Multi-Similarity Based Multi-Source Transfer Learning and Its Applications

Zhen Liu, Jun-an Yang, Hui Liu, and Wei Wang
Electronic Engineering Institute, Hefei 230037, China
Key Laboratory of Electronic Restriction of Anhui Province, Hefei 230037, China
Email: {ahulz, liuhui1983eei}@163.com; yangjunan@ustc.edu.cn; wwei009@mail.ustc.edu.cn

Abstract —In this paper, a novel multi-source transfer learning method based on multi-similarity ($(MS)^2TL$) is proposed. First, we measure the similarities between domains at two levels, i.e., "domain-domain" and "sample-domain". With the multi-similarities, $(MS)^2TL$ can explore more accurate relationship between the source domains and the target domain. Then, the knowledge of the source domains is transferred to the target based on the smoothness assumption, which enforces the requirement that the target classifier shares similar decision values with the relevant source classifiers on the unlabeled target samples. $(MS)^2TL$ can increase the chance of finding the sources closely related to the target to reduce the "negative transfer" and also imports more knowledge from multiple sources for the target learning. Furthermore, $(MS)^2TL$ only needs the pre-learned source classifiers when training the target classifier, which is suitable for large datasets. We also employ a sparsity-regularizer based on the ε-insensitive loss to enforce the sparsity of the target classifier with the support vectors only from the target domain such that the label prediction on any test sample is very fast. We also use the ε-insensitive loss function to enforce the sparsity of the decision function for fast label prediction. Validation of $(MS)^2TL$ is performed with toy and real-life datasets. Experimental results demonstrate that $(MS)^2TL$ can more effectively and stably enhance the learning performance. Finally, $(MS)^2TL$ is also applied to the communication specific emitter identification task and the result is also satisfying.

Index Terms—Transfer learning, multiple source transfer, manifold assumption

I. Introduction

Transfer learning [1]-[2] can effectively exploit and transfer the knowledge from different but similar source domains for target domain learning. Recently, transfer learning has been applied to many real-world applications, such as text processing [3], computer vision [4]-[5], network identification [6], automatic control [7], etc.

For the single-source domain setting, much work has been developed [1]. In general, the effectiveness of the knowledge transfer from a source domain to the target domain depends on how well they are related. The stronger the relationship, the more usable will be the source knowledge. Often in practice, one may be offered more than one source domain for learning. If we only use one source domain for learning, it is wasteful and we also can't ensure that the selected source domain is well related with the target domain. Brute force transferring in case of weak relationships may lead to performance deterioration of the target domain learning, i.e., "negative transfer". In this paper, we propose a novel multi-source transfer learning method called $(MS)^2TL$ (Multi-Similarity based Multi-Source Transfer Learning). $(MS)^2TL$ explores the relationships between the source domains and the target domain by multi-similarity metric. Then, the knowledge of the source domains is transferred to the target based on the smoothness assumption, which enforces that the target classifier shares similar decision values with the relevant source classifiers on the unlabeled target samples.

We summarize the main contributions of this paper as follows: We propose a novel multi-source transfer learning method called $(MS)^2TL$, which can not only improve the ability to avoid the problem of "negative transfer" but also explore more knowledge from the source domains for the target domain learning. In $(MS)^2TL$, we measure the similarities between domains at two levels, i.e., "domain-domain" and "sample-domain". With the multi-similarities, we then define a multi-source transfer manifold regularizer and add it into the optimal function of $(MS)^2TL$ for knowledge transfer. We also use the ε-insensitive loss function to enforce the sparsity of the decision function for fast label prediction. Furthermore, $(MS)^2TL$ only needs the pre-learned source classifiers when training the target classifier, which is suitable for large datasets. $(MS)^2TL$ can be readily introduced to many kernel methods and extend these methods to the corresponding transfer learning methods [8]. In this paper, we give our method under the framework of least square SVM (LS-SVM) [9]. We evaluate our method in two multiple transfer learning related applications, i.e., target recognition and document retrieval. Experimental results demonstrate that $(MS)^2TL$ can more effectively and stably enhance the learning performance. Finally, the proposed algorithm is applied to the communication specific emitter identification task and the result is also satisfying.

The rest of the paper is organized as follows: In Section II, we briefly review the related work; In Section III, the proposed method $(MS)^2TL$ is introduced; In

Manuscript received January 5, 2016; revised June 20, 2016.
This work was supported by the National High-tech R&D Program (863 Program) of China, and Anhui Provincial Natural Science Foundation (NO.1308085QF99, NO.1408085MKL46).
Corresponding author email: ahulz@163.com.
doi:10.12720/jcm.11.6.539-549

Section IV, extensive experiments are performed; Some conclusions are given in Section V.

II. Related Works

Chattopadhyay *et al*. [10] proposed a weighting scheme which gives higher weights to those source domains with similar conditional probability distributions to the target. Based on [10], Sun *et al*. [11] proposed a two-stage transfer methodology in which the source samples are first weighted based on the marginal probability differences and then re-weighted by the weighting scheme in [10].

Ref. [10] and [11] use the source domain samples to train a target classifier whenever a new task is conducted. It is not efficient when the size of dataset is large. A more efficient way is to train a classifier in each source domain and combine these source classifiers based on the relationships between the source domains and the target domain. Schweikert *et al*. [12] proposed a multi-source transfer learning algorithm by combining the pre-learned source classifiers and target classifier through a so-called multiple convex combination. Yang *et al*. [13] proposed adaptive support vector machine (A-SVM) to learn a new SVM classifier for the target domain, which is adapted from the existing classifiers trained with the source domain samples.

However, in [13], equal weights were used for all source classifiers without considering the differences among the source domains. Besides, numerous unlabeled samples in the target domain are also not exploited in A-SVM. Duan *et al*. [14] proposed the data-dependent regularizer and proposed a multi-source transfer learning method called Domain Adaptation Machine (DAM). DAM could assign different weights based on the similarities between the source domains and the target domain and use all the samples for learning.

III. Multi-Similarity Based Multi-Source Transfer Learning

Let us represent the sth source domain as $D^s = \left\{(\mathbf{x}_i^s, y_i^s)\right\}_{i=1}^{N_s}$, where $\boldsymbol{x}_i^s$ and y_i^s are the ith feature vector and the corresponding label respectively, N_s is the number of instances in the sth source data set D^s, $s = 1, \cdots, M$ and M is the number of source domains. In the target domain, the labeled set is $D_l^T = \left\{(\mathbf{x}_i^T, y_i^T)\right\}_{i=1}^{N_l}$ while the unlabeled set is $D_u^T = \left\{\mathbf{x}_i^T\right\}_{i=N_l+1}^{N_l+N_u}$. The whole target set is denoted as $D^T = D_l^T \cup D_u^T$ with the size $N_T = N_l + N_u$. In the problem setting, the joint distributions $P(\mathbf{x}, y)$ of the feature vector $\mathbf{x}$ with its label y among domains are different.

Here, we assume f^T is the target domain classifier and f^s is the pre-learned source domain classifier in the sth source domain. Any type of classifier can be readily used as the source classifier. For the target domain sample $\boldsymbol{x}_i^T$, we denote the decision values as $f_i^T = f^T(\boldsymbol{x}_i^T)$ and $f_i^s = f^s(\boldsymbol{x}_i^T)$.

