Machine Vision and Applications (2014) 25:1635–1645
DOI 10.1007/s00138-013-0556-3

SPECIAL ISSUE PAPER

Semi-supervised Unified Latent Factor learning with multi-view data

Yu Jiang · Jing Liu · Zechao Li · Hanqing Lu

Received: 6 February 2013 / Revised: 7 July 2013 / Accepted: 19 September 2013 / Published online: 25 October 2013

Abstract Explosive multimedia resources are generated on web, which can be typically considered as a kind of multi-view data in nature. In this paper, we present a Semi-supervised Unified Latent Factor learning approach (SULF) to learn a predictive unified latent representation by leveraging both complementary information among multiple views and the supervision from the partially label information. On one hand, SULF employs a collaborative Nonnegative Matrix Factorization formulation to discover a unified latent space shared across multiple views. On the other hand, SULF adopts a regularized regression model to minimize a prediction loss on partially labeled data with the latent representation. Consequently, the obtained parts-based representation can have more discriminating power. In addition, we also develop a mechanism to learn the weights of different views automatically. To solve the proposed optimization problem, we design an effective iterative algorithm. Extensive experiments are conducted for both classification and clustering tasks on three real-world datasets and the compared results demonstrate the superiority of our approach.

Y. Jiang · J. Liu (✉) · H. Lu
National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences,
Beijing, China
e-mail: jliu@nlpr.ia.ac.cn

Y. Jiang
e-mail: yjiang@nlpr.ia.ac.cn

H. Lu
e-mail: luhq@nlpr.ia.ac.cn

Z. Li
School of Computer Science, Nanjing University of Science
and Technology, Nanjing, China
e-mail: zechao.li@gmail.com

Keywords Multi-view learning · Semi-supervised learning · Unified latent factor learning · Nonnegative matrix factorization

1 Introduction

With the popularity of internet technologies, there are explosive multimedia resources available online. This necessitates effective techniques on data analysis and management [13,14,20,26]. We deem such massive web resources have the multi-view attribute in nature, i.e., they are possibly correlated to reflect a common topic. For instance, a news document can be reported in multiple languages, an image can be described from different visual aspects, such as color, texture, shape, etc., and a video is a natural aggregation of textual, visual and audio information. Accordingly, the proper exploration of the complementary information (or correlations) across multiple views is helpful to boost the performance of data analysis on web data.

Numerous efforts have been made on learning from multi-view data. Generally speaking, there are two main directions. One direction is the co-training style scheme. A prominent achievement in this area is the Co-training algorithm [3], which trains two classifiers separately on two different views and uses the predictions of one classifier on unlabeled examples to augment the training set of the other. The Co-EM algorithm [23], a probabilistic version of Co-training, uses hypotheses learned in one view to probabilistically label the samples in the other one. The main limitation of the Co-training style algorithms is that they treat information from different views on equal terms. It is unreasonable, since data qualities of different views are varied. Moreover, in each active learning process, they typically require retraining with

all available data. This is a heavy burden, especially when there is a large amount of training data.

The other main direction is the unified latent factor or subspace learning approach, which aims to obtain a compact latent representation by taking advantage of inherent structure and relations across multiple views. A classical example is Canonical Correlation Analysis (CCA) [5,15]. It is designed to discover a common subspace representation shared by multiple views. Two spectral clustering algorithms [16] are proposed with co-regularizing the clustering hypotheses across views and apply to multi-view clustering. Most of these methods are formulated as an unsupervised learning problem, and so the discovered latent factor or subspace has weak prediction ability. Though sufficient labeled data can be very expensive to obtain, more often than not there is partially labeled information available with the rapid increase of free on-line information such as user tagging, rating, etc. It is no doubt that the latent representations would be more discriminating if partially label information can be integrated. Some semi-supervised approaches have been proposed to learn a shared or common latent representation [2,6]. In this work, we propose a new semi-supervised algorithm based on Nonnegative Matrix Factorization (NMF) to exploit multi-view data. To our best knowledge, it is the first time to handle multi-view data for semi-supervised learning based on NMF.

Nonnegative Matrix Factorization is an effective factor learning method, and the nonnegative constraint leads to the parts-based representation of objects, which accords with the cognitive process of human brain from the psychological and physiological evidences [17]. Recently, several variants of NMF have been proposed [4,8,10,19,22]. However, most work focus on the applications with the single view data or by simply concatenating the multi-view features. Moreover, NMF, as an unsupervised learning algorithm, cannot incorporate the class information effectively. Thus, it would be of great benefit to extend the usage of NMF to explore the complementary information across multi-view data, and be able to import some label information into its learning process.

In this paper, we develop a novel algorithm called Semi-supervised Unified Latent Factor learning (SULF), to learn a compact unified latent subspace as well as a discriminating linear classifier for multi-view data. In SULF, we consider two aspects of minimizing problems. First, a multi-view collaborative NMF model is proposed to jointly minimize the reconstruction errors of the multi-view data matrices, while a unified latent space is required to be shared across multiple views. Second, a $l_{2,1}$-norm regularized regression model is employed to minimize the prediction loss on partially labeled data with the unified latent representation. Besides, the proposed model can also automatically estimate how much each view should be trusted to accommodate noisy or unreliable views. To jointly consider the above issues, the obtained parts-based representation with multi-view data is incorporated with the complementary information among different views as well as the supervision from the partially labeled data, and, therefore, can have more discriminating power. A multiplicative-based alternative algorithm is developed to solve the joint optimization problem. At last, experimental results on three real-world datasets verify the effectiveness of our method.

The remainder of this paper is organized as follows: In Sect. 2, we overview some related work on unified latent factor learning and semi-supervised latent factor learning. Section 3 gives a brief review of NMF and Semi-NMF. Sections 4 and 5 elaborate the model of SULF and an efficient iterative algorithm in detail. The experimental evaluations and discussions are presented in Sect. 6. Section 7 concludes this paper.

2 Related work

In recent years, a number of unified latent factor learning algorithms emerged to discover inherent structure and relations among multiple views. Among them, CCA and CCA-based algorithms are representative and widely used. Chaudhuri et al. [5] projected the data into a latent subspace aiming to find the directions that maximize the correlation between the two sets of projected representations. Sun et al. [25] developed the discriminating CCA for multi-view data, aiming to reduce feature dimension with discrimination by maximizing the within-class correlation while minimizing the between-class correlation. There are many other types of unified latent factor algorithms. Kumar et al. [16] proposed a multi-view clustering approach in the framework of spectral clustering. In essence, it learns a latent representation by using the philosophy of co-regularization. Chen et al. [6] learned a predictive subspace representation underlying multiple views by a large-margin approach which jointly maximizes data likelihood and minimizes a prediction loss on training data. Different from the above methods, our SULF is based on NMF and tries to learn a parts-based representation. Besides, most of the above methods are designed to deal with two-view conditions, while SULF is able to deal with multiple views. What is more, SULF could learn the weights of different views automatically which is very crucial, especially when we have no prior knowledge which view is the best.

