LEARNING DISCRIMINATIVE GEODESIC FLOW KERNEL FOR UNSUPERVISED DOMAIN ADAPTATION

Jianze Wei[†], *Jian Liang*[§,¶,*], *Ran He*[§,‡,¶], *Jinfeng Yang*[†]

[†] Tianjin Key Lab for Advanced Signal Processing, CAUC
[§] CRIPAC & NLPR, CASIA [¶] University of Chinese Academy of Sciences (UCAS)
[‡] CAS Center for Excellence in Brain Science and Intelligence Technology (CEBSIT)
caucweijianze@163.com, {jian.liang, rhe}@nlpr.ia.ac.cn, jfyang@cauc.edu.cn

ABSTRACT

Extracting the domain-invariant features provides an important intuition for unsupervised domain adaptation. Due to the unavailable target labels, it is difficult to guarantee that the learned domain-invariant features are good for target instances classification. In this paper, we extend the classic geodesic flow kernel method by leveraging the pseudo labels during the training process to learn a discriminative geodesic flow kernel for unsupervised domain adaptation. Specifically, the proposed method alternately discovers the pseudo target labels and builds the geodesic flow from a discriminative source subspace to another 'discriminative' target subspace. More specially, the pseudo target labels are inferred via the learned kernel based on an easy yet effective label propagation strategy. Hence, the proposed method not only holds the property of domain-invariance, but also maximizes the consistency between pseudo label structure and data structure. Experimental results illustrate that the proposed method outperforms the state-of-the-art unsupervised domain adaptation methods for object recognition and sentiment analysis.

Index Terms— unsupervised domain adaptation, geodesic flow kernel, label propagation.

1. INTRODUCTION

Supervised learning has shown phenomenal performance in recent years, however, it relies heavily on labeled data. In order to alleviate the affect of manual labeling, it is vital to utilize extensive given labeled data from related domains to perform classification. With a little or no help of target labels, domain adaptation leverages the prior knowledge from source domain on the similar task of target domain and has been extensively studied in many areas due to its characteristics. Presently, there are two settings for domain adaptation technology, including semi-supervised domain adaptation where a few labeled target data are available and unsupervised domain adaptation where the target data are completely unlabeled. Without the supporting of target labels, there exists a performance gap between semi- and un-supervised domain adaptation settings. Even being challenging, unsupervised domain adaptation is appealing because you do not need any prior knowledge from target domain. No matter unsupervised or semi-supervised domain adaptation, the greatest challenge they face is the large distribution difference between different domains, which is named as distribution shift or domain shift. This shift always leads to a bad situation where classifiers trained on source domain often exhibit significant degradation in recognition accuracy of another target task.

To tackle the distribution shift, one intuition is to discover the domain-invariant feature representation for both source and target domains. In this manner, Daumè III [1, 2] and Li [3] proposed Easy Adaptation (EA or EA++) and Heterogeneous Feature Augmentation (HFA), respectively, and these methods built an augmented space to include common, source-specific and target-specific components. However, both EA++ and HFA were semi-supervised methods, the dependence on target labels limited their application. Gopalan [4] investigated manifold learning and proposed Sampling Geodesic Flow (SGF). SGF contributed the geodesic flow curve to connect the source and target domains, then sampled a fixed number of subspaces derived from the flow and concatenated the features from these subspaces to generate an augmented features. However, there was no easy strategy to tune the parameters for SGF, which limited its performance. Gong [5] extended SGF and proposed an unsupervised domain adaptation method based on Grassmann manifolds to build the augmented space. This feature space included all intermediate subspaces derived from a geodesic flow between source space and target space, and it was insensitive to idiosyncrasies in either domain. However, due to lack of the prior knowledge of target labels, it was hard to ensure that feature spaces are class-discriminative for the target task.

To address this problem, we put forward Discriminative Geodesic Flow Kernel (D-GFK), a simple but effective unsupervised domain adaptation method that considers the domain-invariant feature representation and the geometric relationships in the reproducing kernel Hilbert spaces (RKHS). The proposed method aims to guide the geodesic flow ker-