Duan *et al* [14] proposed DAM which simultaneously minimizes the loss of the labeled training data from the target domain as well as a data-dependent regularizer defined on the unlabeled data. The proposed framework DAM is then formulated as follows:

$$\min_{f^T} \quad \Omega(f^T) + \lambda_L \Omega_L(f^T) + \lambda_D \Omega_D(f^T) \tag{1}$$

where $\Omega(f^T)$ is a regularizer to control the complexity of the target domain classifier f^T, $\Omega_L(f^T)$ is a loss function of f^T on the labeled samples of target domain, $\Omega_D(f^T)$ is the data-dependent regularizer defined on the unlabeled samples of target domain, and $\lambda_L, \lambda_D > 0$ are the regularization parameters.

In DAM, the key of knowledge transfer is the data-dependent regularizer $\Omega_D(f^T)$.

$$\Omega_D(f^T) = \frac{1}{2}\sum_{s=1}^{M}\gamma_s \sum_{i=N_l+1}^{N_T}(f_i^T - f_i^s)^2 \tag{2}$$

where γ_s is the similarity weight of the sth source D^s and is computed as follows

$$\gamma_s = \exp(-MMD^2(D^s, D^T)/\beta_1) \tag{3}$$

where $MMD^2(D^s, D^T) = \left\| \frac{1}{N_T}\sum_{i=1}^{N_T}\phi(\boldsymbol{x}_i^T) - \frac{1}{N_s}\sum_{j=1}^{N_s}\phi(\boldsymbol{x}_i^s) \right\|_{\mathcal{H}}^2$ is the maximum mean discrepancy (MMD) [15] for measuring the data distributions between the sth source domain and the target domain. *MMD* is an effective nonparametric distance metric for comparing data distributions in the reproducing kernel Hilbert space. $\beta_1 > 0$ is the bandwidth parameter to control the spread of *MMD* and is usually fixed as the mean of *MMD* among domains.

DAM uses γ_s to measure the relationships between the source domains and the target domain. However, the γ_s only consider the relationships between domains as a whole. The similarity measurement is not enough detailed and accurate. As we know, it is important to find and measure the relationships between the source domains and the target domain for transfer learning. To explore the relationships between domains better, we give a multi-similarity measurement at two levels, i.e., "domain-domain" and "sample-domain". With the defined multi-similarity weights, we modify the regularizer in (2) and then give our method (MS)^{2}TL under the framework of DAM.

A. Multi-Similarity

In Fig. 1, the triangles and the circles represent one class respectively. Fig. 1 shows that the classification model learned in the biased source domain is not reliable in the target domain. The key of transfer learning is to find and measure the relationships between the source domains and the target domain. Here we measure the similarities between domains at two levels, i.e., "domain-domain" and "sample-domain".

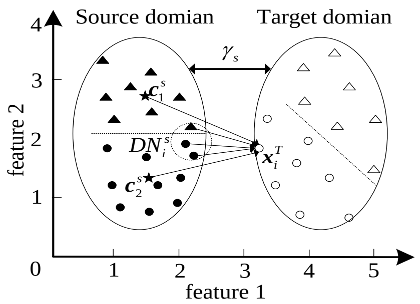


Fig. 1. The multi-similarity

Firstly, we concern the overall similarities between the source domains and the target domain, i.e., the similarity at the level of "domain-domain". Here, we use the similarity weight γ_s in (3) as the measurement.

Since the samples in the target domain are different, their relevancies to a source domain are also different. To describe the relationship between the target domain and the source domains further in detail, we concern the similarities at the level of "sample-domain". Here, two kinds of distance are first given: the average distance in the neighborhood (i.e., DN_i^s) and the minimum distance to the class center (i.e., DC_i^s).

DN_i^s is the average distance of the target sample $\boldsymbol{x}_i^T$ to its neighbors in the sth source domain D^s.

$$DN_i^s = \frac{1}{N_k}\sum_{k=1}^{N_k} d(\boldsymbol{x}_i^T, \boldsymbol{x}_k^s) \tag{4}$$

where $\boldsymbol{x}_k^s$ is the kth neighbor of $\boldsymbol{x}_i^T$ in D^s, N_k is the number of neighbors, $d(\bullet)$ is a general distance metric. If DN_i^s is small, $\boldsymbol{x}_i^T$ is more likely to occur in D^s and thus more similar to D^s.

DC_i^s is the minimum distance of the target domain sample $\boldsymbol{x}_i^T$ to the class centers in the sth source domain D^s.

$$DC_i^s = \min_j d(\boldsymbol{x}_i^T, \boldsymbol{c}_j^s) \tag{5}$$

where $\boldsymbol{c}_j^s$ is the mean of jth class samples in D^s. If DC_i^s is small and $\boldsymbol{x}_i^T$ is most close to the jth class center of D^s, $\boldsymbol{x}_i^T$ probably belongs to the jth class in D^s.

The distances defined in (4)-(5) measure the similarity from the marginal and conditional distribution of the data respectively. Combining them together, we have the following similarity weight A_{is} of $\boldsymbol{x}_i^T$ in the sth source domain at the level of "sample-domain"

$$A_{is} = \exp(-(d_i^s)^2 / \beta_2) \tag{6}$$

where $d_i^s = 0.5\times(DN_i^s + DC_i^s)$, $\beta_2 > 0$ is the bandwidth parameter to control the spread of d_i^s and is usually fixed as the mean of d_i^s in the whole unlabeled target set.

With (3) and (6), we defined the similarities between the source domains and the target domain at the two levels of "domain-domain" and "sample-domain". The multi-similarities can measure the distribution relevance in more detail, which is in favor of the transfer learning.

B. *Multi-Source Transfer Manifold Regularizer*

Belkin *et al*. [16] proposed the manifold assumption which enforces the decision function to be smooth on the data manifold, namely, the two samples in a high-density region should share similar decision values. Motivated from the manifold assumption, we similarly assume that the target domain classifier f^T should have similar decision values on the unlabeled target samples with the pre-learned classifiers from the relevant source domains. Based on the similarities defined in section III.A, the multi-source transfer manifold regularizer $\Omega_M(f^T)$ is given as follows

$$\Omega_M(f^T) = \frac{1}{2}\sum_{s=1}^{M}\gamma_s \sum_{i=N_l+1}^{N_T} A_{is}(f_i^T - f_i^s)^2 \tag{7}$$

where γ_s and A_{is} are the similarity weights defined in (3) and (6) respectively. As shown in Fig. 2, the regularizer $\Omega_M(f^T)$ builds the connections between the sources and the target through the similarity weights γ_s and A_{is}. If γ_s and A_{is} are large, the decision value of f^T and f^s on $\boldsymbol{x}_i^T$ will be similar. Thus, we can transfer the knowledge from the sources to the target under this assumption of "domain relevance-decision constraint".