Semi-supervised latent factor learning, which tries to find compact latent representation using small amount of labeled data together with large amount of unlabeled data, is of great interest both in theory and in practice. Chen et al. [7] proposed a new latent factor algorithm termed as semi-paired and semi-supervised generalized correlation analysis,

which can deal with semi-paired and semi-supervised multi-view data. Ando and Zhang. [2] presented a framework for semi-supervised learning, where a generative model is used to learn effective parametric feature representations for discriminating learning. There are also some semi-supervised latent factor algorithms base on NMF. Chen et al. [8] proposed a semi-supervised NMF framework for data clustering. Users are able to provide supervision in terms of pairwise constraints on a few data objects specifying whether they "must" or "cannot" be clustered together. Cai et al. [4] presented a graph regularized NMF (GNMF) approach to encode the geometrical information of the data space. When labeled information is available, it can be naturally incorporated into the graph structure. This gives rise to semi-supervised GNMF. Liu et al. [22] developed a semi-supervised matrix decomposition method, called Constrained NMF (CNMF). The central idea of this approach is that the data points from the same class should be merged together in the new representation space. Different from the first two work, SULF tries to learn a parts-based unified latent representation. And other than these above NMF-based work, SULF is able to not only utilize the local discriminating information captured from the limited labeled data, but also discover the inherent structures and relations among different views.

3 A brief review of NMF and semi-NMF

Given an input nonnegative data matrix $\mathbf{X} \in \mathbb{R}^{M\times N}$, each column of $\mathbf{X}$ is a sample of vector. NMF aims to find two nonnegative matrices $\mathbf{U} \in \mathbb{R}^{M\times K}$ and $\mathbf{V} \in \mathbb{R}^{K\times N}$ whose product can well approximate the original matrix $\mathbf{X}$. The cost function of standard NMF is defined as

$$\begin{aligned} &\min \quad ||\mathbf{X}-\mathbf{UV}||_F^2 \\ &\text{s.t.} \quad \mathbf{U},\mathbf{V} \geq 0 \end{aligned} \tag{1}$$

Although the objective functions (1) are convex in $\mathbf{U}$ only or $\mathbf{V}$ only, they are not convex in both variables together. Therefore, it is unrealistic to expect an algorithm to find the global minimum. Lee and Seung [18] proposed an iterative undate algorithm to find the locally optimal solution as follows:

$$\begin{aligned} u_{mk} &\leftarrow u_{mk}\frac{(\mathbf{XV}^T)_{mk}}{(\mathbf{UVV}^T)_{mk}} \\ v_{kn} &\leftarrow v_{kn}\frac{(\mathbf{U}^T\mathbf{X})_{kn}}{(\mathbf{U}^T\mathbf{UV})_{kn}} \end{aligned} \tag{2}$$

When the data matrix $\mathbf{X}$ is unconstrained (i.e., it may have mixed signs), Semi-NMF restricts $\mathbf{V}$ to be nonnegative while placing no restriction on the signs of $\mathbf{U}$. The cost function of Semi-NMF is defined as

$$\begin{aligned} &\min \quad ||\mathbf{X}-\mathbf{UV}||_F^2 \\ &\text{s.t.} \quad \mathbf{V} \geq 0 \end{aligned} \tag{3}$$

Ding et.al. [10] computes the Semi-NMF factorization via updating $\mathbf{U}$ and $\mathbf{V}$ alternatively:

$$\mathbf{U} = \mathbf{XV}^T(\mathbf{VV}^T)^{-1} \tag{4}$$

$$v_{kj} \leftarrow v_{kj}\sqrt{\frac{(\mathbf{U}^T\mathbf{X})_{kj}^+ + ((\mathbf{U}^T\mathbf{U})^-\mathbf{V})_{kj}}{(\mathbf{U}^T\mathbf{X})_{kj}^- + ((\mathbf{U}^T\mathbf{U})^+\mathbf{V})_{kj}}}, \tag{5}$$

where we separate the positive and negative parts of a matrix $\mathbf{B}$ as

$$\mathbf{B}^+ = (|\mathbf{B}|+\mathbf{B})/2 \quad \mathbf{B}^- = (|\mathbf{B}|-\mathbf{B})/2 \tag{6}$$

4 Semi-supervised Unified Latent Factor learning

Figure 1 illustrates the framework of SULF. Suppose there are N data points with P views, among which the first R data points are labeled. $\mathbf{X}^p \in \mathbb{R}^{M^p\times N}$ is the data matrix of pth view. $\mathbf{Y} \in \mathbb{R}^{C\times R}$ is the label matrix, which encodes the label information as

$$y_{cr} = \begin{cases} 1 & \text{if the } r\text{th data point belongs to the } c\text{th class} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

To achieve the ultimate target, SULF involves two components: the one to exploit the complementary information among multiple views, and the other to incorporate the label information from partially labeled data.

4.1 Multi-view collaborative NMF

Single-view information is usually only a kind of unilateral or partial reflection of data properties. In order to exploit multi-view information collaboratively, SULF simultaneously performs NMF with different view data matrices based on the underlying assumption that the distributions of samples in basis (topic) spaces are consistent across different views. In other words, different views share a common factor matrix $\mathbf{V}$.

Though all views share the same $\mathbf{V}$, it is not reasonable for them to play the same role during the learning process. This is because different views suffer from varying degrees of information loss and noise pollution. It is reasonable to expect that the best view is dominant, but usually we lack the prior knowledge which view is the best. SULF tries to learn the weights of different views automatically according to the reconstruction precision of data matrices.

