Semi-supervised Multi-view Individual and Sharable Feature Learning for Webpage Classification

Fei Wu*
College of Automation, Nanjing University of Posts and Telecommunications
China
wufei_8888@126.com

Xiao-Yuan Jing
School of Computer, Wuhan University,
College of Automation, Nanjing University of Posts and Telecommunications
China
jingxy_2000@126.com

Jun Zhou
College of Automation, Nanjing University of Posts and Telecommunications
China
darkness127@163.com

Yimu Ji
College of Computer, Nanjing University of Posts and Telecommunications
China

Chao Lan
Department of Computer Science, University of Wyoming
United States

Qinghua Huang
School of Science, Nanjing University of Posts and Telecommunications
China

Ruchuan Wang
College of Computer, Nanjing University of Posts and Telecommunications
China

ABSTRACT

Semi-supervised multi-view feature learning (SMFL) is a feasible solution for webpage classification. However, how to fully extract the complementarity and correlation information effectively under semi-supervised setting has not been well studied. In this paper, we propose a semi-supervised multi-view individual and sharable feature learning (SMISFL) approach, which jointly learns multiple view-individual transformations and one sharable transformation to explore the view-specific property for each view and the common property across views. We design a semi-supervised multi-view similarity preserving term, which fully utilizes the label information of labeled samples and similarity information of unlabeled samples from both intra-view and inter-view aspects. To promote learning of diversity, we impose a constraint on view-individual transformation to make the learned view-specific features to be statistically uncorrelated. Furthermore, we train a linear classifier, such that view-specific and shared features can be effectively combined for classification. Experiments on widely used webpage datasets demonstrate that SMISFL can significantly outperform state-of-the-art SMFL and webpage classification methods.

CCS CONCEPTS

• **Information systems** → **World Wide Web**; *Data extraction and integration*; • **Computing methodologies** → machine learning.

KEYWORDS

Webpage classification, semi-supervised multi-view feature learning (SMFL), view-specific feature, view-shared feature

ACM Reference Format:
Fei Wu, Xiao-Yuan Jing, Jun Zhou, Yimu Ji, Chao Lan, Qinghua Huang, and Ruchuan Wang. 2019. Semi-supervised Multi-view Individual and Sharable Feature Learning for Webpage Classification. In *Proceedings of the 2019 World Wide Web Conference (WWW'19), May 13–17, 2019, San Francisco, CA, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3308558.3313492

1 INTRODUCTION

In World Wide Web, there exists a huge repository of data in the form of webpages, and the size of the web is growing day-by-day. This makes it challenging for users to search relevant contents accurately. Classification of webpages, that is allocating a webpage to one category that can describe its contents [14], is regarded to be able to help improve the quality of search results [7]. Webpage usually contains multiple types of information, such as images, hyperlinks, URLs, HTML tags, etc. Different types of information is combined to describe the same webpage. Thus, webpage classification is a typical kind of multi-view data-based classification task [20, 25, 29, 37, 38]. In webpage classification task, we can only make use of a limited number of labeled webpages with a large number of unlabeled pages, since manually labeling a large repository of webpages is time and labor consuming and is infeasible. Hence, webpage classification is a semi-supervised learning problem [39, 40]. Besides, since a page usually contains rich information, the webpage data is high-dimensional [16]. Due to these characteristics, semi-supervised multi-view feature learning (SMFL) is a feasible solution for webpage classification.

In recent years, some SMFL methods have been presented. Chen et al. [1] developed a semi-paired and semi-supervised generalized correlation analysis (S^2GCA) algorithm. Volpi et al. [23] addressed manifold-regularized semi-supervised kernel canonical correlation analysis (MR-skCCA) for hyperspectral data classification. Jing

*corresponding author

WWW '19, May 13–17, 2019, San Francisco, CA, USA

ACM ISBN 978-1-4503-6674-8/19/05.
https://doi.org/10.1145/3308558.3313492

et al. presented the uncorrelated semi-supervised intra-view and inter-view manifold discriminant (USI^{2}MD) [5] method and further designed the semi-supervised multi-view correlation feature learning (SMCFL) [6] method for webpage classification. Wan et al. [24] addressed the cost sensitive semi-supervised canonical correlation analysis (CS^{3}CCA) algorithm. Zhang et al. [36] presented generalized semi-supervised structured subspace learning (GSS-SL) for cross-modal retrieval. To our knowledge, three SMFL methods, i.e., S^{2}GCA, USI^{2}MD and SMCFL, have been successfully used for webpage classification.

1.1 Motivation and Contribution

Most of existing SMFL methods generally map data of multiple views into a holistically latent space either through pursuing one projective transformation for each view, e.g., S^{2}GCA, USI^{2}MD, GSS-SL, etc., or by searching for a common transformation, e.g., SMCFL. As discussed in [6], correlation information exploration is important for webpage data. However, the first type of SMFL methods cannot make full use of the correlation information across multiple views. For the second type of SMFL methods, they cannot well preserve specific characteristic of each view. Therefore, how to effectively explore and utilize both the complementary information (particularity) and the correlation information (commonality) of different views is a significant research topic.

A few supervised/unsupervised multi-view learning methods have tried to preserve the specific characteristic of each view and uncover the shared component simultaneously [3, 12, 30]. Inspired by these works, we propose a novel SMFL approach for webpage classification. The contributions of our study can be summarized as following four points:

(1) We design a semi-supervised multi-view similarity preserving term. It jointly learns multiple view-individual transformations and one sharable transformation for exploring the particularity and the commonality of different views in the semi-supervised scenario. With this term, label information of labeled samples and the similarity information of unlabeled samples can be well utilized from both intra-view and inter-view aspects.