978-1-5386-1737-3/18$31.00

nel learning toward a class-discriminative direction according to the geometric relationships, which is precious prior knowledge for unsupervised domain adaptation. The proposed D-GFK with two steps can be optimized by an alternative strategy. The first step is to learn the Geodesic Flow Kernel (GFK). According to the source basis and the current target basis, we compute the kernel for geodesic flow between source and target domains, which is almost similar to GFK [5]. The second step learns the target basis points on Grassmann manifold. It first estimates the pseudo labels for target data with the graph-based transition in RKHS, then utilizes them to generate a set of new target basis vectors which is better at classification than before. These two steps are executed alternately until convergence. The resulted kernel of D-GFK not only has a good property of domain-invariance, but also contains the geometric relationships between data from both domains. Moreover, the empirical experiments show that the proposed D-GFK can quickly converge within several iterations. The main contributions of this paper are summarized as follows: 1). We propose a better domain-invariant feature learning method to expect that the learned feature spaces are latently discriminative with the aid of pseudo labels. 2). Based on the kernel similarities on the learned invariant feature space, the label propagation method is natively utilized and shows the potential on domain-invariant feature learning. 3). The experimental results demonstrate that the proposed method outperforms state-of-the-art unsupervised domain adaptation methods on the visual and sentiment adaptation benchmarks.

The rest of this paper is organized as follows. In Section 2, we provide a brief review of related works, especially methods for specific domain or tasks. Section 3 presents our proposed method and algorithm. In Section 4, we give the introduction of datasets and the details of experimental evaluation. Finally, the conclusion is given in Section 5.

2. RELATED WORK

Recently, domain adaptation has drawn a widespread attention of researchers and become a hot research field, which is considered as an important direction in machine learning.

The most commonly used domain adaptation approaches can be classified into two categories, instance-based adaptation methods and feature-based adaptation methods. Instance-based adaptation methods weight the labeled source data to fit the distribution of target domain, then apply the classifier trained from the weighed source data to the target domain. TrAdaBoost [6] and Transfer Joint Matching (TJM) [7] can be attributed to this category. TrAdaBoost weights the source instances to filter out the data that is irrelevant to target domain, and TJM implements instance-weighting using a $l_{2,1}$-norm constraint to select sparse instances that can be reused on target domain.

For feature-based methods, some of them, like Transfer Component Analysis (TCA) [8], Joint Distribution Analysis (JDA) [9], Subspace Alignment (SA) [10], CORrelation ALignment (CORAL) [11], Adaptation Regularization Transfer Learning (ARTL) [12], Balanced Distribution Adaptation (BDA) [13], Joint Geometrical and Statistical Alignment (JGSA) [14] etc., aim to map data into a common space to match the distribution of the source domain and that of the target domain. The others are to build an augmented space that contains original source and target spaces, and the representative work of this part includes EA++ , HFA, SGF and GFK.

Of particular relevance to our work is GFK, a domain-invariant feature representation which belongs to the second category. GFK extends SGF and overcomes the limitation of parameter tuning using a kernel-based framework. It uses all collections of d-dimensional subspaces from a smooth Riemannian manifold to represent data from source and target domain. The merit of this idea is that the feature representation of GFK covers the source domain, target domain and intermediate domains between source and target domain, guaranteeing the domain-invariance. However, the lack of target labels makes it difficult to ensure that the GFK feature is class-discriminative for target task.

3. PROPOSED APPROACH

In this section, we begin with the definition of terminologies, then detail the proposed method.

The source data $\mathbf{X}_s = [\mathbf{x}_1^s, \mathbf{x}_2^s, \ldots, \mathbf{x}_{n_s}^s] \in \mathbb{R}^{D \times n_s}$ are drawn from the distribution $\mathcal{P}_s(x_s)$, and $\mathbf{X}_t = [\mathbf{x}_1^t, \mathbf{x}_2^t, \ldots, \mathbf{x}_{n_t}^t] \in \mathbb{R}^{D \times n_t}$ denotes the target data drawn from the distribution $\mathcal{P}_t(x_t)$, where D represents the feature dimensionality of the data, n_s and n_t are the number of samples from source and target domain respectively. For unsupervised domain adaptation, there are labeled data available from source domain $\mathcal{D}_s = \{(\mathbf{x}_i, y_i)\}_{i=1}^{n_s}$ and unlabeled target instances $\mathcal{D}_t = \{(\mathbf{x}_j)\}_{j=1}^{n_t}$ in the training stage. Assuming the feature space and label space are the same, $\mathcal{X}_s = \mathcal{X}_t$ and $\mathcal{Y}_s = \mathcal{Y}_t$, while $\mathcal{P}_s(X_s) \neq \mathcal{P}_t(X_t)$ due to the domain shift.