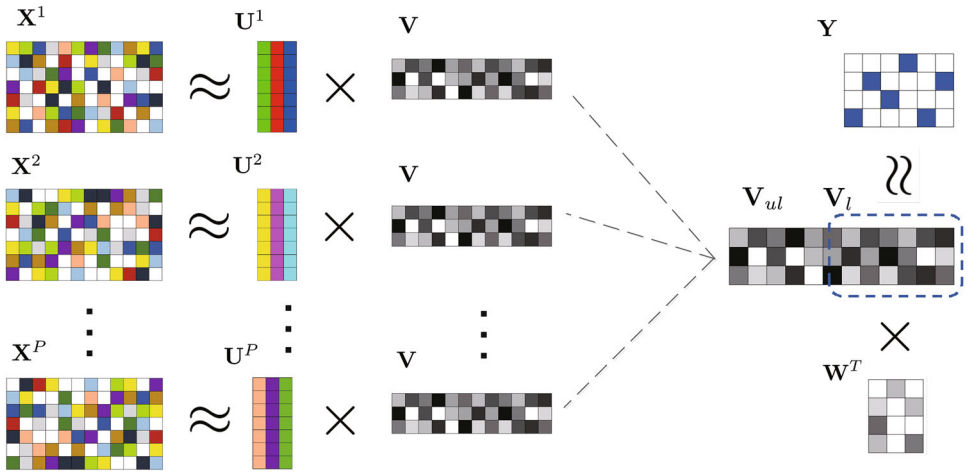


Fig. 1 The illustration of SULF

Thus, the objective function for the multi-view collaborative NMF is defined as

$$\begin{aligned} \min \quad & \sum_{p=1}^{P} \pi^p ||\mathbf{X}^p - \mathbf{U}^p \mathbf{V}||_F^2 + \lambda ||\mathbf{\Pi}||^2 \\ \text{s.t.} \quad & \mathbb{U}, \mathbf{V}, \mathbf{\Pi} \geq 0, \quad \sum_{p=1}^{P} \pi^p = 1, \end{aligned} \tag{8}$$

where $\mathbb{U} = \{\mathbf{U}^1, \mathbf{U}^2, \ldots, \mathbf{U}^P\}$ is the set of different view basis matrices. $\mathbf{\Pi} = (\pi^1, \pi^2, \ldots, \pi^P)$ is the weight vector of different views. The parameter λ controls the smoothness of $\mathbf{\Pi}$. The larger value of λ leads to the smoother view weights.

4.2 $l_{2,1}$-norm regularized regression with latent representation

In order to enhance the discriminating power of parts-based representation, SULF attempts to incorporate the partially label information. $\mathbf{V}_l \in \mathbb{R}^{K \times R}$, the first R columns of $\mathbf{V}$, is the compact representation of the first R labeled data, and $\mathbf{V}_{ul} \in \mathbb{R}^{K \times (N-R)}$ is the compact representation of the other $N-R$ unlabeled data, i.e., $\mathbf{V} = [\mathbf{V}_l, \mathbf{V}_{ul}]$. SULF learns a linear classifier $\mathbf{W} \in \mathbb{R}^{K \times C}$ with respect to $\mathbf{V}_l$ to fit the partially label information $\mathbf{Y}$ by minimize the following problem:

$$\begin{aligned} \min \quad & ||\mathbf{W}^T \mathbf{V}_l - \mathbf{Y}||_F^2 + \gamma ||\mathbf{W}||_{2,1} \\ \text{s.t.} \quad & \mathbf{V}_l \geq 0 \end{aligned} \tag{9}$$

where

$$||\mathbf{W}||_{2,1} = \sum_{k=1}^{K} \sqrt{\sum_{c=1}^{C} w_{kc}^2} \tag{10}$$

The $l_{2,1}$-norm regularization term [11,21] is introduced to ensure $\mathbf{W}$ sparse in rows. In that way, $\mathbf{W}$ does a feature selection during the fitting process that leads to a better reconstruction of label matrix $\mathbf{Y}$.

The minimization problem in Eq. 9 is similar to the problem of Semi-NMF [10]. That is, we should factorize the nonnegative matrix $\mathbf{Y}$ into a nonnegative matrix $\mathbf{V}_l$ and a matrix $\mathbf{W}$ with mixed signs.

4.3 Unified objective function

Considering the objective for multi-view Information and partially label information simultaneously, we obtain a unified objective function for SULF:

$$\begin{aligned} \min \quad & \sum_{p=1}^{P} \pi^p ||\mathbf{X}^p - \mathbf{U}^p \mathbf{V}||_F^2 + \lambda ||\mathbf{\Pi}||^2 \\ & + \beta ||\mathbf{W}^T \mathbf{V}_l - \mathbf{Y}||_F^2 + \gamma ||\mathbf{W}||_{2,1} \\ \text{s.t.} \quad & \mathbb{U}, \mathbf{V}, \mathbf{\Pi} \geq 0, \quad \sum_{p=1}^{P} \pi^p = 1, \end{aligned} \tag{11}$$

where β is a nonnegative parameter to trade off the aforementioned two objectives.

5 Optimization

The joint optimization function in (11) is not convex over all variables $\mathbb{U}$, $\mathbf{V}$, $\mathbf{W}$ and $\mathbf{\Pi}$ simultaneously. Thus, we propose an iterative optimization algorithm. For the ease of representation, we define

$$\begin{aligned} \mathcal{O}(\mathbb{U}, \mathbf{V}, \mathbf{W}, \mathbf{\Pi}) = & \sum_{p=1}^{P} \pi^p ||\mathbf{X}^p - \mathbf{U}^p \mathbf{V}||_F^2 + \lambda ||\mathbf{\Pi}||^2 \\ & + \beta ||\mathbf{W}^T \mathbf{V}_l - \mathbf{Y}||_F^2 + \gamma ||\mathbf{W}||_{2,1} \end{aligned} \tag{12}$$

Then, the joint optimization problem can be iteratively solved by the following four reduced subproblems: (1) fix $\mathbf{V}$,

Algorithm 1 Algorithm of SULF

Input:
P view data matrices $\mathbf{X}^1, \mathbf{X}^2, \ldots, \mathbf{X}^P$, label matrix $\mathbf{Y}$, parameters $\beta, \gamma, \lambda, K$
Output:
P view basis matrices $\mathbb{U} = \{\mathbf{U}^1, \mathbf{U}^2, \ldots, \mathbf{U}^p\}$, factor matrices $\mathbf{V}$, linear classifier $\mathbf{W}$,
Initialize $\mathbf{U}^1, \mathbf{U}^2, \ldots, \mathbf{U}^P, \mathbf{V}$;
Initialize $(\pi^1, \pi^2, \ldots, \pi^P) = (1/P, 1/P, \ldots, 1/P)$;
loop
  Fix $\mathbf{V}$, update $\mathbf{W}$ as in(15);
  for $p = 1$ to P **do**
    Fix $\mathbf{V}$, update $\mathbf{U}^p$ as in(19);
  end for
  Fix $\mathbb{U}$, $\mathbf{\Pi}$, update $\mathbf{V}_l$ as in(26);
  Fix $\mathbb{U}$, update $\mathbf{V}_{ul}$ as in(27);
  for $p = 1$ to P **do**
    Fix $\mathbb{U}$, $\mathbf{V}$, computer $c^p = ||\mathbf{X}^p - \mathbf{U}^p\mathbf{V}||_F^2$;
  end for
  Update $\mathbf{\Pi}$ using CVX;
end loop until convergence

minimize $\mathcal{O}(\mathbf{W})$; (2) fix $\mathbf{V}$, minimize $\mathcal{O}(\mathbb{U})$; (3) fix $\mathbb{U}$ and $\mathbf{\Pi}$, minimize $\mathcal{O}(\mathbf{V})$; and (4) fix $\mathbb{U}$ and $\mathbf{V}$, minimize $\mathcal{O}(\mathbf{\Pi})$. We summarize the updating algorithm in Algorithm 1.