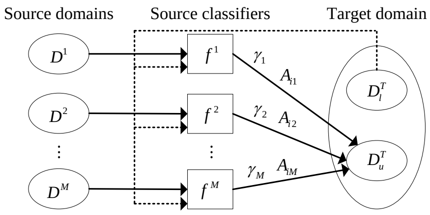


Fig. 2. Transfer learning based on the multi-similarities

C. *The Solution*

The minimizer of the optimization problem in (1) admits a form of $f^T(\boldsymbol{x}) = \boldsymbol{w}'\phi(\boldsymbol{x}) + b$ and then the regularizer $\Omega(f^T) = \|\boldsymbol{w}\|^2/2$. In addition, $\Omega_L(f^T)$ is modeled as the square error of the target domain classifier f^T on the labeled target samples, which is analogous to the LS-SVM [9]. Under the framework of DAM in (1), the optimal function of (MS)^{2}TL is then formulated as follows

$$\begin{aligned}\min_{f^T} \; & \frac{1}{2}\|\boldsymbol{w}\|^2 + \frac{\lambda_L}{2}\sum_{i=1}^{N_l}(f_i^T - y_i^T)^2 \\ & + \frac{\lambda_D}{2}\sum_{s=1}^{M}\gamma_s \sum_{i=N_l+1}^{N_T} A_{is}(f_i^T - f_i^s)^2\end{aligned} \tag{8}$$

To solve (8) efficiently, the ε-insensitive loss function in SVR [17] is introduced into (8). Then, we rewrite (8) as follows

$$\begin{aligned}\min_{f_i^T,\mathbf{w},b} \; & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N_T}\ell_\varepsilon(\mathbf{w}'\phi(\mathbf{x}_i)+b-f_i^T)+ \\ & \frac{\lambda_L}{2}\sum_{i=1}^{N_l}(f_i^T - y_i^T)^2 + \frac{\lambda_D}{2}\sum_{s=1}^{M}\gamma_s \sum_{i=N_l+1}^{N_T} A_{is}(f_i^T - f_i^s)^2\end{aligned} \tag{9}$$

where the ε-insensitive loss function $\ell_\varepsilon(t) = |t| - \varepsilon$ if $|t| > \varepsilon$, otherwise 0. C is the regularization parameter. Since $\ell_\varepsilon(\bullet)$ is non-smooth, (9) is usually transformed as a constrained optimization problem

$$\begin{aligned}\min_{f_i^T,\boldsymbol{w},b,\xi_i,\xi_i^*} \; & \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N_T}(\xi_i+\xi_i^*) \\ & + \frac{\lambda_L}{2}\sum_{i=1}^{N_l}(f_i^T - y_i^T)^2 + \frac{\lambda_D}{2}\sum_{s=1}^{M}\gamma_s \sum_{i=N_l+1}^{N_T} A_{is}(f_i^T - f_i^s)^2\end{aligned} \tag{10}$$

$$s.t. \quad \boldsymbol{w}'\phi(\boldsymbol{x}_i^T)+b-f_i^T \le \varepsilon+\xi_i, \xi_i \ge 0 \tag{11}$$

$$f_i^T - \boldsymbol{w}'\phi(\boldsymbol{x}_i^T) - b \le \varepsilon+\xi_i^*, \xi_i^* \ge 0 \tag{12}$$

$$i = 1,\cdots,N_T$$

By introducing the Lagrange multipliers α_i's and η_i's (resp., α_i^*'s and η_i^*'s) for the constraints in (11) (resp., (12)), we obtain the following Lagrangian

$$\begin{aligned}L = \; & \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N_T}(\xi_i+\xi_i^*) + \frac{\lambda_L}{2}\sum_{i=1}^{N_l}(f_i^T - y_i^T)^2 \\ & + \frac{\lambda_D}{2}\sum_{s=1}^{M}\gamma_s \sum_{i=N_l+1}^{N_T} A_{is}(f_i^T - f_i^s)^2 \\ & - \sum_{i=1}^{N_T}\alpha_i(\varepsilon+\xi_i+f_i^T-\boldsymbol{w}'\phi(\boldsymbol{x}_i^T)-b) - \sum_{i=1}^{N_T}\eta_i\xi_i \\ & - \sum_{i=1}^{N_T}\alpha_i^*(\varepsilon+\xi_i^*-f_i^T+\boldsymbol{w}'\phi(\boldsymbol{x}_i^T)+b) - \sum_{i=1}^{N_T}\eta_i^*\xi_i^*\end{aligned} \tag{13}$$

Setting the derivatives of (13) w.r.t. the primal variables ($\boldsymbol{w}$, b, ξ_i, ξ_i^*, f_i^T) to zeros, respectively, we have

$$\begin{aligned}&\frac{\partial L}{\partial \boldsymbol{w}} = 0 \Rightarrow \boldsymbol{w} = \sum_{i=1}^{N_T}\phi(\boldsymbol{x}_i^T)(\alpha_i^*-\alpha_i) = -\boldsymbol{\Phi}(\boldsymbol{\alpha}-\boldsymbol{\alpha}^*) \\ &\frac{\partial L}{\partial b} = 0 \Rightarrow \boldsymbol{1}_{N_T}'\boldsymbol{\alpha} = \boldsymbol{1}_{N_T}'\boldsymbol{\alpha}^* \\ &\frac{\partial L}{\partial \xi_i} = 0 \Rightarrow C = \alpha_i + \eta_i \\ &\frac{\partial L}{\partial \xi_i^*} = 0 \Rightarrow C = \alpha_i^* + \eta_i^* \\ &\frac{\partial L}{\partial f_i^T} = 0 \Rightarrow f_i^T = \tilde{y}_i + q_i(\alpha_i - \alpha_i^*)\end{aligned} \tag{14}$$

where $\boldsymbol{1}_{N_T}$ is a column vector of all ones with size N_T, $\boldsymbol{\alpha} = [\alpha_1,\cdots,\alpha_{N_T}]'$, $\boldsymbol{\alpha}^* = [\alpha_1^*,\cdots,\alpha_{N_T}^*]'$, $\boldsymbol{\Phi} = [\phi(\boldsymbol{x}_1^T),\cdots,\phi(\boldsymbol{x}_{N_T}^T)]$. If $i = 1,\cdots,N_l$, $q_i = 1/\lambda_L$ and $\tilde{y}_i = y_i^T$. If $i = N_l+1,\cdots,N_T$, $q_i = \frac{1}{\lambda_D\sum_{s=1}^{M}\gamma_s A_{is}}$ and $\tilde{y}_i = \frac{1}{\sum_{s=1}^{M}\gamma_s A_{is}}\sum_{s=1}^{M}\gamma_s A_{is} f_i^s$.

Substituting them back into (13), we arrive at the following dual formulation

$$\begin{aligned}\min_{\boldsymbol{\alpha},\boldsymbol{\alpha}^*} \; & \frac{1}{2}(\boldsymbol{\alpha}-\boldsymbol{\alpha}^*)'\tilde{\mathbf{K}}(\boldsymbol{\alpha}-\boldsymbol{\alpha}^*) + \tilde{\mathbf{y}}'(\boldsymbol{\alpha}-\boldsymbol{\alpha}^*) + \varepsilon\mathbf{1}_{N_T}'(\boldsymbol{\alpha}+\boldsymbol{\alpha}^*) \\ s.t. \; & \mathbf{1}_{N_T}'\boldsymbol{\alpha} = \mathbf{1}_{N_T}'\boldsymbol{\alpha}^*, \mathbf{0}_{N_T} \le \boldsymbol{\alpha},\boldsymbol{\alpha}^* \le C\mathbf{1}_{N_T}\end{aligned} \tag{15}$$

where $\tilde{\boldsymbol{K}} = \boldsymbol{K} + diag(\boldsymbol{q})$, $\boldsymbol{K} = \boldsymbol{\Phi}'\boldsymbol{\Phi}$, $\boldsymbol{q} = [q_1,\cdots,q_{N_T}]'$.