(2) To reduce the redundancy information across views, we design a multi-view statistical uncorrelation term, which makes the learned view-specific features to be statistically uncorrelated.

(3) We train a linear classifier based on view-specific and shared features, which further improves the discriminability of view-specific and shared features.

(4) By combining these designed terms, we propose a novel SMFL approach, named semi-supervised multi-view individual and sharable feature learning (SMISFL). Experiments on the widely used WebKB [1] and Internet Advertisements [8] datasets demonstrate the proposed approach can bring significant performance improvement for webpage classification.

2 RELATED WORK

2.1 Webpage Classification Methods

A number of webpage classification methods have been developed over the past decade, and most of these methods adopt machine learning-based techniques for performing the task [11, 28, 32]. Li et al. [10] presented the two-view transductive SVM (TTSVM) algorithm, which improves transductive SVM by employing a multi-view learning paradigm, and evaluated the effectiveness of the algorithm with webpage data. Wang et al. [26] studied large-scale cross-language webpage classification via dual knowledge transfer by using the nonnegative matrix trifactorization technique. Shivashankar et al. [19] built a positive and unlabeled learning model to determine whether two endpoint URLs refer to the same entity with the distant supervision. Saleh et al. [15] introduced a strategy called classification using multilayered domain ontology (CMDO) for vertical web page classification. These webpage classification works do not take all three characteristics of webpage classification into consideration.

In recent years, three SMFL methods, i.e., S^{2}GCA [1], USI^{2}MD [5], and SMCFL [6], have been applied to the webpage classification task. S^{2}GCA performs canonical correlation analysis on paired data, and learns features with the aim of preserving global structure of unlabeled data and achieving separability of different classes. USI^{2}MD learns a projective transformation for each view, and employs the semi-supervised intra-view and inter-view manifold discriminant learning schema and the semi-supervised uncorrelation constraint for discriminant features learning. SMCFL only learns a multi-view shared transformation that projects samples of multiple views to a common space, by considering the discriminant correlation of labeled and unlabeled samples. These three methods do not consider to explore both the particularity and the commonality across views for helping classification. In addition, they do not consider classifier design when training the classification models.

2.2 SMFL Methods

Multi-view learning in semi-supervised scenario is a hot research topic [2, 4, 13, 17, 31, 34, 35]. In recent years, some SMFL methods have been presented. MR-skCCA [23] performs kernel canonical correlation analysis between views on labeled and unlabeled data with the graph Laplacian-based manifold regularization. CS^{3}CCA [24] first uses the L2-norm manner to obtain the soft label for unlabeled data, and then embeds the misclassification cost into canonical correlation analysis. GSS-SL [36] designs the label graph constraint to help predict relevant class labels for unlabeled data, and takes the label space as a linkage to model the correlation among different modalities. Existing SMFL methods do not fully take exploration and utilization of both the view-specific and shared information into consideration.

Recently, a few supervised/unsupervised multi-view learning methods have tried to make use of the view-specific and shared information [3, 12, 30]. Xue et al. [30] presented a multi-view correlated feature learning with shared component algorithm, which learns individual information in each view and tries to capture feature correlation among multiple views by learning a shared component. Hu et al. [3] presented a sharable and individual multi-view deep metric learning (MvDML) algorithm, which learns an individual distance metric for each view to retain its specific property and learns a shared representation for different views to preserve their common properties based on deep neural network. These works are supervised/unsupervised methods without considering the joint use of labeled and unlabeled data. And these works do not well study either the aspect of how to effectively extract discriminant features with low redundancy or the aspect of how to learn classifier with the aim of effectively using both the view-specific and shared features jointly.

3 PROPOSED APPROACH

3.1 Semi-supervised Multi-view Similarity Preserving

Let $X_m \in \mathbb{R}^{d\times N}$ be the training webpage sample set of the m^{th} ($m=1,\cdots,M$) view, where M denotes the number of views and N is the number of samples in each view. Suppose that X_m consists of two parts of samples, i.e., $X_m = \{x_m^i\}_{i=1}^N = \left[\bar{X}_m, \widehat{X}_m\right]$, where $\bar{X}_m \in \mathbb{R}^{d\times N_l}$ denotes the labeled data set in the m^{th} view and $\widehat{X}_m \in \mathbb{R}^{d\times N_u}$ contains unlabeled data in the m^{th} view. Here, N_l and N_u separately denote the number of samples in $\bar{X}_m$ and $\widehat{X}_m$. We aim to learn the transformation $V_m = \begin{pmatrix} V_m^s \\ V^c \end{pmatrix} \in \mathbb{R}^{(d_v+d_c)\times d}$, $m=1,\cdots,M$, such that sample x_m^i can be mapped into a discriminant subspace through $V_m x_m^i$. Here, $V_m^s \in \mathbb{R}^{d_v\times d}$ is the view-specific transformation, which embodies the intrinsic characteristic of the m^{th} view, and $V^c \in \mathbb{R}^{d_c\times d}$ is the view-sharable transformation, which reflects the correlation information of M views. For simplicity, d_c is considered to be equal to d_v in this paper.