5.1 The updating rule for $\mathbf{W}$

There is no nonnegative constraint for $\mathbf{W}$. Requiring the derivative of $\mathcal{O}(\mathbf{W})$ w.r.t. $\mathbf{W}$, we have

$$\frac{\partial \mathcal{O}}{\partial \mathbf{W}} = 2\left(\beta \mathbf{V}_l \left(\mathbf{W}^T\mathbf{V}_l - \mathbf{Y}\right)^T + \gamma \mathbf{E}\mathbf{W}\right) \tag{13}$$

Here, $\mathbf{E}$ is a diagonal matrix with $e_{kk} = \frac{1}{2||\mathbf{w}_k||_2}$. [1] Let $\frac{\partial \mathcal{O}}{\partial \mathbf{W}} = 0$, we get the following updating rule for $\mathbf{W}$:

$$\mathbf{W} = \left(\beta \mathbf{V}_l\mathbf{V}_l^T + \gamma \mathbf{E}\right)^{-1} \beta \mathbf{V}_l\mathbf{Y}^T \tag{14}$$

Let $\mathbf{A} = \left(\beta \mathbf{V}_l\mathbf{V}_l^T + \gamma \mathbf{E}\right)$, then

$$\mathbf{W} = \beta \mathbf{A}^{-1}\mathbf{V}_l\mathbf{Y}^T \tag{15}$$

5.2 The updating rule for $\mathbb{U}$

Since each basis matrix in $\mathbb{U}$ is completely symmetrical, we give a detailed analysis of the pth view, and other views can be derived analogously. Let ψ_{ik}^p is the Lagrange multiplier for constraint $u_{ik}^p \geq 0$, and $\mathbf{\Psi}^p = [\psi_{mk}^p]$, the Lagrange function $\mathcal{L}(\mathbf{U}^p)$ is defined as

$$\mathcal{L}(\mathbf{U}^p) = \mathcal{O}(\mathbf{U}^p) + Tr((\mathbf{\Psi}^p)^T\mathbf{U}^p) \tag{16}$$

[1] $\mathbf{w}_k$ is the kth row of $\mathbf{W}$. In practice, $||\mathbf{w}_k||_2$ could be close to zero but not zero. Theoretically, it could be zeros. For this case, we can let ε is very small constant, and regularize $e_{kk} = \frac{1}{2\sqrt{\mathbf{w}_k^T\mathbf{w}_k + \varepsilon}}$.

The partial derivatives of $\mathcal{L}(\mathbf{U}^p)$ with respect to $\mathbf{U}^p$ is

$$\frac{\partial \mathcal{L}(\mathbf{U}^p)}{\partial \mathbf{U}^p} = -2\mathbf{X}^p\mathbf{V}^T + 2\mathbf{U}^p\mathbf{V}\mathbf{V}^T + \mathbf{\Psi}^p \tag{17}$$

Using the Karush–Kuhn–Tucker condition $\psi_{mk}^p u_{mk}^p = 0$, we get the following equations for $\mathbf{U}^p$:

$$-(\mathbf{X}^p\mathbf{V}^T)_{mk}u_{mk}^p + (\mathbf{U}^p\mathbf{V}\mathbf{V}^T)_{mk}u_{mk}^p = 0 \tag{18}$$

This equation leads to the following updating rule for $\mathbf{U}^p$:

$$u_{mk}^p \leftarrow u_{mk}^p \frac{(\mathbf{X}^p\mathbf{V}^T)_{mk}}{(\mathbf{U}^p\mathbf{V}\mathbf{V}^T)_{mk}} \tag{19}$$

5.3 The updating rule for $\mathbf{V}$

$\mathbf{V}$ includes two parts $\mathbf{V}_l$ and $\mathbf{V}_{ul}$ as already stated. For the ease of representation, we also divide each data matrix $\mathbf{X}^p$ into two parts $\mathbf{X}_l^p$ and $\mathbf{X}_{ul}^p$ accordingly. Let ϕ_{kn} be the Lagrange multiplier for constraint $v_{kn} \geq 0$, and $\mathbf{\Phi} = [\mathbf{\Phi}_l, \mathbf{\Phi}_{ul}] = [\phi_{kn}]$, the Lagrange function $\mathcal{L}(\mathbf{V})$ be defined as

$$\mathcal{L}(\mathbf{V}) = \mathcal{O}(\mathbf{V}) + Tr(\mathbf{\Phi}^T\mathbf{V}) \tag{20}$$

Substituting $\mathbf{W}$ by Eq. (15) in to (20) and noticing $\mathbf{A}$ is a symmetric matrix, we have

$$\begin{aligned}\mathcal{L}(\mathbf{V}) = &\sum_{p=1}^{P} \pi^p ||\mathbf{X}^p - \mathbf{U}^p\mathbf{V}||_F^2 \\ &-\beta^2 Tr\left(\mathbf{Y}\mathbf{V}_l^T\mathbf{A}^{-1}\mathbf{V}_l\mathbf{Y}^T\right) + Tr(\mathbf{\Phi}\mathbf{V}) + F,\end{aligned} \tag{21}$$

where F is a constant. The partial derivatives of $\mathcal{L}(\mathbf{V})$ with respect to $\mathbf{V}_l$ and $\mathbf{V}_{ul}$ are [2]

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \mathbf{V}_l} = &2\sum_{p=1}^{P} \pi^p(-(\mathbf{U}^p)^T\mathbf{X}_l^p + (\mathbf{U}^p)^T\mathbf{U}^p\mathbf{V}_l) \\ &-2\beta^2\mathbf{A}^{-1}\mathbf{V}_l\mathbf{Y}^T\mathbf{Y} + \mathbf{\Phi}_l\end{aligned} \tag{22}$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{V}_{ul}} = 2\sum_{p=1}^{P} \pi^p(-(\mathbf{U}^p)^T\mathbf{X}_{ul}^p + (\mathbf{U}^p)^T\mathbf{U}^pV_{ul}) + \mathbf{\Phi}_{ul} \tag{23}$$

Using the Karush–Kuhn–Tucker condition $\phi_{kn}v_{kn} = 0$, we get the following equations for $\mathbf{V}_l$ and $\mathbf{V}_{ul}$:

$$\begin{aligned}&\sum_{p=1}^{P} \pi^p((\mathbf{U}^p)^T(\mathbf{U}^p\mathbf{V}_l - \mathbf{X}_l^p)_{kn}(v_l)_{kn} \\ &-\beta^2(\mathbf{A}^{-1}\mathbf{V}_l\mathbf{Y}^T\mathbf{Y})_{kn}(v_l)_{kn} = 0\end{aligned} \tag{24}$$

$$\sum_{p=1}^{P} \pi^p((\mathbf{U}^p)^T(\mathbf{U}^p\mathbf{V}_{ul} - \mathbf{X}_{ul}^p)_{kn}(v_{ul})_{kn} = 0 \tag{25}$$

[2] For convenience, $\mathbf{A}$ is approximately as constant matrix when requiring the derivatives of $\frac{\partial \mathcal{L}}{\partial \mathbf{V}_l}$.