Since the dual form of (15) is similar to the dual of ε-SVR [18], the objective function of (MS)^{2}TL in (9) can be solved efficiently by using state-of-the-art SVM solvers such as LIBSVM [19]. For any test sample $\boldsymbol{x}$, the decision value of the target classifier f^T is

$$\begin{aligned}f^T(\boldsymbol{x}) &= \boldsymbol{w}'\phi(\boldsymbol{x}) + b \\ &= \sum_{i:a_i-a_i^*\ne 0}(a_i-a_i^*)k(\boldsymbol{x}_i^T,\boldsymbol{x}) + b\end{aligned} \tag{16}$$

which is a linear combination of $k(\mathbf{x}_i^T,\mathbf{x})$'s without involving any base classifiers. According to the Karush–Kuhn–Tucker (KKT) conditions in (14), if the target sample $\mathbf{x}_i^T$ has the value $|\mathbf{w}'\phi(\mathbf{x}_i)+b-f_i^T| < \varepsilon$, then its corresponding coefficient $(a_i - a_i^*)$ in (16) becomes zero. Therefore, the computation for the prediction can be greatly reduced by using the sparse representation in (16).

IV. Experiments and Discussions

If we only concern the similarity at the level of "domain-domain", namely, set all the A_{is} equal to 1, (MS)^{2}TL would change to be similar to the DAM algorithm [14]. Considering their relations, we compare our method (MS)^{2}TL with DAM in the experiments. To show the improvement by the transfer leaning progress, we also compare our (MS)^{2}TL with a non-transfer learning classification strategy represented as "Base" in the experiments. The Base means that the source domain classifiers f^s's are used to predict the unlabeled target samples directly and the average accuracy is the final result.

To demonstrate that (MS)^{2}TL can use different type of classifier as source classifiers f^s, we also conduct experiments with three types of source classifiers respectively, i.e., LS-SVM, C4.5, and Naïve Bayes.

In the experiments, we need to fix some parameters empirically, i.e., the number of neighbors (i.e., N_k) in (4), and the regularization parameters C, λ_L, and λ_D in (9). In the default setting, we set $\lambda_L = \lambda_D = 1$, $C = 1$, and $N_k = 8$. The parameters in the source classifiers are set as the default values in Weka [20]. Gaussian kernel (i.e., $k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp(-\|\boldsymbol{x}_i, \boldsymbol{x}_j\|^2 / (2\sigma^2))$ is used as the default kernel in which the kernel parameter σ is set as the mean distance between samples in the target domain. In the target domain, n samples per class are randomly selected as the labeled target set for the experiments. n is tuned in the range [0, 2, 4, 6, 10, 15, 20]. The experiments are repeated for 20 times with different source samples and target samples. The average classification accuracy is used as the evaluation measure.

In order to fully evaluate the algorithm performance, we evaluate (MS)^{2}TL for two multi-source transfer learning related applications: 1) target recognition, and 2) document retrieval. Finally, (MS)^{2}TL is also applied to the communication specific emitter identification task.

A. Experiments on Target Recognition

The experiments on target recognition use three datasets as the sources [21]: Amazon (images downloaded from online merchants), Webcam (low-resolution images by a web camera), and DSLR (high-resolution images by a digital SLR camera). We regard each dataset as a source domain. Caltech-256 [22] is used as the target domain. There are totally 10 classes which are common among all four datasets: Backpack, Touring-bike, Calculator, Head, Phones, Computer-keyboard, Laptop-101, Computer-monitor, Computer-mouse, Coffee-mug, and Video-projector. There are 8 to 151 samples per category per domain, and 2533 images in total. Fig. 3 highlights the differences among these domains with example images from the category of Computer-monitor.

(a) Amazon (b) Webcam

(c) DSLR (d) Caltech-256

Fig. 3. Example images from the Computer-monitor category in different domains

We extract the 4096 dimensional DeCAF$_6$ features [23] from the raw images. Then, these features from different domains are used to learn a classification model for the target domain. The classification accuracies of the proposed method compared with the Base and DAM are recorded in Table I. The highest accuracy among different methods is highlighted in bold.

TABEL I: CLASSIFICATION ACCURACIES ON TARGET RECOGNITION EXPERIMENTS

n	C4.5			Naïve Bayes			LS-SVM		
	Base	DAM	(MS)^{2}TL	Base	DAM	(MS)^{2}TL	Base	DAM	(MS)^{2}TL
0	0.9202	0.9319	**0.9427**	0.6836	0.7254	**0.8000**	0.5125	0.5125	**0.5482**
2	0.9404	0.9625	**0.9795**	0.7023	0.7622	**0.7708**	0.4902	0.4902	**0.5355**
4	0.9559	0.9701	**0.9876**	0.6764	0.7239	**0.7764**	0.5465	0.5519	**0.5677**
6	0.9103	0.9439	**0.9622**	0.6534	0.7505	**0.8528**	0.5940	0.6974	**0.7085**
10	0.9342	0.9672	**0.9744**	0.6275	0.8577	**0.9483**	0.6124	0.7992	**0.8602**
15	0.9316	0.9747	**0.9753**	0.6488	0.9515	**0.9555**	0.5865	0.9121	**0.9504**
20	0.9371	0.9760	**0.9772**	0.7015	0.9708	**0.9798**	0.5753	0.9477	**0.9657**
Ave.	0.9328	0.9609	**0.9713**	0.6705	0.8203	**0.8691**	0.5596	0.7015	**0.7337**

In Table I, (MS)^{2}TL can effectively improve the classification accuracy regardless of source classifiers and the number of the labeled target samples. The experimental results demonstrate that (MS)^{2}TL could better explore the relevant relationship between the source domains and the target domain, and transfer more knowledge from the source domains to promote the target domain learning. As the Base method uses the source classifiers directly without considering the difference between domains, its classification results are always bad. Besides, the average accuracies of (MS)^{2}TL are also higher than Base and DAM (last row in each table). In addition, it is can be found that the accuracies of (MS)^{2}TL and DAM generally increase along with the increasing of the number of the labeled target domain samples.

B. Experiments on Document Retrieval

In this section, the experiments are conducted for the application of document retrieval. The experimental dataset is the 20 Newsgroups dataset [24] which contains 18774 documents, and has a hierarchical structure with 6 main categories and 20 subcategories. To use the dataset for the purpose of multi-source transfer learning experiments, we regard the subcategories per main category as the samples of the common class in different domains. We choose the samples from three main categories with at least four subcategories and generate three settings to evaluate the algorithms (see Table II for the detailed settings). In every setting, we consider one main category as the positive class and use another one as the negative class, and employ all the samples from two subcategories to construct one domain.

TABEL II: DESCRIPTION OF THE 20 NEWSGROUPS DATASET

Settings	Source domains	Target domain
rec vs. *sci*	rec.autos & sci.crypt rec.motorcycles & sci.electronics rec.sport.baseball & sci.med	rec.sport.hockey & sci.space
comp vs. *rec*	comp.graphics & rec.autos comp.os.ms-windows.misc & rec.motorcycles comp.sys.ibm.pc.hardware & rec.sport.basebal	comp.sys.mac.hardware & rec.sport.hockey
sci vs. *comp*	sci.crypt & comp.graphics sci.electronics & comp.os.ms-windows.misc sci.med & comp.sys.ibm.pc.hardware	sci.space & comp.sys.mac.hardware

