf{y}_u\}$. Among them, D_U will not be seen during training. It is worth mentioning that, in addition to a limited amount of target-domain labels, the feature representations for source and target-domain data are also different (i.e., $d_s \neq d_t$). In a nutshell, we are given d_s*-dimensional* source domain data D_S and few labeled d_t*-dimensional* target domain data D_L, and the goal of HDA is to predict the labels for d_t*-dimensional* data $\mathbf{X}_U$.

3.2. Label and Structure-consistent Unilateral Projection

To associate and recognize heterogeneous data across domains, we aim to learn a transformation matrix $\mathbf{A} \in \mathbb{R}^{d_s \times d_t}$ for projecting source-domain data to the target domain, while both label and consistency can be preserved for performance guarantees.

Let the new projected source-domain data as $\mathbf{A}^\top \mathbf{X}_S$, which can be viewed as additional labeled data in the target domain, and we need the distribution of such data to be matched to that of the target-domain data. That is, our goal is to solve the following optimization problem:

$$\min_{\mathbf{A}} E_C(\mathbf{A}, D_S, D_T) + E_S(\mathbf{A}, D_S), \tag{1}$$

where we have E_C as the *maximum mean discrepancy* (MMD) term for associating cross-domain class-conditional distribution, and E_S as the label and structure-preserving term.

For the class-conditional MMD term, we match the class-conditional data distributions of the projected source-domain and target-domain data (only the labeled ones) as suggested in [15]. That is, the empirical estimates of class-conditional means are applied to approximate such distributions, and thus E_C is calculated as

$$E_C(\mathbf{A}, D_S, D_T) = \sum_{c=1}^{C} \left\| \frac{1}{n_s^c} \sum_{i=1}^{n_s^c} \mathbf{A}^\top \mathbf{x}_s^{i,c} - \frac{1}{n_l^c} \sum_{i=1}^{n_l^c} \mathbf{x}_l^{i,c} \right\|^2 + \lambda \left\| \mathbf{A} \right\|^2, \tag{2}$$

where $\mathbf{X}_S^c = [\mathbf{x}_s^{1,c}, \mathbf{x}_s^{2,c}, \cdots, \mathbf{x}_s^{n_s^c,c}]$ and $\mathbf{X}_L^c = [\mathbf{x}_l^{1,c}, \mathbf{x}_l^{2,c}, \cdots, \mathbf{x}_l^{n_l^c,c}]$ denote source and labeled target domain data of class c, respectively, and $\{n_s^c, n_l^c\}$ indicate their corresponding numbers of data. Similar to [8, 9], we impose a regularizer $R(\mathbf{A}) = \lambda \left\| \mathbf{A} \right\|^2$ to prevent overfitting when learning the transformation $\mathbf{A}$.

For the label and structure-preserving term, we impose a class-wise locality constraint on the projected source data,

which emphasizes the structure of labeled source-domain data after projection. Thus, the term E_S is defined as

$$E_S(\mathbf{A}, D_S) = \sum_{i=1}^{n_s}\sum_{j=1}^{n_s} w_{ij} \left\|\mathbf{A}^\top \mathbf{x}_s^i - \mathbf{A}^\top \mathbf{x}_s^j\right\|^2, \quad (3)$$

where

$$w_{ij} = \begin{cases} \exp\left(-\left\|\mathbf{x}_s^i - \mathbf{x}_s^j\right\|^2 / \sigma^2\right) & \text{if } \{\mathbf{x}_s^i, \mathbf{x}_s^j\} \in \mathbf{X}_S^c \\ 0 & \text{otherwise} \end{cases}$$

denotes structural similarity between $\mathbf{x}_s^i$ and $\mathbf{x}_s^j$ (where σ is calculated by the standard deviation of source-domain data). It is worth noting that, different from [5], we particularly preserve within-class data similarity instead of the structure of the entire source-domain data. This would improve the classification ability after the adaptation is completed.

With (2) and (3), our proposed LS-UP is formulated as follows:

$$\begin{aligned}\min_{\mathbf{A}} \sum_{c=1}^{C} \left\| \frac{1}{n_s^c}\sum_{i=1}^{n_s} \mathbf{A}^\top \mathbf{x}_s^{i,c} - \frac{1}{n_l^c}\sum_{i=1}^{n_l^c} \mathbf{x}_l^{i,c} \right\|^2 + \lambda \|\mathbf{A}\|^2 \\ + \sum_{i=1}^{n_s}\sum_{j=1}^{n_s} w_{ij} \left\|\mathbf{A}^\top \mathbf{x}_s^i - \mathbf{A}^\top \mathbf{x}_s^j\right\|^2.\end{aligned} \quad (4)$$