For samples x_m^i and x_l^j, we define the following weight to characterize their similarity

$$w_{m,l}^{i,j} = \begin{cases} 1, \text{if } x_m^i \text{ and } x_l^j \text{ are from the same class} \\ -1, \text{if } x_m^i \text{ and } x_l^j \text{ are from different classes} \\ \exp\left(-\left\|x_m^i - x_l^j\right\|_2^2 \Big/ (2\sigma^2)\right), \text{if } x_m^i \text{ or } x_l^j \text{ is unlabeled} \end{cases}$$

where σ is calculated as the standard variation of samples in the m^{th} and the l^{th} views. $W_{m,l}$ ($m=1,\cdots,M;\ l=1,\cdots,M$) is the similarity matrix between the m^{th} and l^{th} views. And the overall similarity matrix can be defined as $W = \begin{bmatrix} W_{1,1} & W_{1,2} & \cdots & W_{1,M} \\ W_{2,1} & W_{2,2} & \cdots & W_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ W_{M,1} & W_{M,2} & \cdots & W_{M,M} \end{bmatrix}$.

In order to learn discriminant transformations, such that in the projected space, limited number of labeled samples from the same class cluster together, the labeled samples with different class labels can be well separable, and unlabeled samples can preserve the global structure, we define the following semi-supervised multi-view similarity preserving term

$$\begin{aligned} J_1(V_m|_{m=1}^M) &= \frac{1}{2M^2N^2}\sum_{m=1}^M\sum_{i=1}^N\sum_{l=1}^M\sum_{j=1}^N w_{m,l}^{i,j}\left\|\begin{pmatrix} V_m^s \\ V^c \end{pmatrix}x_m^i - \begin{pmatrix} V_l^s \\ V^c \end{pmatrix}x_l^j\right\|_F^2 \\ &= \frac{1}{M^2N^2}\sum_{m=1}^M\sum_{l=1}^M tr\left(V_m X_m L_{m,l} X_l^T V_l^T\right) \end{aligned} \quad (1)$$

where L is the Laplacian matrix containing the sub-matrices $L_{m,l}$ ($m=1,\cdots,M;\ l=1,\cdots,M$), and L is defined as $L = D - W$. Here, D is a diagonal matrix with the i^{th} diagonal element being calculated as the sum of the i^{th} row of the overall similarity matrix W. We should minimize J_1.

This term can fully utilize the label information of labeled samples and similarity information of unlabeled samples from both intra-view and inter-view aspects.

3.2 Multi-view Statistical Uncorrelation

In order to reduce the redundancy information across views and learn pure view-specific features with view-individual transformations, we design the following multi-view statistical uncorrelation term

$$J_2(V_m^s|_{m=1}^M) = \frac{1}{M(M-1)}\sum_{m=1}^M\sum_{l=1,m\neq l}^M \left\|V_m^s Cov_{m,l} V_l^{s\,T}\right\|_F^2 \quad (2)$$

where $Cov_{m,l} = \frac{1}{N}\sum_{i=1}^N \left(x_m^i - \mu_m\right)\left(x_l^i - \mu_l\right)^T$ is the cross-covariance matrix between the m^{th} and the l^{th} views. Here, μ_m and μ_l are separately the mean samples of X_m and X_l. We should also minimize this term to enhance the complementarity of features extracted from different views.

3.3 View-specific and Shared Features-based Classification Loss

We also design a linear classifier, such that the learned view-specific and view-shared features can be effectively combined for classification. The designed view-specific and shared features-based classification loss term is given as

$$J_3(P,b) = \left\|P\left(\left(V_1\bar{X}_1\right)^T,\cdots,\left(V_M\bar{X}_M\right)^T\right)^T + b\mathrm{I}^T - Y\right\|_F^2 \quad (3)$$

where $Y \in \mathbb{R}^{C\times N_l}$ is a label matrix with each column being a one-hot label vector, $P = [P_1, P_2, \cdots, P_{2m-1}, P_{2m}, \cdots, P_{2M-1}, P_{2M}] \in \mathbb{R}^{C\times(2Md_v)}$ is the projective matrix, $b \in \mathbb{R}^C$ is a bias term, and $\mathrm{I} \in \mathbb{R}^{N_l}$ is a vector full of 1. Here, C denotes the number of classes, and each $P_{2m-1} \in \mathbb{R}^{C\times d_v}$ and $P_{2m} \in \mathbb{R}^{C\times d_v}$ are sub-matrices of P. With this term, the discriminant ability of learned features can be further improved.

3.4 Total Objective of SMISFL

By combining the terms of semi-supervised multi-view similarity preserving, multi-view statistical uncorrelation, and the classification loss, we formulate the overall objective as

$$\begin{aligned} \min_{V_1,\cdots,V_M,P,b} J = &\frac{1}{2M^2N^2}\sum_{m=1}^M\sum_{i=1}^N\sum_{l=1}^M\sum_{j=1}^N w_{m,l}^{i,j}\left\|\begin{pmatrix} V_m^s \\ V^c \end{pmatrix}x_m^i - \begin{pmatrix} V_l^s \\ V^c \end{pmatrix}x_l^j\right\|_F^2 \\ &+\alpha\frac{1}{M(M-1)}\sum_{m=1}^M\sum_{l=1,m\neq l}^M\left\|V_m^s Cov_{m,l} V_l^{s\,T}\right\|_F^2 \\ &+\beta\left\|P\left(\left(V_1\bar{X}_1\right)^T,\cdots,\left(V_M\bar{X}_M\right)^T\right)^T + b\mathrm{I}^T - Y\right\|_F^2 \\ &+\gamma\left(\frac{1}{M}\sum_{m=1}^M\left\|V_m^s\right\|_{2,1} + \|V^c\|_F^2\right) \end{aligned} \quad (4)$$

where α, β and γ are balance factors. $\frac{1}{M}\sum_{m=1}^M\left\|V_m^s\right\|_{2,1} + \|V^c\|_F^2$ is a regularization term to avoid the overfitting problem. We employ the $\ell_{2,1}$-norm based regularizer, which makes V_m^s, $m=1,\cdots,M$ to be sparse in rows, such that discriminant features can be selected for each view.