In this paper, all the matrices are written as uppercase bold, and lowercase letters bold for vectors. $\mathbf{M}_{ij}$ denotes the (i,j)-th element of matrix $\mathbf{M}$. $\mathbf{M}^{-1}$ and $\mathbf{M}^{\mathrm{T}}$ represent the inverse and transpose of $\mathbf{M}$ respectively.

3.1. Revisiting GFK

In statistical learning, there is a basic assumption that data can be embedded in a low-dimensional linear subspace. One of the most typical examples is PCA [15], where the original data are mapped into a low-dimensional subspace. This mapping preserves the most important information of the original data and avoids influence from noise or other irrelevant features. The proposition of GFK is also based on this assumption. Here, we make a brief introduction, and details can be

found in [5].

Assume that the source and target dataset correspond to two points $\mathbf{\Phi}(0)$ and $\mathbf{\Phi}(1)$ on a Grassmann manifold. $\mathbf{P}_s \in \mathbb{R}^{D\times d}$ and $\mathbf{P}_t \in \mathbb{R}^{D\times d}$ denote two sets of basis of the subspaces for the source and target domain. $\mathbf{R}_s \in \mathbb{R}^{D\times(D-d)}$ presents the orthogonal complement to $\mathbf{P}_s$, where d is the dimensionality of the subspace. Then we construct a geodesic flow [4, 5] between the two points and integrate an infinite number of subspaces along the flow $\mathbf{\Phi}(t)$. This geodesic flow is parameterized as

$$\begin{aligned}\mathbf{\Phi}(t) &= \mathbf{P}_s\mathbf{U}_1\mathbf{\Gamma}(t) - \mathbf{R}_s\mathbf{U}_2\mathbf{\Sigma}(t) \\ &= \begin{bmatrix} \mathbf{P}_s & \mathbf{R}_s \end{bmatrix} \begin{bmatrix} \mathbf{U}_1 & 0 \\ 0 & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \mathbf{\Gamma}(t) \\ \mathbf{\Sigma}(t) \end{bmatrix},\end{aligned} \tag{1}$$

where $\mathbf{U}_1 \in \mathbb{R}^{D\times d}$ and $\mathbf{U}_2 \in \mathbb{R}^{D\times d}$ are two orthonormal matrices, and can be computed by

$$\mathbf{P}_S^{\mathrm{T}}\mathbf{P}_T = \mathbf{U}_1\mathbf{\Gamma}\mathbf{V}^{\mathrm{T}}, \mathbf{R}_S^{\mathrm{T}}\mathbf{P}_T = -\mathbf{U}_2\mathbf{\Sigma}\mathbf{V}^{\mathbf{T}}. \tag{2}$$

To ensure the domain-invariance, all subspace features are concatenated together to form the GFK feature $\mathbf{z}^\infty$, and a kernel-based framework is implemented for optimization. The kernel matrix can be computed in a closed-form

$$\mathbf{G} = \begin{bmatrix} \mathbf{P}_s\mathbf{U}_1 & \mathbf{R}_s\mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda}_1 & \mathbf{\Lambda}_2 \\ \mathbf{\Lambda}_2 & \mathbf{\Lambda}_3 \end{bmatrix} \begin{bmatrix} \mathbf{U}_1^{\mathrm{T}}\mathbf{P}_s^{\mathrm{T}} \\ \mathbf{U}_2^{\mathrm{T}}\mathbf{R}_s^{\mathrm{T}} \end{bmatrix}, \tag{3}$$

where $\mathbf{\Lambda}_1$, $\mathbf{\Lambda}_2$ and $\mathbf{\Lambda}_3$ are three diagonal matrices with elements

$$\lambda_{1i} = 1+\frac{\sin(2\theta_i)}{2\theta_i}, \lambda_{2i} = \frac{\cos(2\theta_i)-1}{2\theta_i}, \lambda_{3i} = 1-\frac{\sin(2\theta_i)}{2\theta_i}. \tag{4}$$

With the help of GFK, the data from source and target domain can be embedded into a shared domain-invariant space.

3.2. Label propagation with the kernel transition matrix

Since the target labels are unavailable, it is difficult to ensure that the target basis obtained in GFK is always class-discriminative. To address this problem, [16, 17, 18] have obtained promising performance in semi-supervised learning by exploiting the geometric relationship of the data structure. Following their works, we propose a label propagation with the kernel transition matrix.