TABLE III: CLASSIFICATION ACCURACIES ON DOCUMENT RETRIEVAL EXPERIMENTS

n	*rec* vs. *sci*			*comp* vs. *rec*			*sci* vs. *comp*		
	Base	DAM	(MS)^{2}TL	Base	DAM	(MS)^{2}TL	Base	DAM	(MS)^{2}TL
0	0.8562	0.8597	**0.8613**	0.8977	**0.9054**	0.9045	0.8049	**0.8219**	0.8209
2	0.8564	0.8618	**0.8630**	0.8977	**0.9057**	0.9052	0.8051	0.8215	**0.8226**
4	0.8562	0.8626	**0.8637**	0.8983	0.9060	**0.9071**	0.8052	0.8222	**0.8233**
6	0.8563	0.8629	**0.8645**	0.8971	0.9078	**0.9089**	0.8048	0.8234	**0.8250**
10	0.8563	0.8661	**0.8670**	0.8972	0.9101	**0.9112**	0.8066	0.8341	**0.8378**
15	0.8564	0.8696	**0.8704**	0.8992	0.9111	**0.9123**	0.8055	0.8259	**0.8281**
20	0.8564	0.8738	**0.8745**	0.8987	0.9164	**0.9175**	0.8050	0.8418	**0.8450**
Ave.	0.8562	0.8761	**0.8771**	0.8978	0.9147	**0.9152**	0.8052	0.8371	**0.8383**

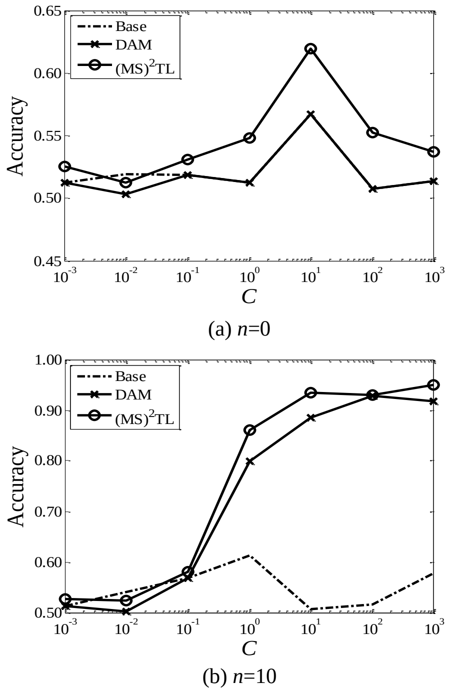
In the 20 Newsgroups dataset, each document is represented by the 61188 dimensional word-frequency features. Since the feature dimension is very high, we only perform the experiments with one kind of source classifier (i.e. LS-SVM). The classification accuracies of the methods under different number of the labeled target domain samples are recorded in Table III. The highest accuracy among different methods is highlighted in bold.

Table III shows that our method (MS)^{2}TL outperforms other algorithms in most cases except that it performs slightly worse than DAM in two cases when setting n=0, and 2 (see setting "*comp* vs. *rec*" and "*comp* vs. *sci*" in Table III). When the number of lab`eled samples per class (i.e., n) from the target domain increases, the performances of (MS)^{2}TL and DAM both improve. We observe that the Base method also achieves good results by only using the source domain classifiers directly, possibly because the source domains are highly relevant to the target domain. This conjecture is also supported by measuring the similarities between the sources and the target according to section III.A.

C. Parameter Analysis

For further studying the performance of the proposed (MS)^{2}TL, the influences of the parameters are considered. In this section, we evaluate the performance variations with respect to the regularization parameters C, λ_L, λ_L, and the number of neighbors N_k by using the datasets of target recognition described in Section IV.A. When evaluating the performance variations with respect to one parameter, we fix the other parameters as their default values (see the beginning of the Section IV).

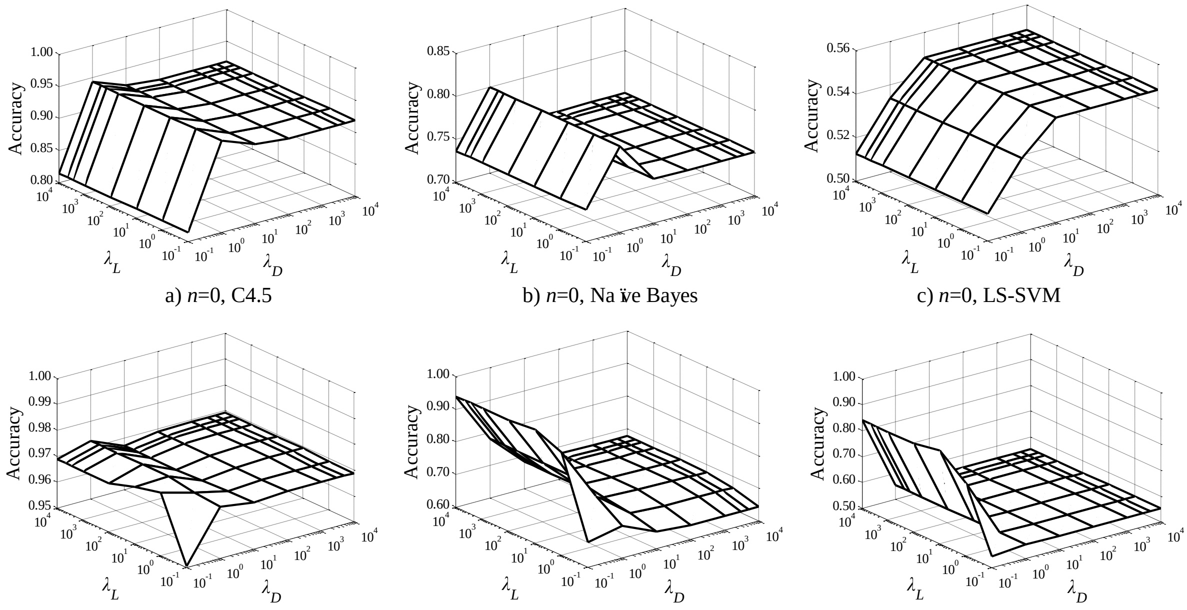
First, we consider the performance variations w.r.t. the regularization parameter C. We choose the LS-SVM as the source domain classifier since it also has the parameter C. In the experiments, C is tuned in the range [10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3]. Here, the classification results of all methods with different number of the labeled target samples (i.e., n) are shown in Fig. 4. We observe that our method (MS)^{2}TL is better than other methods by using different C's in most cases. If there is no labeled samples in the target domain (i.e., n=0), DAM has no improvements compared with Base while (MS)^{2}TL still achieves the highest classification accuracy in most cases. In the case of labeled samples existing in the target domain (i.e., n=10), the performances of DAM and (MS)^{2}TL tend to saturate when C becomes large while the classification results of Base are always not good. To sum up, fixing the value of C at [10^{-3}, 10^{-1}] is recommended.



(a) n=0

(b) n=10

Fig. 4. Classification accuracies of all methods on the target recognition dataset with different C



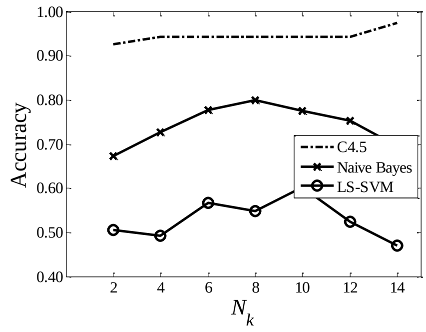
a) n=0, C4.5 b) n=0, Naïve Bayes c) n=0, LS-SVM

d) n=10, C4.5 e) n=10, Naïve Bayes f) n=10, LS-SVM

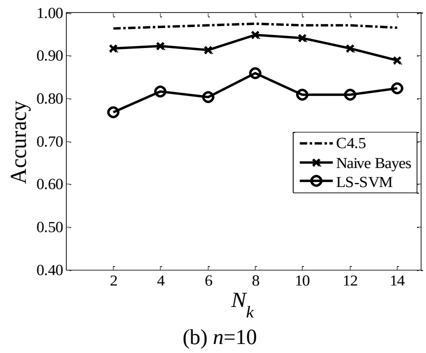
Fig. 5. Classification accuracy on the target recognition dataset with different λ_L and λ_D

The performance variations with respect to different λ_L and λ_D are shown in Fig. 5. Specifically, we set λ_L and λ_D for (MS)^{2}TL as 0.1, 1, 10, 10^2, 10^3, $3*10^3$, $5*10^3$, and 10^4 respectively. From Fig. 5, we observe that the performance of (MS)^{2}TL changes a little along with the variation of λ_L while it changes dramatically along with the variation of λ_D. It demonstrates that the regularizer $\Omega_M(f^T)$ has a big influence on the performance of (MS)^{2}TL. Compared with the two settings in Fig. 5 (i.e., n=0 or 10), we also observe that (MS)^{2}TL achieves the highest accuracy at a larger value of λ_D when there is no labeled target domain samples (i.e., n=0). It can be explained that (MS)^{2}TL depends more on the $\Omega_M(f^T)$ when no labeled target domain samples exist. We also observe that the performances become stable when setting $\lambda_D \geq 10$ in all cases.