3.5 Optimization

The objective (4) is not jointly convex with regard to variables $\left\{V_m^s|_{m=1}^M, V^c, P, b\right\}$ simultaneously. We thus employ an alternating optimization strategy to optimize these variables.

(1) Fix $\left\{V_1^s, V_2^s, \cdots, V_M^s\right\}$ **and** V^c**, and update** P, b

The objective function (4) is reduced to

$$\min_{P,b} J_3 = \left\| P\left(\left(V_1 \bar{X}_1\right)^T, \cdots, \left(V_M \bar{X}_M\right)^T \right)^T + b\mathrm{I}^T - Y \right\|_F^2 \quad (5)$$

When updating $P(b)$, $b(P)$ is fixed. By setting the derivative of J_3 w.r.t. P, b to zero, we have

$$P = -BA^T\left(AA^T\right)^{-1}, b = -\frac{1}{N_l} D\mathrm{I} \quad (6)$$

where $A = \left(\left(V_1\bar{X}_1\right)^T, \cdots, \left(V_M\bar{X}_M\right)^T\right)^T$, $B = b\mathrm{I}^T - Y$, and $D = P\left(\left(V_1\bar{X}_1\right)^T, \cdots, \left(V_M\bar{X}_M\right)^T\right)^T - Y$.

(2) Fix P **and** b**, and update** $\left\{V_1^s, V_2^s, \cdots, V_M^s\right\}$ **and** V^c

We update V_m^s view by view. When updating V_m^s, V^c and $\left\{V_l^s\right\}_{l \neq m}$ are fixed. The objective function (4) is reduced to

$$\begin{aligned} \min_{V_m^s} \frac{1}{M^2N^2} & \left(\begin{array}{l} tr\left(V_m^s X_m L_{m,m} X_m^T V_m^{sT}\right) \\ +2 \sum\limits_{l \neq m}^{M} tr\left(V_m^s X_m L_{m,l} X_l^T V_l^{sT}\right) \end{array} \right) \\ & + \alpha \frac{2}{M(M-1)} \sum_{l \neq m}^{M} \left\| V_m^s Cov_{m,l} V_l^{sT} \right\|_F^2 \\ & + \beta \left\| P_{2m-1} V_m^s \bar{X}_m + E_1 \right\|_F^2 + \gamma \frac{1}{M} \left\| V_m^s \right\|_{2,1} \end{aligned} \quad (7)$$

where $E_1 = \sum\limits_{l \neq m}^{M} P_{2l-1} V_l^s \bar{X}_l + \sum\limits_{m=1}^{M} P_{2m} V^c \bar{X}_m + \left(b\mathrm{I}^T - Y\right)$.

Differentiating (7) with respect to V_m^s and setting it to zero, we have

$$\begin{aligned} & \frac{1}{M^2N^2} \left(V_m^s \left(X_m L_{m,m} X_m^T\right) + \sum_{l \neq m}^{M} \left(V_l^s X_l L_{m,l}^T X_m^T\right) \right) \\ & + \alpha \frac{2}{M(M-1)} \sum_{l \neq m}^{M} \left(V_m^s Cov_{m,l} V_l^{sT} V_l^s Cov_{m,l}^T\right) \\ & + \beta \left(P_{2m-1}^T P_{2m-1} V_m^s \bar{X}_m \bar{X}_m^T + P_{2m-1}^T E_1 \bar{X}_m^T\right) + \gamma \frac{1}{M} G_m^s V_m^s = 0 \end{aligned} \quad (8)$$

where G_m^s is a diagonal matrix with its i^{th} diagonal element as $G_m^s(i,i) = 1/\left(2\left\|V_m^s(i,:)\right\|\right)$. Here $V_m^s(i,:)$ represents the i^{th} row of V_m^s. According to [33], the element in G_m^s is usually calculated as $G_m^s(i,i) = \frac{1}{2\sqrt{\|V_m^s(i,:)\|^2 + \varepsilon}}$, where ε is set as a small positive value.

Then, we can obtain V_m^s by solving the following problem

$$V_m^s H_m^s + Q_m^s V_m^s + K_m^s V_m^s O_m^s = R_m^s \quad (9)$$

where $H_m^s = \frac{1}{M^2N^2} X_m L_{m,m} X_m^T + \frac{2\alpha}{M(M-1)} \sum\limits_{l \neq m}^{M} Cov_{m,l} V_l^{sT} V_l^s Cov_{m,l}^T$, $Q_m^s = \gamma \frac{1}{M} G_m^s$, $K_m^s = \beta P_{2m-1}^T P_{2m-1}$, $O_m^s = \bar{X}_m \bar{X}_m^T$, and $R_m^s = -\left(\frac{1}{M^2N^2} \sum\limits_{l \neq m}^{M} \left(V_l^s X_l L_{m,l}^T X_m^T\right) + \beta P_{2m-1}^T E_1 \bar{X}_m^T\right)$. The optimization problem in (9) can be solved by using the algorithm in the literatures [9, 22].