By taking the derivative of (4) with respect to $\mathbf{A}$, the closed form and optimal solution $\mathbf{A}$ can be easily derived as

$$\mathbf{A} = \left(\lambda \mathbf{I}_{d_s} + \mathbf{X}_S \left(\mathbf{C} + \mathbf{S}\right) \mathbf{X}_S{}^\top\right)^{-1} \left(\mathbf{X}_S \mathbf{H} \mathbf{X}_L{}^\top\right), \quad (5)$$

where $\mathbf{I}_{d_s}$ is a d_s-dimensional identity matrix, matrices $\{\mathbf{C} \in \mathbb{R}^{n_s \times n_s}, \mathbf{H} \in \mathbb{R}^{n_s \times n_l}\}$ are derived by (2), and $\mathbf{S} \in \mathbb{R}^{n_s \times n_s}$ is calculated by (3). More precisely, we have

$$\mathbf{C}_{ij} = \begin{cases} \frac{1}{n_s^c n_s^c} & \text{if } \{\mathbf{x}_s^i, \mathbf{x}_s^j\} \in \mathbf{X}_S^c \\ 0 & \text{otherwise,} \end{cases}$$

$\mathbf{S} = \mathbf{D} - \mathbf{W}$ where $(\mathbf{W})_{ij} = w_{ij}$ and $\mathbf{D}$ is a diagonal matrix with $(\mathbf{D})_{ii} = \Sigma_j w_{ij}$, and

$$\mathbf{H}_{ij} = \begin{cases} \frac{1}{n_s^c n_l^c} & \text{if } \mathbf{x}_s^i \in \mathbf{X}_S^c \text{ and } \mathbf{x}_l^j \in \mathbf{X}_L^c \\ 0 & \text{otherwise.} \end{cases}.$$

Note that, following [8, 9], we fix $\lambda = \frac{1}{2}$ in our experiments.

3.3. Classification

After the optimal $\mathbf{A}$ of (4) is obtained, we now have new feature representation $\mathbf{Z}_S = \mathbf{A}^\top \mathbf{X}_S$ as the projected source-domain data in the target domain, and we denote $D_S^\bigstar = \{\mathbf{A}^\top \mathbf{x}_s^i, y_s^i\}_{i=1}^{n_s} = \{\mathbf{z}_s^i, y_s^i\}_{i=1}^{n_s} = \{\mathbf{Z}_S, \mathbf{y}_s\}$. Given projected source domain data $D_S^\bigstar$ and labeled target domain data D_L, standard classifier like SVM can be directly applied to recognize the remaining unlabeled target domain data $\{\mathbf{x}_u^i\}_{i=1}^{n_u}$ for completing the classification task.

Table 1. Classification results (%) with standard deviations for cross-domain object recognition using the *Office* dataset.

Source Domain	SVM$_t$	MMDT [9]	SHFA [10]	LS-UP
Amazon	52.3±2.1	59.7±2.8	60.0±2.8	**61.5±2.8**
Webcam		58.6±3.4	59.6±3.4	**61.1±3.4**

Table 2. Classification results (%) with standard deviations for multilingual text categorization using the *Reuters Multilingual* dataset (note that $m = 10$).

Source Domain	SVM$_t$	MMDT [9]	SHFA [10]	LS-UP
English	60.4±3.9	63.8±3.5	66.7±2.7	**68.5±2.5**
French		63.3±3.9	67.1±2.5	**69.1±2.6**
German		64.1±3.3	66.8±2.6	**67.8±2.6**
Italian		66.2±2.9	67.0±2.4	**68.5±2.6**

4. EXPERIMENTS

4.1. Dataset Settings

In this section, we evaluate our LS-UP on two HDA benchmark datasets: cross-domain object recognition and multilingual text categorization, as we describe below.

Office Dataset [3] is a dataset widely used for both homogeneous and heterogeneous domain adaptation (for the task of cross-domain object recognition). It contains $4,106$ images with 31 categories in three data domains: Amazon (Images downloaded from the Internet), DSLR (high-resolution images captured by digital SLR cameras), and Webcam (low-resolution images captured by Webcams).

To describe each image in this dataset, *SURF* interest points are first extracted from the images, followed by k-means clustering for converting each image into a Bag-of-Words (BoW) feature vector. Images in Amazon and Webcam are represented in 800-dimensional histogram features (i.e., 800 visual words in the source domain), and those in DSLR are represented in 600-dimensional histogram features (i.e., 600 visual words in the target domain).