N_k is the number of neighbors for the calculation of DN_i^s (see (4)). We show the performance of (MS)^{2}TL with different N_k in Fig. 6, where N_k is set as 2, 4, 6, 8, 10, 12, and 14. In both two settings of Fig. 5 (i.e., n=0 or 10), we can see that the performance of (MS)^{2}TL depends on the setting of parameter N_k. Especially, this dependence is evident when no labeled samples exist in the target domain. This may be because that (MS)^{2}TL will depend more on the knowledge from the sources if there are no labeled target domain samples, then N_k will have a bigger influence on (MS)^{2}TL since N_k is a key parameter for the knowledge transfer. In most cases, it is observed that the learning performance will be badly hurt if the value of N_k is too high or too low. The reason can be concluded as: if the value of N_k is set too small, the local scope can not cover all the affinitive examples; on the contrary, if the value is fixed beyond normal scope, the similarity measure may suffer interfere from the false distribution of the irrelevant data. Thus, fixing the value of N_k at [6, 10] is recommended. In addition, the influence of N_k is small when C4.5 is used as the source classifier.

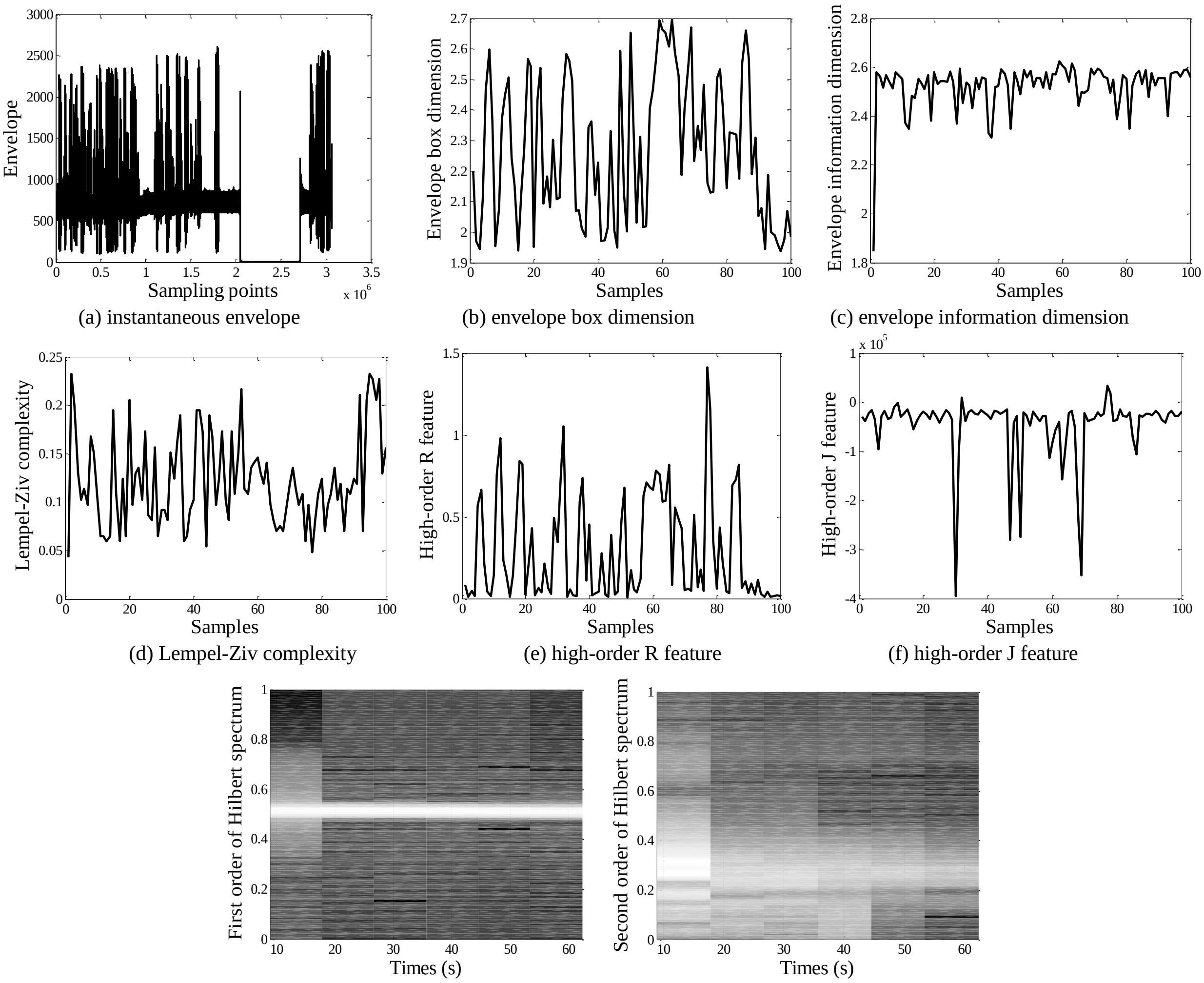


(a) n=0



(b) n=10

Fig.6. Classification accuracy on the target recognition dataset with different N_k



(a) instantaneous envelope (b) envelope box dimension (c) envelope information dimension

(d) Lempel-Ziv complexity (e) high-order R feature (f) high-order J feature

(g) first order of Hilbert spectrum (h) second order of Hilbert spectrum

Fig. 7. Radio emitter signal and the extracted features

D. *Application to Communication Specific Emitter Identification*

Communication specific emitter identification [25]-[26] is widely used in applications such as spectrum management, cognitive radio, network intrusion detection, intelligence gathering, etc. This system discerns wire-less radio emitters of interest only based on the external signal feature measurements. However, in the real-world application, the feature of the emitter signal always changes along with different operation modes, different times and other conditions. It is difficult to make sure that the training data collected previously is suitable for the current target task. Here, the proposed transfer learning algorithm $(MS)^2$TL is applied to this application.

Digital radios with the same type and same modulation mode are selected as the specific emitters, whose transmitting signal bandwidth is 25 KHz. The signal is sampled at the sampling frequency of 204.8 KHz under different conditions, e.g., different work frequencies (160MHz or 410MHz), different speakers (speaker 1, speaker 2 or speaker 3), and different receive distances (short distance with direct wave or long distance without direct wave). After the raw data of emitter signal are obtained, we extract the widely adopted emitter features such as envelope box dimension, envelope information dimension, Lempel-Ziv complexity, high-order spectrum, and Hilbert spectrum. Fig. 7 shows the instantaneous envelope and the extracted features of one radio emitter's signal. To validate the performance of transfer learning, we select data sets under various conditions as the source domains and target domain. Information on these datasets is tabulated in Table IV.