Table 1 Statistics of three datesets

dataset	# of size (N)	# of class (C)	# of dimensionality for each view (M^1 / M^2 / M^3)
NUS-WIDE-Object	3100	31	1000 / 1000 / 1000
Corel5K	5000	50	1000 / 1000 / 1000
Reuters multilingual	1876	6	21531 / 11547/ 24893

Leting $\mathbf{B} = \mathbf{A}^{-1}\mathbf{V}_l\mathbf{Y}^T\mathbf{Y}$, these equations lead to the following updating rules for $\mathbf{V}_l$ and $\mathbf{V}_{ul}$:

$$(v_l)_{kn} \leftarrow (v_l)_{kn} \sqrt{\frac{\left(\sum_{p=1}^{P} \pi^p (\mathbf{U}^p)^T \mathbf{X}_l^p + \beta^2 \mathbf{B}^+\right)_{kn}}{\left(\sum_{p=1}^{P} \pi^p (\mathbf{U}^p)^T \mathbf{U}^p \mathbf{V}_l + \beta^2 \mathbf{B}^-\right)_{kn}}} \tag{26}$$

$$(v_{ul})_{kn} \leftarrow (v_{ul})_{kn} \sqrt{\frac{\left(\sum_{p=1}^{P} \pi^p (\mathbf{U}^p)^T \mathbf{X}_{ul}^p\right)_{kn}}{\left(\sum_{p=1}^{P} \pi^p (\mathbf{U}^p)^T \mathbf{U}^p \mathbf{V}_{ul}\right)_{kn}}} \tag{27}$$

The convergence of updating rules for $\mathbf{V}_l, \mathbf{V}_{ul}$ can be theoretically proved by a similar strategy to [10].

5.4 The updating rule for $\mathbf{\Pi}$

When $\mathbb{U}$, $\mathbf{V}$ are fixed, minimization of $\mathcal{O}(\mathbf{\Pi})$ is reduced to a simple convex optimization problem as follows:

$$\begin{aligned} \min \quad & \sum_{p=1}^{P} c^p \pi^p + \lambda ||\mathbf{\Pi}||^2 \\ \text{s.t.} \quad & \mathbf{\Pi} \geq 0, \quad \sum_{p=1}^{P} \pi^p = 1, \end{aligned} \tag{28}$$

where $c^p = ||\mathbf{X}^p - \mathbf{U}^p\mathbf{V}||_F^2$ is the reconstruction error of the pth view. We solve this convex optimization problem with CVX,[3] a Matlab-based modeling system for convex optimization.

The computational complexity of our proposed method is $O(\sum_{p=1}^{P} M_p N K)$, where P is the number of view, M_p is the dimension of the p-view feature, N is the number of instances and K is the dimension of latent space. It is also worth mentioning that the computational complexity for performing NMF with P view feature is also $O(\sum_{p=1}^{P} M_p N K)$. Thus, the computational cost of our approach is comparable with that of NMF.

6 Experiments

In this section, we evaluate the effectiveness of our proposed SULF algorithm on both classification and clustering tasks.

6.1 Date sets

Three real-world datesets are used in our experiments, including two image datasets and one document corpus. The important statistics of these datesets are summarized in Table 1. A brief description of each dataset is presented as follows:

NUS-WIDE-Object [9]: It consists of 31 object categories and 30,000 images in total. Images which belong to more than one category are eliminated. From the remaining 23,953 images, we randomly sample 100 images per category for model learning and evaluation. 1,000-dimension OpponentSIFT, C-SIFT and rgSIFT [24] are extracted by ColorDescriptor Software[4] as three views.

Corel5K [12]: It contains 5,000 images in 50 groups, such as fox, flower and bridge. Each group is composed of 100 images. Corel5K is represented by the same features as NUS-WIDE-Object dataset.

Reuters [1]: This collection contains documents originally written in five different languages (English, French, German, Spanish and Italian), and their translations, over a common set of 6 categories. The documents are represented as a bag of words using a TFIDF-based weighting scheme. We randomly sample 10 % documents from 18,758 documents originally in English. We use their original representation as the first view, their Spanish translation as the second view and their French translation as the third view. The vocabulary size of English is 21,531, while that of Spanish and French are 11,547 and 24,893, respectively.

6.2 Experimental setup

In order to validate the performance of learned latent factor, we compare the proposed SULF with several algorithms. The compared schemes are listed as follows:

NMF [18]: Nonnegative Matrix Factorization.

SGNMF [4]: Semi-supervised Graph Regularized Nonnegative Matrix Factorization. In SGNMF, an affinity graph is constructed to encode the geometrical information, and the label information is integrated into the graph structure as described in [22].

CNMF [22]: Constrained Nonnegative Matrix Factorization. CNMF incorporates the label information as additional constraints.

[3] http://cvxr.com/cvx/.

[4] http://koen.me/research/colordescriptors/.

Table 2 Characteristics of different algorithms

	NMF	SGNMF	CNMF	CCA	PCSC	CoTra	$SULF_1$	$SULF_2$	$SULF_3$	SULF
Unsupervised	✓			✓	✓		✓			
Semi-supervised		✓	✓			✓		✓	✓	✓
Single-view	✓	✓	✓					✓		
Two-view				✓						
Multi-view					✓	✓	✓		✓	✓

CCA [5]: Canonical Correlation Analysis. PCA is adopted as the pretreatment to denoising and reducing dimensions.

PCSC [16]: Pairwise Co-regularized Spectral Clustering. A spectral clustering algorithm encourages the pairwise similarities of examples under the new representation to be similar across all the views. It is considered as a clustering baseline.

CoTra [3]: Co-training scheme with regularized least square regression as the basis classification. PCA is adopted as the pretreatment to denoising and reducing dimensions. It is used as a classification baseline.

$SULF_1$: Our proposed method without labeled data utilized. That is, β is set to 0, and SULF degenerates into a unsupervised algorithm.

$SULF_2$: Our proposed method with single-view data available. That is, there is only one coefficient in $\mathbf{\Pi}$ is set to 1, others are set to 0, and SULF degenerates into a single-view algorithm.