Regarding instances as different nodes in domain-invariant space, a node is more likely to have the same label as the nearest node. For $\mathbf{x}_t$, there is a larger probability of having the same label with the nearest $\mathbf{x}_s$ in the support set $\{(\mathbf{x}_i^s)\}_{i=1}^{n_s}$. In order to describe this probability, soft-label $\mathbf{L} = \left[\mathbf{l}_1^{\mathrm{T}}, \mathbf{l}_2^{\mathrm{T}}, \ldots, \mathbf{l}_{n_s+n_t}^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{R}^{(n_s+n_t)\times c}$ is employed to present the label $\left[\mathbf{y}_s^{\mathrm{T}}, \mathbf{y}_t^{\mathrm{T}}\right]^{\mathrm{T}}$, and c is the number of categories. The index corresponding to the largest value in $\mathbf{l}_i$ is y_i^t. In our model, $\mathbf{l}_j$ can be expressed as

$$\mathbf{l}_j = \sum_{i=1}^{n_s+n_t} h(\mathbf{x}_j, \mathbf{x}_i)\mathbf{l}_i, \tag{5}$$

where $h(\mathbf{x}_j, \mathbf{x}_i)$ denotes the transfer probability from $\mathbf{x}_i^s$ to $\mathbf{x}_j^t$. Here, we define $h(\mathbf{x}_j, \mathbf{x}_i)$ as a normalized gaussion distance between $\mathbf{z}_i^\infty$ and $\mathbf{z}_j^\infty$,

$$\begin{aligned}h(\mathbf{x}_j, \mathbf{x}_i) &= \frac{\exp\{-\frac{(\mathbf{z}_i^\infty - \mathbf{z}_j^\infty)^2}{\sigma^2}\}}{\sum_{i=1}^{n_s+n_t}\exp\{-\frac{(\mathbf{z}_i^\infty - \mathbf{z}_j^\infty)^2}{\sigma^2}\}} \\ &= \frac{\exp\{-\frac{(\mathbf{x}_i-\mathbf{x}_j)^T\mathbf{G}(\mathbf{x}_i-\mathbf{x}_j)}{\sigma^2}\}}{\sum_{i=1}^{n_s+n_t}\exp\{-\frac{(\mathbf{x}_i-\mathbf{x}_j)^T\mathbf{G}(\mathbf{x}_i-\mathbf{x}_j)}{\sigma^2}\}},\end{aligned} \tag{6}$$

where σ is the scale parameter that controls the transfer probability $h(\mathbf{x}_j, \mathbf{x}_i)$.

With above definition, the soft-label $\mathbf{L}$ can be computed in a matrix form

$$\begin{aligned}\mathbf{L} &= \begin{bmatrix} \sum_{i=1}^{n_s+n_t} h(\mathbf{x}_1, \mathbf{x}_i)\mathbf{l}_i \\ \vdots \\ \sum_{i=1}^{n_s+n_t} h(\mathbf{x}_{n_s+n_t}, \mathbf{x}_i)\mathbf{l}_i \end{bmatrix} \\ &= \begin{bmatrix} h(\mathbf{x}_1, \mathbf{x}_1) & \cdots & h(\mathbf{x}_1, \mathbf{x}_{n_s+n_t}) \\ \vdots & & \vdots \\ h(\mathbf{x}_{n_s+n_t}, \mathbf{x}_1) & \cdots & h(\mathbf{x}_{n_s+n_t}, \mathbf{x}_{n_s+n_t}) \end{bmatrix} \begin{bmatrix} \mathbf{l}_1 \\ \vdots \\ \mathbf{l}_{n_s+n_t} \end{bmatrix} \\ &= \mathbf{H}^{\mathrm{T}}\mathbf{L},\end{aligned} \tag{7}$$

where $\mathbf{H}$ represents the probabilistic transition matrix and each element denotes $\mathbf{H}_{ij} = h(\mathbf{x}_j, \mathbf{x}_i)$. According to [16], $\mathbf{L}$ can converge to an unique fixed point within multiple iterations. Obtaining the pseudo labels for the target data, it is easy to obtain more class-discriminative target basis using Partial Least Squares (PLS) [19]. The complete learning algorithm is summarized in Algorithm 1.

Algorithm 1 Discriminative Geodesic Flow Kernel

Input: Source data $\{\mathbf{X}_s, \mathbf{y}_s\}$ and target data $\mathbf{X}_t$, $\mathbf{X} = [\mathbf{X}_s, \mathbf{X}_t]$; subspace dimensionality d, scale parameter σ, # maximum iteration T.