Algorithm 1 SMISFL

(1) **Input**: Multi-view training samples X_m, $m = 1, \cdots, M$, test sample $y = \{y_m\}|_{m=1}^M$.
(2) **Output**: The class label of y.
 (a) Randomly initialize $V_m^s|_{m=1}^M$, V^c, P and b.
 (b) **Repeat**:
 (c) Update P and b with Eq. (6).
 (d) Update $V_m^s|_{m=1}^M$ and V^c by solving (9) and (11).
 (e) **Until Convergence**
 (f) Map $\{y_m\}|_{m=1}^M$ with the learned transformations $V_m^s|_{m=1}^M$ and V^c by $\left\{\begin{array}{c} V_m^s \\ V^c \end{array}\right\} y_m$, and then utilize the classifier parameters P and b for classification.

When updating V^c, $\left\{V_1^s, V_2^s, \cdots, V_M^s\right\}$ is fixed. The objective function (4) can be reduced to

$$\begin{aligned} \min_{V^c} & \frac{1}{M^2N^2} \sum_{m=1}^{M} \sum_{l=1}^{M} tr\left(V^c X_m L_{m,l} X_l^T V^{cT}\right) \\ & + \beta \left\| \sum_{m=1}^{M} P_{2m} V^c \bar{X}_m + E_2 \right\|_F^2 + \gamma \left\|V^c\right\|_F^2 \end{aligned} \quad (10)$$

where $E_2 = \sum\limits_{m=1}^{M} P_{2m-1} V_m^s \bar{X}_m + \left(b\mathrm{I}^T - Y\right)$.

By firstly differentiating (10) with respect to V^c and setting it to zero, we can obtain V^c by solving the following problem

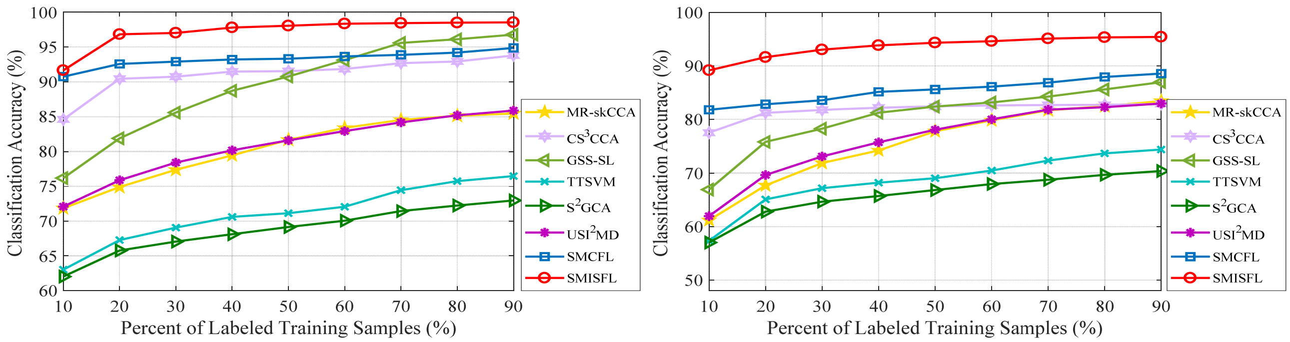
$$V^c H^c + Q^c V^c + \sum_{m=1}^{M} \sum_{l=1}^{M} K_{ml}^c V^c O_{ml}^c = R^c \quad (11)$$

where $H^c = \frac{1}{M^2N^2} \sum\limits_{m=1}^{M} \sum\limits_{l=1}^{M} X_m L_{m,l} X_l^T$, $Q^c = \gamma \mathrm{I}$, $K_{ml}^c = \beta P_{2m}^T P_{2l}$, $O_{ml}^c = \bar{X}_l \bar{X}_m^T$, and $R^c = -\beta \sum\limits_{m=1}^{M} P_{2m}^T E_2 \bar{X}_m^T$. The problem (11) can also be solved by using the algorithm in the literatures [9, 22]. Algorithm 1 summarizes the proposed SMISFL approach.