$SULF_3$: Our proposed method without view weight learning mechanism. The weights is setup averagely, i.e., $\pi^p = 1/P$.

SULF: Our proposed method.

Table 2 summarizes the characteristics of all these algorithms. Except for CoTra and PCSC, all the other approaches are latent factor learning algorithms and aim to learn the compact latent representations. These latent representations are evaluated by both classification and clustering tasks. The dimensionality of the latent space for NUS-WIDE and Corel5K is set to 200 and that for Reuters is set to 50. CoTra is a multi-view classification method, and PCSC is a multi-view clustering one. The former is utilized as a classification baseline, while the latter is a clustering baseline.

For single-view algorithms, experiments are performed with both each single-view feature and concatenation features of three views. We report the best single-view result with the subscript "$_a$", and the result of concatenation features with subscript "$_b$". For example NMF_a indicates the best result of single-view feature achieved by NMF, while NMF_b indicates the result gotten from concatenation features. As for CCA, a two-view algorithm, we perform the algorithm with each pair of two-view data. And the result of best pair is reported.

In the classification task, for NMF, SGNMF, CNMF, CCA and $SULF_1$, we train a regularized least square regression model as the classifier with the latent representation of labeled samples. For $SULF_2$, $SULF_3$ and SULF, we perform classification through inputting the obtained latent representation of unlabeled data $\mathbf{V}_{ul}$ to the learned linear classifier $\mathbf{W}$ directly. As for the clustering task, the effectiveness of these latent representations is evaluated by K-means. And the clustering algorithm is only performed on the unlabeled data. That is to say, we discard the labeled samples after getting the latent representation.

For both classification and clustering task, 5 random train-test splits are applied, and 10 test runs are conducted for each split. The mean of the performance is reported. For each algorithm, we carry out threefold cross-validation to select the appropriate parameters.

6.3 Classification results

Table 3 shows the classification accuracy on the NUS-WIDE, Corel5K and Reuters datasets with different portion of labeled data. These experiments reveal a number of interesting points:

- SULF achieves good performance on all these datasets. Especially, when small portion (10, 20 %) labeled data are available, SULF outperforms all the other baselines. In addition, $SULF_2$ and $SULF_3$ also achieve good performance, which suggests the validity of our algorithm.
- Multi-view algorithms are superior to the single-view algorithms on three datesets in general. All the best performance is obtained by multi-view algorithms on three datasets with different label ratio. And SULF also achieves different degrees of improvement comparing with $SULF_2$. That illustrates there is indeed complementary information (or relations) embodied in multiple views.
- On the Reuters dataset, SULF as a semi-supervised algorithm is superior to all the unsupervised algorithms. And on other two image datasets, SULF also achieves the best performance when there is a small proportion of labeled data, while the performance of SULF is a little lower than that of CCA when there are abundant labeled data. This may be because in the complex feature space, it is more challenging to incorporate the label information.

Table 3 Classification performance on NUS-WIDE, Corel5K and Reuters datasets

	Unsupervised			Semi-supervised								
τ	NMF_a	NMF_b	CCA	$SULF_1$	CoTra	$CNMF_a$	$CNMF_b$	$SGNMF_a$	$SGNMF_b$	$SULF_2$	$SULF_3$	SULF
NUS-WIDE-Object												
0.1	11.54	13.59	12.22	12.83	10.25	12.88	14.22	14.09	15.15	15.90	17.11	**18.17**
0.2	13.96	16.30	17.78	16.51	15.32	14.39	15.96	14.95	16.47	17.04	18.89	**19.60**
0.3	15.58	17.75	19.96	18.52	18.63	17.93	18.70	16.46	17.47	17.82	19.41	**20.10**
0.4	17.36	19.65	**22.31**	20.45	21.18	18.89	20.13	16.56	17.58	18.95	20.16	21.29
0.5	18.59	20.69	**23.68**	21.21	21.55	19.10	21.18	18.84	18.90	18.55	21.71	22.13
Corel5K												
0.1	18.82	21.57	26.58	24.02	24.38	22.00	22.70	23.62	25.24	28.86	30.26	**31.18**
0.2	27.69	30.28	33.70	31.60	33.68	26.62	27.19	30.46	31.11	32.08	34.23	**34.58**
0.3	30.14	31.44	**35.95**	34.28	36.07	27.13	28.91	32.98	32.58	32.57	35.45	35.61
0.4	32.28	32.21	**36.67**	36.22	36.17	28.22	32.87	32.98	35.67	33.80	35.38	35.90
0.5	32.92	33.33	**38.72**	36.34	37.04	30.69	33.33	33.47	36.31	33.76	36.30	36.52
Reuters												
0.1	70.25	71.78	73.29	74.61	74.05	73.90	71.00	72.18	72.50	74.08	75.27	**76.36**
0.2	72.00	72.93	75.37	76.33	76.07	74.50	74.20	73.05	73.83	77.30	76.57	**78.67**
0.3	72.44	73.88	76.77	76.82	77.91	76.39	72.85	74.22	73.84	79.78	77.49	**80.05**
0.4	73.16	74.33	77.14	77.51	78.40	76.22	73.47	74.26	74.02	80.36	78.27	**80.80**
0.5	73.12	74.78	78.34	77.93	77.40	76.17	75.91	74.21	74.11	80.17	78.73	**80.70**

τ is the proportion of labeled data
Bold values indicate the best performance in the comparison

Besides, classification actually is a supervised task. For the unsupervised algorithm, though the labeled data do not participate into latent factor learning directly, a discriminating classifier is learned from abundant labeled data.

- On the Reuters dataset, $CNMF_a$ overmatches $CNMF_b$, and $SGNMF_a$ overmatches $SGNMF_b$ when label ratio is more than 20 %. $SULF_2$ also obtains better performance than $SULF_3$. All of the above phenomena attest that it is very critical to leverage information from different views properly.
- The performance comparison of SULF and $SULF_3$ shows the effectiveness of the mechanism to estimate the weights of different view.

6.4 Clustering results

Two metrics, the accuracy and the normalized mutual information are used to measure the clustering performance [22].

Accuracy (ACC): Given a document d_i, let l_i and r_i be the cluster label and the label provided by the document corpus, respectively. The ACC is defined as follows:

$$\mathrm{ACC} = \frac{\sum_{i=1}^{n} \delta(r_i, map(l_i))}{n}, \tag{29}$$

where n denotes the total number of documents in the test, $\delta(x, y)$ is the delta function that equals one if $x = y$ and equals zero otherwise, and $map(l_i)$ is the mapping function that maps each cluster label l_i to the equivalent label from the document corpus. The best mapping can be found using the Kuhn–Munkres algorithm.