Output: The kernel of D-GFK $\mathbf{G}$; classifier f.

```
Initializ the source basis P_s and the target basis P_t using PLS and PCA respectively.
while not converge and iter ≤ T do
    Compute G according to Eq. (2) and Eq (3).
    Construct the probabilistic transition matrix H according to Eq. (6).
    while not converge and iter ≤ T do
        Compute the soft label L according to Eq. (7).
    end while
    Update the pseudo target labels y_t. Then update the target basis P_s using PLS
end while
Re-train a 1-NN classifier f on {(x_i, y_i)}_{i=1}^{n_s}.
```

4. EXPERIMENT

In this section, we compare our approach with several state-of-the-art unsupervised domain adaptation approaches on two real-world datasets.

4.1. Benchmarks

*4.1.1. **Office-Caltech Dataset***

The Office-Caltech dataset [5] is publically available in order to support the development of visual adaptation. We use this dataset for evaluating the performance of our method on object recognition. This dataset consists of four domains including Caltech (C), Amazon (A), Dslr (D), and Webcam (W). Each domain contains 10 categories, which are overlapped between the Office dataset and Caltech-256 dataset. The dataset employs SURF BoW histogram features to represent original images.

*4.1.2. **Multi-Domain Sentiment Dataset***

This dataset is collected by John Blitzer [20] and is a benchmark dataset for sentiment analysis or sentiment adaptation. The dataset includes the reviews about Kitchen appliances (K), DVDs (D), Books (B) and Electronics (E), the reviews of each product can be regarded as data from the same domain. There are 1000 positive and 1000 negative instances on each domain. For facilitate comparison with recent studies, we follow the feature generation method from [21, 22], which exploits Marginalized Denoising Autoencoders (mSDA) [21] to improve the feature representations[1].

4.2. Experiment setting

In order to evaluate the proposed D-GFK method, we compare it with eight methods of the literature, including 1) TCA [8] + NN; 2) JDA [9] + NN; 3) TJM [7] + NN; 4) SA [10] + NN; 5) CORAL [11] + SVM; 6) GFK [5] + NN; 7) BDA [13] + NN; 8) JGSA [14] + NN. In Table 1, all the reported performance scores of the eight methods are directly collected from the authors' publication. In Table 2 and 3, the scores are gained by running the algorithms provided by the authors. We assume that these results are their best performance.

For different problems of domain adaptation, it is unreasonable to configure a unified set of hyper-parameters for them. In our proposed method, there are two hyper-parameters, i.e. the subspace dimensionality d and the scale parameter σ. In our experiments, we set 1) d = 20 and σ = 0.2 for the Office-Caltech dataset, 2) d = 10 and σ = 0.5 for the Multi-Domain Sentiment dataset. Moreover, the 'fulling training' protocol is adopted for each dataset, which means that all instances from source domain are used in training stage. In the experiment on the Office-Caltech dataset, we also employ a widely adopted training protocol [5, 11], called 'splitting sampling'. For each domain adaptation task under this protocol, p instances per class from source domain are randomly sampled for training, p = 20 for W , C, A and p = 8 for D. Following the 20 publicly standard splits, the reported accuracy is the average accuracy of 20 splits.

[1] https://github.com/GRAAL-Research/domain_adversarial_neural_network

4.3. Object recognition

Using 'fulling training' protocol, the classification accuracies of the proposed D-GFK method and the eight methods for comparison are listed in Table 1. In addition, the recognition results using 'splitting protocol' are given in Table 2. We highlight the highest accuracy for each task in bold.

In Table 1, we note that the proposed method outperforms the recently proposed unsupervised domain adaptation methods and reaches the overall accuracy of 50.24%. Compared with the result of GFK, the proposed method shows different levels of improvement in all tasks. We can also make a similar conclusion from Table 2, which reports the recognition result adopting 'splitting protocol'.

This improvement is achieved due to the fact that the geometric information plays a positive role in object recognition. Under the guide of the geometric information, the target basis is reconstructed, then the kernel matrix can be updated continuously along the correct direction.

4.4. Sentiment adaptation

For sentiment adaptation problem, we compare the proposed method with the eight methods mentioned above. The detailed accuracies of all methods are reported in Table 3.

In Table 3, most of the methods have a significant improvement compared to no adaptation methods (i.e. LP and 1-NN), while our proposed method has the highest average accuracy, 74.05%. Besides, the improvement of D-GFK over GFK is up to 7.25%. This improvement benefits from the geometric relationship, maximizing consistency between estimated label structure and data structure in domain-invariant space. It should be noted that the using of geometric information also smooths the iteration and promotes convergence, as shown in Figure 1(c) and Figure 1(d).