TABLE IV: EXPERIMENTAL RADIO EMITTER DATA

Domains	Work frequency	Speaker	Receive distance
Target	160MHz	Speaker 1	long distance
Source 1	410MHz	Speaker 1	long distance
Source 2	160MHz	Speaker 2	long distance
Source 3	160MHz	Speaker 3	long distance
Source 4	160MHz	Speaker 1	short distance

100 samples of each emitter are randomly chosen from 'Target' in Table IV as the target domain. The other datasets in Table IV are used as the source domains. In the target domain, we also choose n samples per emitter as the labeled target samples for experiments. n is also tuned in the range [0, 2, 4, 6, 10, 15, 20]. For each source domain, we choose a certain number of samples per

emitter for experiments. The number is set as 10, 50, and 100. All the other parameters are set as the default values described at the beginning of Section IV. The experiments are also repeated for 20 times with different source samples and target samples. The average classification accuracies are recorded in Table V to Table VII.

It can be seen that the $(MS)^2TL$ has generally achieved higher classification accuracies compared with Base and DAM. The highest classification accuracy of $(MS)^2TL$ can be as high as 93.76%. All of these experimental results show that $(MS)^2TL$ is more suitable for communication specific emitter identification task.

TABLE V: CLASSIFICATION ACCURACY WHEN THE SAMPLE NUMBER PER CLASS OF EACH SOURCE IS 10

n	C4.5			Naïve Bayes			LS-SVM		
	Base	DAM	$(MS)^2TL$	Base	DAM	$(MS)^2TL$	Base	DAM	$(MS)^2TL$
0	0.4769	0.7586	**0.8879**	0.5297	0.6807	**0.8684**	0.3396	0.5875	**0.6884**
2	0.4785	0.7719	**0.8932**	0.4950	0.7437	**0.8773**	0.3179	0.6544	**0.8180**
4	0.4724	0.8243	**0.9035**	0.5300	0.7666	**0.8925**	0.2975	0.7126	**0.8367**
6	0.4945	0.8477	**0.9061**	0.5174	0.8381	**0.9184**	0.3020	0.7677	**0.8632**
10	0.4673	0.8756	**0.9126**	0.5121	0.8569	**0.9122**	0.2981	0.8296	**0.8836**
15	0.4939	0.9248	**0.9336**	0.5125	0.8951	**0.9129**	0.3295	0.8975	**0.9163**
20	0.4986	0.9134	**0.9194**	0.5005	0.9049	**0.9206**	0.3268	0.9153	**0.9203**
Ave.	0.4832	0.8452	**0.9081**	0.5139	0.8123	**0.9003**	0.3159	0.7664	**0.8467**

TABLE VI: CLASSIFICATION ACCURACY WHEN THE SAMPLE NUMBER PER CLASS OF EACH SOURCE IS 50

n	C4.5			Naïve Bayes			LS-SVM		
	Base	DAM	$(MS)^2TL$	Base	DAM	$(MS)^2TL$	Base	DAM	$(MS)^2TL$
0	0.4830	0.7192	**0.8072**	0.5549	0.6871	**0.7545**	0.3805	0.6446	**0.6686**
2	0.4719	0.7983	**0.8765**	0.5503	0.7291	**0.8099**	0.3176	0.6572	**0.7075**
4	0.4821	0.7972	**0.8756**	0.5456	0.7609	**0.8428**	0.3221	0.7267	**0.7909**
6	0.4660	0.8583	**0.8967**	0.5459	0.8260	**0.8822**	0.3329	0.7836	**0.8479**
10	0.4899	0.8855	**0.9099**	0.5358	0.8666	**0.9021**	0.3816	0.8686	**0.8982**
15	0.4907	0.8984	**0.9064**	0.5374	0.8950	**0.9108**	0.4234	0.9169	**0.9255**
20	0.4781	0.9336	**0.9376**	0.5254	0.9139	**0.9205**	0.3748	0.9133	**0.9223**
Ave.	0.4802	0.8415	**0.8871**	0.5422	0.8112	**0.8604**	0.3618	0.7873	**0.8230**

TABLE VII: CLASSIFICATION ACCURACY WHEN THE SAMPLE NUMBER PER CLASS OF EACH SOURCE IS 100

n	C4.5			Naïve Bayes			LS-SVM		
	Base	DAM	$(MS)^2TL$	Base	DAM	$(MS)^2TL$	Base	DAM	$(MS)^2TL$
0	0.4668	0.7669	**0.7755**	0.5559	0.7206	**0.7392**	0.4041	0.6250	**0.6290**
2	0.4678	0.7812	**0.8180**	0.5483	0.7246	**0.7454**	0.3653	0.6305	**0.6451**
4	0.4682	0.8296	**0.8379**	0.5209	0.7644	**0.7767**	0.3720	0.7048	**0.7105**
6	0.4725	0.8579	**0.8707**	0.5575	**0.8150**	0.8088	0.3804	**0.7520**	0.7399
10	0.4973	0.8917	**0.9032**	0.5175	0.8381	**0.8642**	0.3825	0.8736	**0.8833**
15	0.4988	0.9101	**0.9136**	0.5116	**0.8874**	0.8826	0.3592	**0.8798**	0.8771
20	0.4915	0.9047	**0.9068**	0.5537	**0.9234**	0.9215	0.3734	0.9106	**0.9113**
Ave.	0.4804	0.8489	**0.8608**	0.5379	0.8105	**0.8198**	0.3767	0.7681	**0.7709**

V. CONCLUSIONS

In this paper, a novel multi-source transfer learning method called $(MS)^2TL$ is proposed. The method explores the relationships between the source domains and the target domain by multi-similarity metric. Then, the knowledge of the source domains is transferred to the target domain based on the smoothness assumption, which enforces that the target classifier shares similar decision values with the relevant source domain classifiers on the unlabeled target samples. The method can import more knowledge from multiple sources for the target learning and also increase the chance of finding the source domains closely related to the target domain to reduce the “negative transfer”. Furthermore, the proposed method only needs the pre-learned source domain classifiers when training the target domain classifier, which is suitable for large datasets. We also use the ε-insensitive loss function to enforce the sparsity of the decision function for fast label prediction. Extensive experiments on the target recognition and document retrieval clearly demonstrate the effectiveness of our method. For further showing the practicality, $(MS)^2TL$ is also applied to the task of communication specific emitter identification and the result is also satisfying.

ACKNOWLEDGMENT

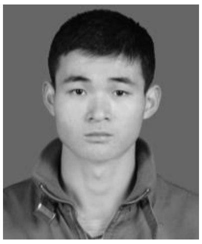
Both authors would like to acknowledge the support of Key Laboratory of Electronic Restriction of Anhui Province, National High-tech R&D Program (863 Program), and Anhui Provincial Natural Science Foundation (NO.1308085QF99, NO.1408085MKL46).

REFERENCES

[1] S. J. Pan and Q. Yang, “A survey on transfer learning,” *IEEE Trans. Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345-1359, Oct. 2010.

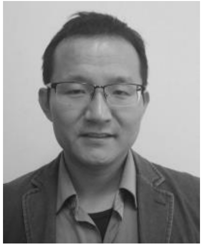
[2] S. L. Sun, H. L. Shi, and Y. B. Wu, “A survey of multi-source domain adaptation,” *Information Fusion*, vol. 24, pp. 84-92, July. 2015.