Normalized Mutual Information (NMI): Let C denote the set of clusters obtained from the ground truth and $\hat{C}$ obtained from our algorithm. Their mutual information metric $\mathrm{MI}(C, \hat{C})$ is defined as follows:

$$\mathrm{MI}(C, \hat{C}) = \sum_{c_i \in C, \hat{c}_j \in \hat{C}} p(c_i, \hat{c}_j) \log_2 \frac{p(c_i, \hat{c}_j)}{p(c_i)p(\hat{c}_j)}, \tag{30}$$

where $p(c_i)$ and $p(\hat{c}_j)$ are the probabilities that a document arbitrarily selected from the corpus belongs to the clusters c_i and $\hat{c}_j$, respectively, and $p(c_i, \hat{c}_j)$ is the joint probability that the arbitrarily selected document belongs to the clusters c_i as well as $\hat{c}_j$ at the same time. In our experiments, we use the normalized mutual information NMI as follows:

$$\mathrm{NMI}(C, \hat{C}) = \frac{\mathrm{MI}(C, \hat{C})}{\max(H(C), H(\hat{C}))}, \tag{31}$$

where $H(C)$ and $H(\hat{C})$ are the entropies of C and $\hat{C}$, respectively. It is easy to check that $\mathrm{NMI}(C, \hat{C})$ ranges from 0 to 1. $\mathrm{NMI} = 1$ if the two sets of clusters are identical, and $\mathrm{NMI} = 0$ if the two sets are independent.

Table 4 reveals the clustering results on these two metrics with 20 % labeled data given for the semi-supervised

Table 4 Clustering performance with 20 % labeled data available

	Unsupervised					Semi-supervised						
	NMF_a	NMF_b	PCSC	CCA	$SULF_1$	$CNMF_a$	$CNMF_b$	$SGNMF_a$	$SGNMF_b$	$SULF_2$	$SULF_3$	SULF
ACC												
NUS-WIDE	13.48	13.17	14.21	11.19	15.45	13.18	13.26	14.59	14.62	14.69	15.54	**16.60**
Corel5K	16.21	18.85	18.58	15.11	19.74	17.20	19.11	17.29	19.58	22.41	25.23	**25.48**
Retuers	39.73	42.80	44.76	37.05	44.08	38.75	40.70	44.96	43.87	55.87	**56.44**	56.25
NMI												
NUS-WIDE	13.70	14.32	16.00	10.51	17.53	14.89	15.01	14.96	16.33	17.32	18.27	**19.22**
Corel5K	25.00	27.67	26.84	20.95	28.42	26.24	28.34	27.04	29.59	30.69	34.02	**34.56**
Retuers	21.06	22.88	32.12	16.28	21.90	19.17	19.24	26.95	24.50	37.64	37.07	**39.17**

Bold values indicate the best performance in the comparison

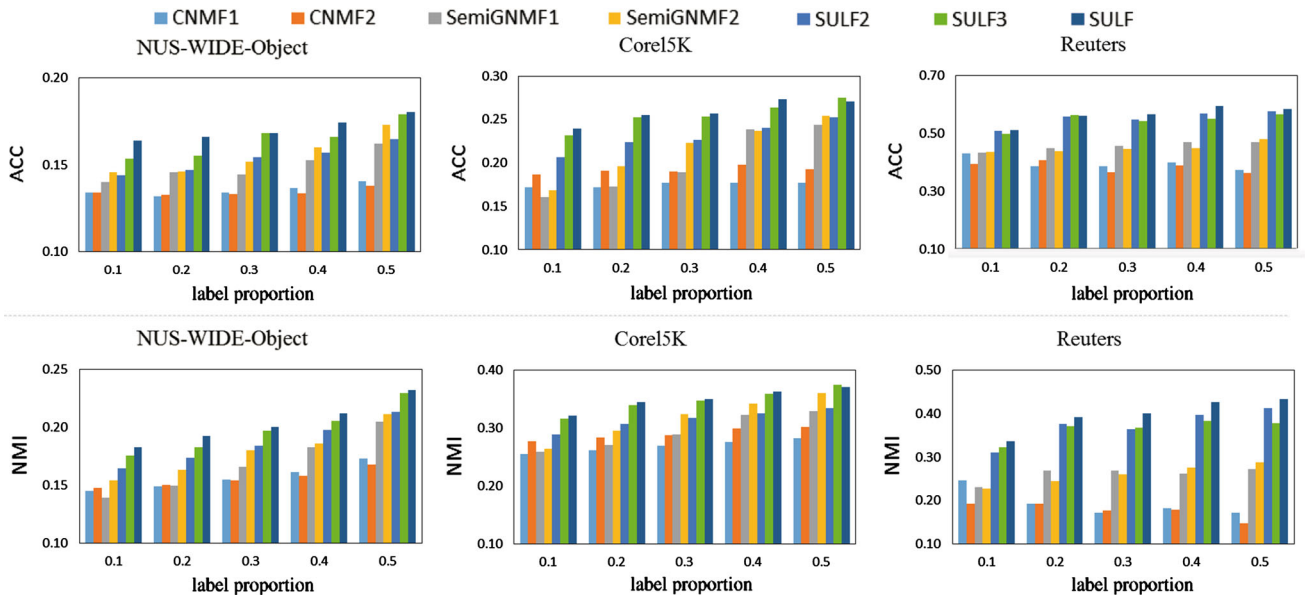


Fig. 2 Clustering performance of semi-supervised algorithms with different label proportion

algorithms. And Fig. 2 illustrates the performance of semi-supervised algorithms with different label proportion. Some interesting points can be observed:

- In the clustering task, SULF displays a greater advantage. On Retuers dataset with 20 % labeled data available, SULF even improves more than 10 % than other methods on both ACC and NMI.
- The performance of SULF is better than that of $SULF_1$. And with the increase of label ratio, the performance of all these semi-supervised methods is improved universally. Both of these points demonstrate the value of partially label information.
- Compared with $SULF_2$, $SULF_3$ and SULF improve the performance obviously, which illustrates complementary information across multiple views is helpful to boost the performance.
- Similar to the performance of classification, the result comparison between SULF and $SULF_3$ again confirms the effectiveness of the mechanism to estimate the weights of different view.
- To those single-view methods, in most conditions, combining features together is an effective way to improve the result.
- The performance of CNMF degenerates too much on Retuers dataset. It may be that the hard constraint of CNMF is too strong for text data.

6.5 Parameter selection

SULF has three essential parameters: λ controls the smoothness of $\mathbf{\Pi}$, β balances the multi-view objective and semi-supervised objective and the γ controls the $l_{2,1}$-norm regularization term.