4.5. Parameter and convergence analysis

In order to find the optimal performance, we conduct parameter sensitivity analysis for our method. There are two parameters set in D-GFK, i.e. the subspace dimensionality d and the scale parameter σ for the transition matrix.

We implement our proposed D-GFK with different values of d. Figure 1(a) plots accuracy curves w.r.t different values of subspace dimensionality d, and d is chosen from the range of 0.1 to 1. In Figure 1(a), the accuracy curve for the Office-Caltech dataset is drawn in red, while the curve for the Multi-Domain Sentiment dataset is painted in blue. Both curves

Table 1: Recognition accuracies (%) on Office-Caltech dataset using the 'fulling training' protocol. Notation: Caltech:C; Amazon:A; Webcam:W; DSLR:D. * means that results are obtained from [14]

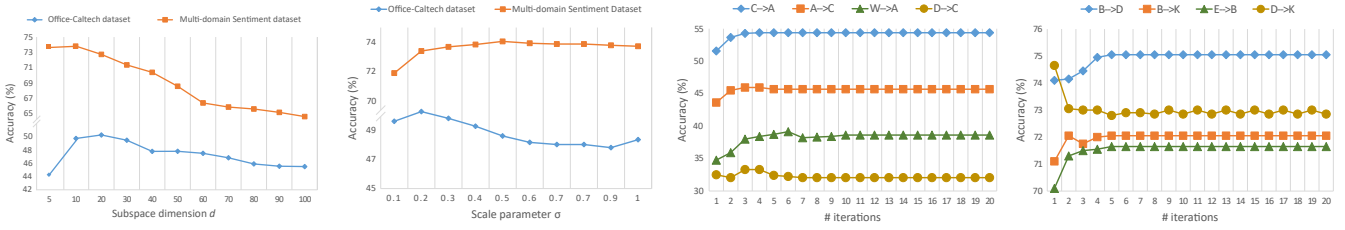
Datasets	C→A	C→W	C→D	A→C	A→W	A→D	W→C	W→A	W→D	D→C	D→A	D→W	Avg.
LP	19.42	20.34	19.11	19.79	28.14	24.84	12.11	20.46	55.41	16.21	27.14	63.05	27.17
1-NN	23.70	25.76	25.48	25.00	29.83	25.48	19.86	22.96	59.24	26.27	28.50	63.39	31.37
TCA*	45.82	31.19	34.39	42.39	36.27	33.76	29.39	28.91	89.17	30.72	31.00	86.10	43.26
JDA	44.78	41.69	45.22	39.36	37.97	39.49	31.17	32.78	89.17	31.52	33.09	89.49	46.31
TJM	46.76	38.98	44.59	39.45	42.03	45.22	30.19	29.96	89.17	31.43	32.78	85.42	46.33
SA	39.00	36.80	39.60	35.30	38.60	37.60	32.30	37.40	80.30	32.40	38.00	83.60	44.24
CORAL	47.20	39.20	40.70	40.30	38.70	38.30	34.60	37.80	84.90	**34.20**	**38.10**	85.90	46.66
GFK*	46.03	36.95	40.76	40.69	36.95	40.13	24.76	27.56	85.35	29.30	28.71	80.34	43.13
BDA	44.89	38.64	47.77	40.78	39.32	43.31	28.94	32.99	**91.72**	32.50	33.09	**91.86**	47.15
JGSA	51.46	45.42	45.86	41.50	**45.76**	**47.13**	33.21	**39.87**	90.45	29.92	38.00	**91.86**	50.04
D-GFK	**54.38**	**46.44**	**49.68**	**45.68**	41.69	46.50	**35.08**	38.62	90.45	32.06	**38.10**	84.41	**50.26**

Table 2: Recognition accuracies (%) on Office-Caltech dataset using the 'splitting training' protocol. Notation: Caltech:C; Amazon:A; Webcam:W; DSLR:D