4 EXPERIMENTS

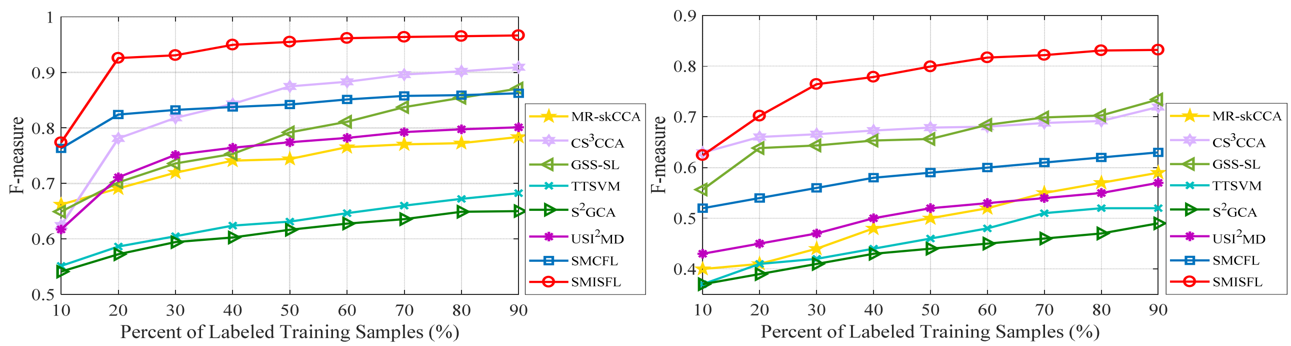
4.1 Dataset

We validate the effectiveness of our approach on two widely used datasets, i.e., WebKB [1] and Internet Advertisements [8] (AD). WebKB contains 1051 webpages, described by two views, i.e., the page view that includes the textual content on the page and link view that includes the hyperlinks pointing to the page. 1051 pages can be categorized into two classes: the course class including 230 pages and the non-course class including 821 pages. We use the preprocessed version of this dataset provided by [18], where 3000-dimensional and 1840-dimensional feature vectors are separately extracted from the page and link documents. The AD dataset includes 3279 samples that can be categorized into two classes: 458 advertisement and 2821 non-advertisement samples. A preprocessed version of this dataset [8] is used, where each sample is regarded as a binary vector with quite large sparsity. As in [5, 6], three views, i.e., 495 base URL features, 472 destination URL features and 457 image URL features, are selected.



(a) On WebKB (b) On AD

Figure 1: Average classification accuracies of compared methods on WebKB and AD datasets.



(a) On WebKB (b) On AD

Figure 2: Average F-measure values of compared methods on WebKB and AD datasets.

Table 1: Average classification results (% for CA) and running time (seconds) of compared methods.

Method	WebKB				AD			
	CA	F1	TR	TE	CA	F1	TR	TE
CS^{3}CCA	91.13	0.84	0.56	0.13	81.80	0.68	5.80	1.37
GSS-SL	89.42	0.78	0.61	0.15	80.53	0.66	3.57	1.21
TTSVM	71.09	0.63	-	0.35	68.62	0.46	-	0.95
S^{2}GCA	68.75	0.61	2.84	0.34	65.96	0.43	7.78	1.12
USI^{2}MD	80.80	0.76	2.76	0.18	76.25	0.51	18.65	1.21
SMCFL	93.27	0.84	0.44	0.09	85.40	0.58	5.08	0.28
SMISFL	**97.25**	**0.93**	**0.21**	**0.05**	**93.63**	**0.77**	**0.45**	**0.08**

4.2 Experimental Settings

In experiments, we compare our SMISFL approach with two categories of state-of-the-art related methods including: (1) Webpage classification methods: **TTSVM** [10], **S^{2}GCA** [1], **USI^{2}MD** [5], and **SMCFL** [6]; (2) SMFL methods: **MR-skCCA** [23], **CS^{3}CCA** [24], and **GSS-SL** [36].

Like in [5, 6], PCA [21] is performed to reduce the dimensionalities of samples in different views to 1050 for WebKB and 456 for AD. On each dataset, we randomly select half of each class for training and use the remaining samples for testing. The training samples are further split into the labeled training sample set and the unlabeled training set according to a certain percentage (e.g., 20% of training samples used as labeled samples) for semi-supervised learning. TTSVM, S^{2}GCA and CS^{3}CCA were originally designed for two views-based problems. Following [5, 6], on AD dataset, two views that yield the best classification results are used for these three methods. For compared methods, we report their results with the listed results in [6] or with the codes provided by the authors. In experiments, the tuning parameters (α, β and γ in Eq. (4), and the dimensionality of V_m, i.e., d_v) of SMISFL are set as $\alpha = 1$, $\beta = 10^2$, $\gamma = 10^2$, and $d_v = 10$ on WebKB, and $\alpha = 1$, $\beta = 10^{-1}$, $\gamma = 10^2$, and $d_v = 10$ on AD.

4.3 Results and Analysis

In this paper, we adopt the measures of classification accuracy (CA) and F-measure (F1) [27] to evaluate the classification effects of our approach and compared methods. We perform webpage classification experiments with the ratio of the number of labeled training samples to the total number of training samples increasing from 10% to 90%. Figs. 1 and 2 show the results measured by CA and F1 on two datasets, respectively. In these figures, each result is the average result across 20 random running. To statistically observe the results in the figures, we also calculate the mean result of each method across different percentages of labeled training samples, and we list these mean results in Table 1.

In Figs. 1 and 2, and Table 1, our SMISFL approach can always obtain better classification results with respect to the CA and F1 measures than compared webpage classification methods and state-of-the-art SMFL methods. From Table 1, SMISFL can improve the classification results by **3.98%**=(97.25%-93.27%) for CA and **0.09**=(0.93-0.84) for F1 on WebKB, and improve the classification results by **8.23%**=(93.63%-85.40%) for CA and **0.09**=(0.77-0.68) for F1 on AD. Better F1 results also mean that SMISFL can well classify the samples of the minority class, since samples used for training the model are class-imbalanced. The reason for the classification result improvement against all compared methods mainly lies in that the complementary information and the correlation information of different views of webpage data is effectively explored and leveraged under semi-supervised setting for classification.

Table 1 also reports the average training (TR) and testing (TE) time (for the total test set) of compared methods in the semi-supervised learning case that 20 percent of training samples are used as labeled samples. Our hardware configuration comprises i7-8700K CPU and 32 GB memory. From the table, SMISFL needs significantly less training and testing time than compared webpage classification and SMFL methods.

4.4 Discussion

In this subsection, we evaluate the important components of SMISFL. We separately call the version of SMISFL without the semi-supervised multi-view similarity preserving term as **SMISFL-s**, the version of SMISFL without the multi-view statistical uncorrelation term as **SMISFL-u**, the version of SMISFL without the view-specific and shared features-based classification loss term as **SMISFL-c**. For SMISFL-c, we concatenate feature representations of multiple views and employ the nearest neighbor classifier for classification. Furthermore, we conduct SMISFL without learning the view-individual transformation or the view-sharable transformation, and separately call these versions as **SMSFL** and **SMIFL**. Table 2 shows the result comparison.