Following [8, 9, 6, 10], Amazon and Webcam are selected as source domains, and DSLR is chosen as the target domain. For training, we randomly select 20 and 8 images per category in Amazon and Webcam as source-domain labeled images, and 3 images per category in DSLR as the labeled ones in the target domain. For testing, we use the remaining images in DSLR as the target-domain testing ones.

Multilingual Reuters Collection [16, 17] is a text dataset with about $11K$ news articles from 6 categories in 5 languages (i.e., English, French, German, Italian, and Spanish). All the articles are represented by bag-of-words model weighted by TF-IDF. Following the same setting of [6, 10],

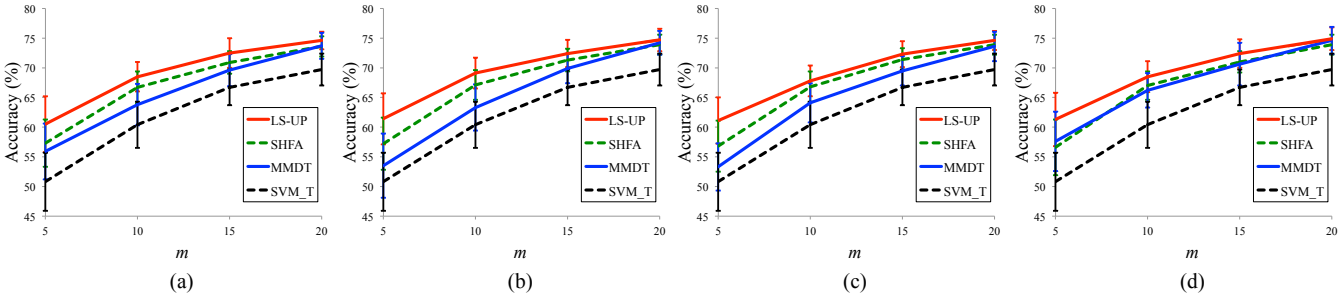

Fig. 2. Classification rates on *Reuters Multilingual* dataset of all methods with respect to different numbers of labeled target domain instances per category ($m = 5, 10, 15$ and 20). Note that Spanish is the target domain, and source domains are (a) English, (b) French, (c) German, and (d) Italian, respectively.

Spanish is the target domain while the articles of the other four different languages are regarded as the source domains.

For the training set, we randomly select 100 source domains articles from each categories and m labeled target articles from Spanish where $m = 5, 10, 15$, and 20. For testing, we randomly select $3,000$ articles from Spanish as the test data. Note that we adopt PCA (with 60% energy preserved) to reduce the feature dimensionality. Compared to the original high-dimensional TF-IDF feature, the dimensions of the reduced feature in each domain are $1,131$, $1,230$, $1,417$, $1,041$, and 807 for English, French, German, Italian, and Spanish, respectively.

4.2. Classification Results

We first compare our LS-UP with $\mathbf{SVM}_t$, which utilizes only labeled target domain data to train standard SVMs for classification purposes (i.e., no adaptation). Besides this baseline approach, we also compare our LS-UP with two state-of-the-art methods: (1) **MMDT** [9], a maximum-margin domain transform method for HDA, and (2) **SHFA**[10], a semi-supervised heterogeneous feature augmentation-based domain adaptation method. It is worth noting that, when conducting the experiments, we directly apply the released code of **MMDT** and **SHFA** for fair comparisons. In our evaluation, we randomly sample the training and test data as described in Section 4.1 by 20 times, and report the averaged classification accuracies as well as the corresponding standard deviations of all methods.

Object recognition:
Table 1 lists the averaged classification results for cross-domain object recognition of all methods. It is obvious that, without performing adaptation, $\mathbf{SVM}_t$ produces the lowest accuracy due to only a limited amount of labeled data is utilized in the target domain. When advancing domain adaptation, such as **MMDT**, **SHFA**, and ours, the results are improved by a clear margin. Since our LS-UP performs favorably against the two recent HDA approaches, it can be verified that our integration of matching class-conditional data distribution with label and structural consistency would be preferable for solving cross-domain object recognition.

Text categorization:
In Table 2, we report the classification results with picking $m = 10$ labeled target domain data. It can be seen that **MMDT** outperforms $\mathbf{SVM}_t$ with 4% improvement in classification rate. Comparing to **MMDT**, **SHFA** further improves 2.8%. Our LS-UP achieves the highest classification rate, and thus is preferable for HDA.

For the completeness of our evaluation, we plot the classification accuracies with different label numbers per class (i.e. $m = 5, 10, 15$, and 20) of labeled target domain data for each source domain in Figure 2. It can be observed that our proposed LS-UP method consistently outperforms other HDA approaches, especially when m is small. Therefore, the effectiveness of our proposed method in solving HDA problems can be successfully verified.