Datasets	C→A	C→W	C→D	A→C	A→W	A→D	W→C	W→A	W→D	D→C	D→A	D→W	Avg.
LP	15.89	19.92	21.11	16.69	26.39	25.48	12.35	20.69	50.22	14.64	25.91	53.44	25.23
1-NN	21.00	19.00	23.60	20.00	23.10	22.30	12.00	14.70	31.30	19.90	23.00	51.70	23.50
TCA	37.94	35.24	35.99	34.77	36.08	32.36	29.96	30.75	81.72	28.45	31.61	79.47	41.20
JDA	38.47	35.80	36.21	34.17	36.59	33.03	30.73	31.29	81.85	27.65	32.36	79.90	41.50
TJM	38.49	38.17	37.87	33.91	37.44	35.00	31.34	29.40	82.68	26.96	31.69	80.73	41.97
SA	37.77	34.81	39.71	35.94	35.58	37.23	31.81	35.08	78.28	31.49	34.48	**82.29**	42.87
CORAL	**46.80**	38.44	40.16	**40.11**	37.58	37.64	**34.20**	**37.61**	83.54	**34.84**	38.07	80.66	45.80
GFK	37.46	34.56	38.25	35.73	34.34	35.13	30.19	32.65	82.96	31.03	32.51	77.42	41.85
BDA	40.10	32.61	39.59	34.37	33.61	32.07	31.91	32.89	82.55	30.51	31.54	79.27	41.75
JGSA	35.14	35.78	28.73	31.03	39.90	35.32	25.73	38.00	66.91	26.06	**39.17**	71.63	39.45
D-GFK	44.27	**40.29**	**42.87**	39.31	**40.02**	**37.87**	34.11	37.49	**85.16**	33.75	37.23	79.20	**45.96**

Table 3: Recognition accuracies (%) on multi-domain sentiment dataset. Notation: Books:B; Dvd:D; Electrics:E; Kitchen:K.

Datasets	B→D	B→E	B→K	D→B	D→E	D→K	E→B	E→D	E→K	K→B	K→D	K→E	Avg.
LP	50.05	50.10	50.05	50.00	50.10	49.90	49.95	50.15	50.75	50.15	50.15	50.85	50.18
1-NN	49.60	49.75	50.30	53.25	51.00	53.10	50.75	50.85	51.20	52.20	51.20	52.25	51.29
TCA	63.60	60.90	64.20	63.30	64.15	69.05	59.50	62.05	74.75	64.05	65.40	74.50	65.45
JDA	64.15	62.10	65.35	62.35	66.25	68.90	59.20	61.55	74.65	62.70	64.30	74.00	65.46
TJM	59.15	61.00	66.95	57.10	64.25	68.95	56.35	60.60	72.85	54.90	57.95	73.00	62.75
SA	67.00	70.75	72.15	67.50	67.05	69.40	61.35	64.85	70.40	64.35	64.60	68.20	67.30
CORAL	71.60	65.05	67.25	70.05	65.55	67.05	67.05	66.15	77.55	68.15	68.85	75.40	69.14
GFK	66.40	65.45	69.20	66.25	63.70	67.70	62.35	63.35	73.75	65.50	65.00	72.95	66.80
BDA	61.85	59.35	62.20	63.50	61.05	61.95	60.40	59.75	69.85	61.00	60.30	68.50	62.48
JGSA	66.55	**74.95**	72.10	55.50	67.30	65.60	51.55	50.75	54.95	58.25	56.40	51.65	60.46
D-GFK	**75.20**	71.95	**72.65**	**74.40**	**71.00**	**72.95**	**71.85**	**72.90**	**80.40**	**73.45**	**72.90**	**79.90**	**74.05**



(a) Subspace dimensionality d. (b) Scale parameter σ. (c) On the Office-Caltech Dataset. (d) On the Multi-domain sentiment Dataset.

Fig. 1: Parameter sensitivity study of D-GFK on several subsets.

increase first and then decrease with d increasing. Obviously, curves decrease due to the fact that a lower dimensional feature has fewer irrelevant items. However, low-dimensional features may increase the risk of losing important classification information, resulting in the decline of performance.

The scale parameter σ affects the performance of the label propagation. In order to explore an optimal σ, the accuracy curves for both datasets are plotted in Figure 1(b), where the value of σ varies from 0 to 1. In Figure 1(b), the effect of σ on the Multi-Domain Sentiment dataset is much smaller that that on the Office-Caltech dataset. This may be due to the fact that the Multi-Domain Sentiment dataset is a binary classification problem and has fewer categories than that of the Office-Caltech dataset. According to our experiments, we recommend $\sigma = 0.2$ for the Office-Caltech dataset and $\sigma = 0.5$ for the Multi-Domain Sentiment dataset.

Empirically, we explore the convergence property of D-GFK, Figure 1(c) and Figure 1(d) illustrate the iteration processes. Specifically, the degradation cases are marked as the lines with dots. Overall, we find that the recognition accuracies change rapidly and converges within 10 iterations.