[3] M. T. Bahadori, Y. Liu, and D. Zhang, “A general framework for scalable transductive transfer learning,” *Knowl. Inf. Syst.*, vol. 38, no. 1, pp. 61-83, Jan. 2014.
[4] B. Cheng, M. Liu, H. Suk, D. Shen, and D. Q. Zhang, “Multimodal manifold-regularized transfer learning for MCI conversion prediction,” *Brain Imaging and Behavior*, vol. 9, no. 4, pp. 913-926, Dec. 2015.
[5] L. Duan, D. Xu, and S. Chang, “Exploiting web images for event recognition in consumer videos: a multiple source domain adaptation approach,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognitio*n, Providence, 2012, pp. 1338-1345.
[6] M. Fang, J. Yin, X. Q. Zhu, and C. Q. Zhang, “TrGraph: Cross-network transfer learning via common signature subgraphs,” *IEEE Trans. Knowledge and Data Engineering*, vol. 27, no. 9, pp. 2536-2549, Mar. 2015.
[7] R. A. Shafik, A. Das, L. A. Maeda-Nunez, S. Yang, G. V. Merrett, and B. M. Al-Hashimi, “Learning transfer-based adaptive energy minimization in embedded systems,” *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, Oct. 2015.
[8] K. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf, “An introduction to kernel-based learning algorithms,” *IEEE Trans. on Neural Netw.*, vol. 12, no. 2, pp. 181-201, Mar. 2001.
[9] T. Van Gestel, J. A. K. Suykens, B. Baesens, S. Viaene, J. Vanthienen, G. Dedene, B. de Moor, and J. Vandewalle, “Benchmarking least squares support vector machine classifiers,” *Mach. Learn.*, vol. 54, no. 1, pp. 5-32, Jan. 2004.
[10] R. Chattopadhyay, J. Ye, S. Panchanathan, W. Fan, and I. Davidson, “Multisource domain adaptation and its application to early detection of fatigue,” *ACM Trans. Knowledge Discov. Data*, vol. 6, no. 4, pp. 717-725, Aug. 2011.
[11] Q. Sun, R. Chattopadhyay, S. Panchanathan, and J. Ye, “A two-stage weighting framework for multi-source domain adaptation,” *Adv. Neural Inform. Process. Syst.*, vol. 24, pp. 505-513, Dec. 2011.
[12] G. Schweikert, G. Rätsch, C. Widmer, and B. Schölkopf, “An empirical analysis of domain adaptation algorithms for genomic sequence analysis,” *Adv. Neural Inform. Process. Syst.*, vol. 21, pp. 1433-1440, Dec. 2009.
[13] J. Yang, R. Yan, and A. Hauptmann, “Cross-domain video concept detection using adaptive svms,” in *Proc. 15th International Conf. Multimedia*, New York, 2007, pp. 188-197.
[14] L. Duan, D. Xu, and I. Tsang, “Domain adaptation from multiple sources: a domain dependent regularization approach,” *IEEE Trans. on Neural Networks Learn. Syst.*, vol. 23, no. 3, pp. 504-518, Mar. 2012.
[15] K. M. Borgwardt, A. Gretton, M. Rasch, H. Kriegel, B. Schölkopf, and A. J. Smola, “Integrating structured biological data by kernel maximum mean discrepancy,” *Bioinformatics*, vol. 22, no. 14, pp. 49-57, Jul. 2006.
[16] M. Belkin, P. Niyogi, and V. Sindhwani, “Manifold regularization: A geometric framework for learning from labeled and unlabeled examples,” *J. Mach. Learn. Res.*, vol. 7, pp. 2399-2434, Dec. 2006.
[17] I. W. Tsang and J. T. Kwok, “Large-scale sparsified manifold regularization,” in *Advances in Neural Information Processing Systems 19*, Cambridge, 2007, pp. 1401-1408.
[18] T. Kato, H. Kashima, M. Sugiyama, and K. Asai, “Multi-task learning via conic programming,” in *Advances in Neural Information Processing Systems 20*, Cambridge, 2008, pp. 737-744.
[19] C. C. Chang and C. J. Lin. (2001). LIBSVM: A Library for Support Vector Machines [Online]. Available: http://www.csie.ntu.edu.tw/~ cjlin/libsvm
[20] E. Frank, M. Hall, P. Reutemann, and L. Trigg. Waikato environment for knowledge analysis. [Online]. Available: http://www.cs.waikato.ac.nz/~ml/weka/
[21] K. Saenko, B. Kulis, M. Fritz, and T. Darrell, “Adapting visual category models to new domains,” in *Proc. 11th European Conf. Computer Vision*, Heraklion, 2010, pp. 213–226.
[22] G. Griffin, A. Holub, and P. Perona, “Caltech-256 object category dataset,” Technical Report of California Institute of Technology, 2007.
[23] 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,” in *Proc. International Conf. Machine Learning*, Beijing, 2014, pp. 647-655.
[24] K. Lang (Jan. 2014). 20Newsgroups. [Online]. Available: http://qwone.com/~jason/20Newsgroups/
[25] Y. M. Chen, C. M. Lin, and C. S. Hsueh, “Emitter identification of electronic intelligence system using type-2 fuzzy Classifier,” *Systems Science & Control Engineering*, vol. 2, no. 1, pp. 389-397, May. 2014.
[26] Y. J. Yuan, Z. T. Huang, H. Wu, and X. Wang, “Specific emitter identification based on Hilbert Huang transform based time frequency energy distribution features,” *IET Communications*, vol. 8, no. 13, pp. 2404-2412, Sept. 2014.

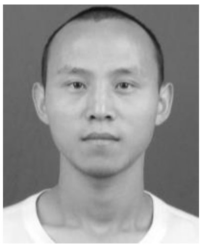


Zhen Liu was born in Anhui Province, China, in 1989. He received the B.S. degree in electrical engineering & automation from Anhui University, Hefei, China, in 2010 and the M.S. degree in circuits & systems from Electronic Engineering Institute, Hefei, China, in 2013. He is currently pursuing the Ph.D. degree with the Department of Communications, Electronic Engineering Institute, Hefei, China. His research interests include communication specific emitter identification, intelligent computing, data mining and machine learning.

Jun-An Yang was born in Anhui Province, China, in 1965. He received the B.S. degree in radio technology from Southeast University, Nanjing, China in 1986 and the M.S. degree in communication & information systems from Electronic Engineering Institute, Hefei, China, in 1991. He received the Ph.D. degree in signal & information processing from University of Science and Technology of China (USTC), Hefei, China, in 2003. He is currently a professor in the Department of Communications at Electronic Engineering Institute, Hefei, China. His research interests include neural computing, large-scale machine and computer vision.



Hui Liu was born in Anhui Province, China, in 1983. He received the B.S. degree in communication engineering from Wuhan University, Wuhan, China, in 2005. He received the M.S. degree and the Ph.D. degree in communication & information systems from Electronic Engineering Institute, Hefei, China, in 2008 and 2011 respectively. He is currently a lecturer in the Department of Communications at Electronic Engineering Institute, Hefei, China. His research interests include intelligent information processing and cognitive communication.



Wei Wang was born in Anhui Province, China, in 1987. He received the B.S. degree in mechanism design, manufacturing & automatization from Northwestern Polytechnical University, Xi'an, China, in 2008. He received the M.S. degree and the Ph.D. degree in precision instrument and machinery from University of Science and Technology of China (USTC), Hefei, China, in 2011 and 2014 respectively. He is currently a lecturer in the Department of Communications at Electronic Engineering Institute, Hefei, China. His research interests include intelligent information processing and computer vision.