From the table, the classification results of SMISFL-s, SMISFL-u and SMISFL-c are obviously inferior to that of SMISFL, and when only one of view-specific and shared transformations is used, SMISFL experiences an obvious performance degradation. This result means that the learned complementary and correlation features from multi-view data are useful for webpage classification, and the joint utilization of view-specific and shared features helps improve the performance. In addition, the designed three terms can help learn features with favorable discriminative capability.

Table 2: Classification results of variants of SMISFL.

Method	WebKB		AD	
	CA(%)	F1	CA(%)	F1
SMISFL-s	83.36	0.69	89.72	0.66
SMISFL-u	96.86	0.89	92.64	0.73
SMISFL-c	90.87	0.78	89.49	0.70
SMSFL	95.43	0.89	90.96	0.65
SMIFL	94.93	0.90	90.12	0.72
SMISFL	**97.25**	**0.93**	**93.63**	**0.77**

5 CONCLUSION

In this paper, we propose a novel SMFL approach SMISFL for webpage classification. It jointly learns the view-individual and sharable transformations by fully using of the labeled and unlabeled webpage data. Discriminant features with complementarity and correlation are well explored and effectively combined for classification. Experimental results on two widely used datasets demonstrate that SMISFL outperforms state-of-the-art SMFL and webpage classification methods. Results also validate the effectiveness of the semi-supervised view-individual and sharable feature learning scheme.

6 ACKNOWLEDGMENT

The work in this paper was supported by the National Natural Science Foundation of China (No. 61702280), National Postdoctoral Program for Innovative Talents (No. BX20180146), Natural Science Foundation of Jiangsu Province (No. BK20170900), and Scientific Research Staring Foundation for Introduced Talents in NJUPT (NUPTSF, No. NY217009).

REFERENCES

[1] Xiaohong Chen, Songcan Chen, Hui Xue, and Xudong Zhou. 2012. A unified dimensionality reduction framework for semi-paired and semi-supervised multi-view data. *Pattern Recognition* 45, 5 (2012), 2005–2018.

[2] Chen Gong, Dacheng Tao, Stephen J Maybank, Wei Liu, Guoliang Kang, and Jie Yang. 2016. Multi-modal curriculum learning for semi-supervised image classification. *IEEE Transactions on Image Processing* 25, 7 (2016), 3249–3260.

[3] Junlin Hu, Jiwen Lu, and Yap-Peng Tan. 2018. Sharable and individual multi-view metric learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40, 9 (2018), 2281–2288.

[4] Yu Jiang, Jing Liu, Zechao Li, and Hanqing Lu. 2014. Semi-supervised unified latent factor learning with multi-view data. *Machine Vision and Applications* 25, 7 (2014), 1635–1645.

[5] Xiao-Yuan Jing, Qian Liu, Fei Wu, Baowen Xu, Yang-Ping Zhu, and Songcan Chen. 2015. Web page classification based on uncorrelated semi-supervised intra-view and inter-view manifold discriminant feature extraction. In *Proceedings of the 24th International Joint Conferences on Artificial Intelligence*. 2255–2261.

[6] Xiao-Yuan Jing, Fei Wu, Xiwei Dong, Shiguang Shan, and Songcan Chen. 2017. Semi-supervised multi-view correlation feature learning with application to webpage classification. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence*. 1374–1381.

[7] Vinod Kumar. 2016. *Design and implementation of an efficient framework for web page classification*. Ph.D. Dissertation. Thapar University.

[8] Nicholas Kushmerick. 1999. Learning to remove internet advertisements. In *Proceedings of the 3rd Annual Conference on Autonomous Agents*. 175–181.

[9] Peter Lancaster. 1970. Explicit solutions of linear matrix equations. *SIAM Rev.* 12, 4 (1970), 544–566.

[10] Guangxia Li, Kuiyu Chang, and Steven CH Hoi. 2012. Multiview semi-supervised learning with consensus. *IEEE Transactions on Knowledge and Data Engineering* 24, 11 (2012), 2040–2051.

[11] Daniel López-Sánchez, Juan M Corchado, and Angélica González Arrieta. 2017. A CBR system for image-based webpage classification: case representation with convolutional neural networks. In *Proceedings of the 30th International Florida Artificial Intelligence Research Society Conference*. 483–488.

[12] Shirui Luo, Changqing Zhang, Wei Zhang, and Xiaochun Cao. 2018. Consistent and specific multi-view subspace clustering. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*. 3730–3737.

[13] Feiping Nie, Jing Li, and Xuelong Li. 2017. Convex multiview semi-supervised classification. *IEEE Transactions on Image Processing* 26, 12 (2017), 5718–5729.

[14] Xiaoguang Qi and Brian D Davison. 2009. Web page classification: Features and algorithms. *Comput. Surveys* 41, 2 (2009), Artical No. 12.

[15] Ahmed I Saleh, Mohammed F Al Rahmawy, and Arwa E Abulwafa. 2017. A semantic based web page classification strategy using multi-layered domain ontology. *World Wide Web* 20, 5 (2017), 939–993.

[16] Esra Saraç. 2010. *Web page classification using ant colony optimization*. Ph.D. Dissertation. MS Thesis, Çukurova University.