5. CONCLUSION

We proposed the D-GFK, a simple but effective unsupervised domain adaptation method. Compared with GFK, the proposed D-GFK not only retained the advantages of domain-invariance, but also adjusted the geodesic flow between source and target space point. Furthermore, an iterative strategy made the proposed method converge quickly. In the further, we will try our best to reduce the influence of domain-invariant features on the basis of domain-invariance.

6. ACKNOWLEDGMENT

This work is supported by National Natural Science Foundation of China (No.61379102, U1433120, 61502498) and the Fundamental Research Funds for the Central Universities (No.3122014C003). This work was performed when Jianze Wei visited the National Laboratory of Pattern Recognition. Wei and Jian Liang are joint first authors for this article. Liang designed the research and is the corresponding author.

7. REFERENCES

[1] Hal Daum III, "Frustratingly easy domain adaptation," in *ACL*, 2009, pp. 256–263.

[2] A. Kumar, A. Saha, and Hal Daum III, "Co-regularization based semi-supervised domain adaptation," in *NIPS*, 2011, pp. 478–486.

[3] W. Li, L. Duan, D. Xu, and I. Tsang, "Learning with augmented features for supervised and semi-supervised heterogeneous domain adaptation," *TPAMI*, vol. 36, no. 6, pp. 1134–1148, 2014.

[4] R. Gopalan, R. Li, and R. Chellappa, "Domain adaptation for object recognition: An unsupervised approach," in *ICCV*, 2011, pp. 999–1006.

[5] B. Gong, Y. Shi, F. Sha, and K. Grauman, "Geodesic flow kernel for unsupervised domain adaptation," in *CVPR*, 2012, pp. 2066–2073.

[6] W. Dai, Q. Yang, G. Xue, and Y. Yu, "Boosting for transfer learning," in *ICML*, 2007, pp. 193–200.

[7] M. Long, J. Wang, G. Ding, Jiaguang Sun, and Philip S Yu, "Transfer joint matching for unsupervised domain adaptation," in *CVPR*, 2014, pp. 1410–1417.

[8] S. Pan, I. Tsang, J. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *TNN*, vol. 22, no. 2, pp. 199–210, 2011.

[9] M. Long, J. Wang, G. Ding, Jiaguang Sun, and Philip S Yu, "Transfer feature learning with joint distribution adaptation," in *ICCV*, 2013, pp. 2200–2207.

[10] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars, "Unsupervised visual domain adaptation using subspace alignment," in *ICCV*, 2013, pp. 2960–2967.

[11] B. Sun, J. Feng, and K. Saenko, "Return of frustratingly easy domain adaptation," in *AAAI*, 2016, pp. 2058–2065.

[12] M. Long, J. Wang, G. Ding, Sinno Jialin Pan, and S Yu Philip, "Adaptation regularization: A general framework for transfer learning," *TKDE*, vol. 26, no. 5, pp. 1076–1089, 2014.

[13] J. Wang, Y. Chen, S. Hao, W. Feng, and Z. Shen, "Balanced distribution adaptation for transfor learning," in *ICDM*, 2017, pp. 1129–1134.

[14] J. Zhang, W. Li, and P. Ogunbona, "Joint geometrical and statistical alignment for visual domain adaptation," in *CVPR*, 2017, pp. 1859–1867.

[15] V. Cheng and C. Li, "Classification probabilistic pca with application in domain adaptation," in *PAKDD*, 2011, pp. 75–86.

[16] X. Zhu and Z. Ghahramani, "Learning from labeled and unlabeled data with label propagation," *Technical Report CMU-CALD-02-107*, 2002.

[17] Fujiwara Y. and Irie G., "Efficient label propagation," in *ICML*, 2014, pp. 784–792.

[18] G. Jun, Y. Quo, X. Kong, and R. He, "Unsupervised feature selection with ordinal locality," in *ICME*, 2017, pp. 1213–1218.

[19] A. Sharma and D. W Jacobs, "Bypassing synthesis: Pls for face recognition with pose, low-resolution and sketch," in *CVPR*, 2011, pp. 593–600.

[20] J. Blitzer, M. Dredze, and F. Pereira, "Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification," in *ACL*, 2007, pp. 440–447.

[21] M. Chen, Z. Xu, K. Weinberger, and F. Sha, "Marginalized denoising autoencoders for domain adaptation," in *ICML*, 2012, pp. 1627–1634.

[22] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *JMLR*, vol. 17, no. 1, pp. 2096–2030, 2016.