5. CONCLUSION

In this paper, we proposed Label and Structure-consistent Unilateral Projection (LS-UP) for solving HDA problems. Our LS-UP is able to project labeled source-domain data into the target domain, while the cross-domain data exhibit distinct feature dimensionality or distributions. In addition to matching cross-domain class-conditional data distribution, we further enforce the label and data consistency observed from the source domain. This is the reason why improved adaptation and classification performance can be jointly expected. Our experiments on object recognition and text categorization verify that our LS-UP achieves satisfactory results and performs favorably against recent HDA approaches.

ACKNOWLEDGEMENTS

This work was supported in part by the Ministry of Science and Technology of Taiwan under Grants MOST103-2221-E-001-021-MY2, MOST104-2221-E-017-016, and MOST104-2119-M-002-039.

6. REFERENCES

[1] Boqing Gong, Yuan Shi, Fei Sha, and Kristen Grauman, "Geodesic flow kernel for unsupervised domain adaptation," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2066–2073.

[2] Sinno Jialin Pan, Ivor W Tsang, James T Kwok, and Qiang Yang, "Domain adaptation via transfer component analysis," *Neural Networks, IEEE Transactions on*, vol. 22, no. 2, pp. 199–210, 2011.

[3] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell, "Adapting visual category models to new domains," in *Computer Vision–ECCV 2010*, pp. 213–226. Springer, 2010.

[4] Xiaoxiao Shi, Qi Liu, Wei Fan, Philip S Yu, and Ruixin Zhu, "Transfer learning on heterogenous feature spaces via spectral transformation," in *Data Mining (ICDM), 2010 IEEE 10th International Conference on*. IEEE, 2010, pp. 1049–1054.

[5] Chang Wang and Sridhar Mahadevan, "Heterogeneous domain adaptation using manifold alignment," in *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, 2011, vol. 22, p. 1541.

[6] Lixin Duan, Dong Xu, and Ivor W. Tsang, "Learning with augmented features for heterogeneous domain adaptation," in *Proceedings of the International Conference on Machine Learning*, Edinburgh, Scotland, June 2012, pp. 711–718, Omnipress.

[7] Wenyuan Dai, Yuqiang Chen, Gui-Rong Xue, Qiang Yang, and Yong Yu, "Translated learning: Transfer learning across different feature spaces," in *Advances in neural information processing systems*, 2008, pp. 353–360.

[8] Brian Kulis, Kate Saenko, and Trevor Darrell, "What you saw is not what you get: Domain adaptation using asymmetric kernel transforms," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 1785–1792.

[9] Judy Hoffman, Erik Rodner, Jeff Donahue, Trevor Darrell, and Kate Saenko, "Efficient learning of domain-invariant image representations," *International Conference on Learning Representations*, 2013.

[10] Wen Li, Lixin Duan, Dong Xu, and Ivor W Tsang, "Learning with augmented features for supervised and semi-supervised heterogeneous domain adaptation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 36, no. 6, pp. 1134–1148, 2014.

[11] Xinxiao Wu, Han Wang, Cuiwei Liu, and Yunde Jia, "Cross-view action recognition over heterogeneous feature spaces," in *Computer Vision (ICCV), 2013 IEEE International Conference on*. IEEE, 2013, pp. 609–616.

[12] Min Xiao and Yuhong Guo, "Feature space independent semi-supervised domain adaptation via kernel matching," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 37, no. 1, pp. 54–66, 2015.

[13] Min Xiao and Yuhong Guo, "Semi-supervised subspace co-projection for multi-class heterogeneous domain adaptation," in *Machine Learning and Knowledge Discovery in Databases*, pp. 525–540. Springer, 2015.

[14] Joey Tianyi Zhou, Ivor W Tsang, Sinno Jialin Pan, and Mingkui Tan, "Heterogeneous domain adaptation for multiple classes," in *International Conference on Artificial Intelligence and Statistics*, 2014, pp. 1095–1103.

[15] Mingsheng Long, Jianmin Wang, Guiguang Ding, Jiaguang Sun, and Philip S Yu, "Transfer feature learning with joint distribution adaptation," in *Computer Vision (ICCV), 2013 IEEE International Conference on*. IEEE, 2013, pp. 2200–2207.

[16] Massih-Reza Amini, Nicolas Usunier, and Cyril Goutte, "Learning from multiple partially observed views - an application to multilingual text categorization," in *NIPS 22*, 2009.

[17] Nicola Ueffing, Michel Simard, Samuel Larkin, and J. Howard Johnson, "NRC's PORTAGE system for WMT 2007," in *In ACL-2007 Second Workshop on SMT*, 2007, pp. 185–188.