[17] XiaoBo Shen and QuanSen Sun. 2014. A novel semi-supervised canonical correlation analysis and extensions for multi-view dimensionality reduction. *Journal of Visual Communication and Image Representation* 25, 8 (2014), 1894–1904.

[18] Vikas Sindhwani, Partha Niyogi, and Mikhail Belkin. 2005. Beyond the point cloud: from transductive to semi-supervised learning. In *Proceedings of the 22nd International Conference on Machine Learning*. ACM, 824–831.

[19] S Subramanian, Timothy Baldwin, Julian Brooke, and Trevor Cohn. 2017. Pairwise webpage coreference classification using distant supervision. In *Proceedings of the 26th International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee, 841–842.

[20] Shiliang Sun. 2013. A survey of multi-view machine learning. *Neural Computing and Applications* 23, 7-8 (2013), 2031–2038.

[21] Matthew Turk and Alex Pentland. 1991. Eigenfaces for recognition. *Journal of Cognitive Neuroscience* 3, 1 (1991), 71–86.

[22] Charles F Van Loan. 2000. The ubiquitous Kronecker product. *J. Comput. Appl. Math.* 123, 1-2 (2000), 85–100.

[23] Michele Volpi, Giona Matasci, Mikhail Kanevski, and Devis Tuia. 2014. Semi-supervised multiview embedding for hyperspectral data classification. *Neurocomputing* 145 (2014), 427–437.

[24] Jianwu Wan, Hongyuan Wang, and Ming Yang. 2017. Cost sensitive semi-supervised canonical correlation analysis for multi-view dimensionality reduction. *Neural Processing Letters* 45, 2 (2017), 411–430.

[25] Huibing Wang, Lin Feng, Xiangzhu Meng, Zhaofeng Chen, Laihang Yu, and Hongwei Zhang. 2017. Multi-view metric learning based on KL-divergence for similarity measurement. *Neurocomputing* 238 (2017), 269–276.

[26] Hua Wang, Feiping Nie, and Heng Huang. 2015. Large-scale cross-language web page classification via dual knowledge transfer using fast nonnegative matrix trifactorization. *ACM Transactions on Knowledge Discovery from Data* 10, 1 (2015), Article No. 1.

[27] Fei Wu, Xiao-Yuan Jing, Ying Sun, Jing Sun, Lin Huang, Fangyi Cui, and Yanfei Sun. 2018. Cross-project and within-project semisupervised software defect prediction: a unified approach. *IEEE Transactions on Reliability* 67, 2 (2018), 581–597.

[28] Ou Wu, Ruiguang Hu, Xue Mao, and Weiming Hu. 2014. Quality-based learning for web data classification. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence*. 194–200.

[29] Chang Xu, Dacheng Tao, and Chao Xu. 2013. A survey on multi-view learning. *arXiv preprint arXiv:1304.5634* (2013).

[30] Xiaowei Xue, Feiping Nie, Sen Wang, Xiaojun Chang, Bela Stantic, and Min Yao. 2017. Multi-view correlated feature learning by uncovering shared component. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence*. 2810–2816.

[31] Yang Yang, De-Chuan Zhan, Xiang-Rong Sheng, and Yuan Jiang. 2018. Semi-supervised multi-modal learning with incomplete modalities. In *Proceedings of the 27th International Joint Conferences on Artificial Intelligence*. 2998–3004.

[32] Yiming Yang, Lei Zhang, Guiquan Liu, and Enhong Chen. 2015. UPCA: An efficient URL-Pattern based algorithm for accurate web page classification. In *Proceedings of the 12th International Conference on Fuzzy Systems and Knowledge Discovery*. 1475–1480.

[33] Qiyue Yin, Shu Wu, and Liang Wang. 2017. Unified subspace learning for incomplete and unlabeled multi-view data. *Pattern Recognition* 67 (2017), 313–327.

[34] Deming Zhai, Hong Chang, Shiguang Shan, Xilin Chen, and Wen Gao. 2012. Multiview metric learning with global consistency and local smoothness. *ACM Transactions on Intelligent Systems and Technology* 3, 3 (2012), Article No. 53.

[35] Chenghao Zhang and Wei-Shi Zheng. 2017. Semi-supervised multi-view discrete hashing for fast image search. *IEEE Transactions on Image Processing* 26, 6 (2017), 2604–2617.

[36] Liang Zhang, Bingpeng Ma, Guorong Li, Qingming Huang, and Qi Tian. 2018. Generalized semi-supervised and structured subspace learning for cross-modal retrieval. *IEEE Transactions on Multimedia* 20, 1 (2018), 128–141.

[37] Lei Zhang, Yao Zhao, Zhenfeng Zhu, Dinggang Shen, and Shuiwang Ji. 2018. Multi-view missing data completion. *IEEE Transactions on Knowledge and Data Engineering* 30, 7 (2018), 1296–1309.

[38] Jing Zhao, Xijiong Xie, Xin Xu, and Shiliang Sun. 2017. Multi-view learning overview: Recent progress and new challenges. *Information Fusion* 38 (2017), 43–54.

[39] Zhi-Hua Zhou, De-Chuan Zhan, and Qiang Yang. 2007. Semi-supervised learning with very few labeled training examples. In *Proceedings of the 22nd AAAI Conference on Artificial Intelligence*. 675–680.

[40] Xiaojin Zhu and Andrew B Goldberg. 2009. Introduction to semi-supervised learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning* 3, 1 (2009), 1–130.