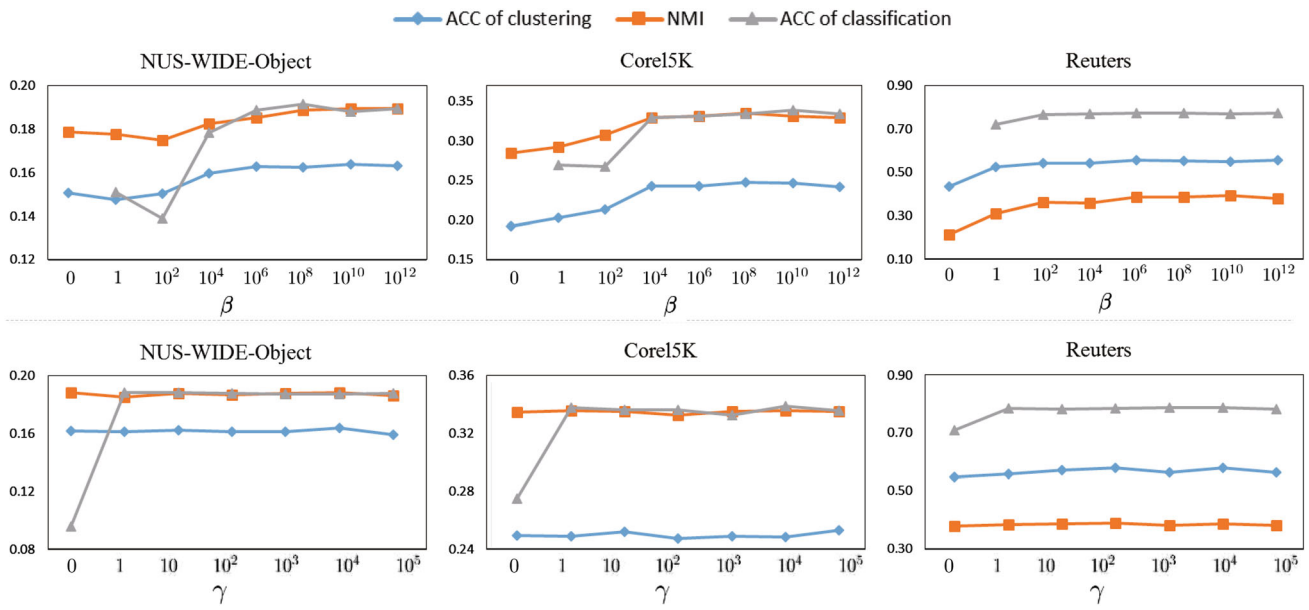


Fig. 3 The performance of SULF versus the parameters β and γ

For convenience, we set λ to 1,000 empirically during our experiments. Figure 3 illustrates how the performance of SULF varies with the parameters β and γ, respectively. The performance of SULF is very stable with respect to β when the value of β is larger than 10,000. In our experiments, we set it to 10^8. The parameter γ has greater influence in the classification task than clustering task, since it affects the learning of linear classification $\mathbf{W}$ directly. In our experiments, we set γ to 10^2.

7 Conclusion

In this paper, we propose a novel algorithm for unified factor learning with partially labeled data, called Semi-supervised Unified Latent Factor learning approach (SULF). SULF assumes that multi-view data matrices share a common unified latent space. In the unified latent space, the prediction loss on the partially labeled data is minimized. What is more, to accommodate noisy or unreliable views, SULF learns the weight of different views automatically. Thus, the obtained parts-based representation can have more discriminating power. An effective multiplicative-based iterative algorithm is developed to solve the proposed optimization problem. The experimental results on three real-world datasets for both classification and clustering tasks have demonstrated the effectiveness of our approach.

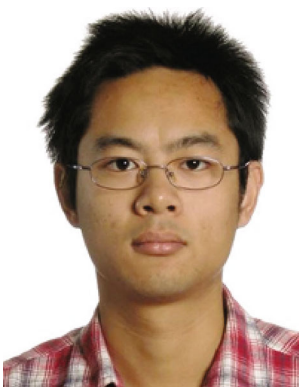
Acknowledgments This work was supported by 973 Program (2012CB316304) and National Natural Science Foundation of China (61272329, 61202325, and 61070104).

References

1. Amini, M.R., Usunier, N., Goutte, C.: Learning from multiple partially observed views—an application to multilingual text categorization. In: Neural Information Processing Systems, pp. 28–36 (2009)
2. Ando, R.K., Zhang, T.: Two-view feature generation model for semi-supervised learning. In: International Conference on Machine Learning, pp. 25–32 (2007)
3. Blum, A., Mitchell, T.M.: Combining labeled and unlabeled data with co-training. In: Computational Learning Theory, pp. 92–100 (1998)
4. Cai, D., He, X., Han, J., Huang, T.S.: Graph regularized nonnegative matrix factorization for data representation. IEEE Trans. Pattern Anal. Mach. Intell. **33**, 1548–1560 (2011)
5. Chaudhuri, K., Kakade, S.M., Livescu, K., Sridharan, K.: Multi-view clustering via canonical correlation analysis. In: International Conference on Machine Learning, pp. 17–136 (2009)
6. Chen, N., Zhu, J., Sun, F., Xing, E.P.: Large-margin predictive latent subspace learning for multiview data analysis. IEEE Trans. Pattern Anal. Mach. Intell. **34**, 2365–2378 (2012)
7. Chen, X., Chen, S., Xue, H., Zhou, X.: A unified dimensionality reduction framework for semi-paired and semi-supervised multi-view data. Pattern Recognit. **45**(5), 2005–2018 (2012)
8. Chen, Y., Rege, M., Dong, M., Hua, J.: Nonnegative matrix factorization for semi-supervised data clustering. Knowl. Inf. Syst. **17**, 355–379 (2008)
9. Chua, T.S., Tang, J., Hong, R., Li, H., Luo, Z., Zheng, Y.: NUS-WIDE: a real-world web image database from National University of Singapore. In: Conference on Image and Video Retrieval, pp. 1–9 (2009)
10. Ding, C.H.Q., Li, T., Jordan, M.I.: Convex and semi-nonnegative matrix factorizations. IEEE Trans. Pattern Anal. Mach. Intell. **32**, 45–55 (2010)
11. Ding, C.H.Q., Zhou, D., He, X., Zha, H.: R1PCA: rotational invariant L1-norm principal component analysis for robust subspace factorization. In: International Conference on Machine Learning, pp. 281–288 (2006)

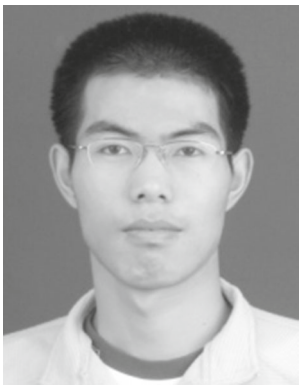
12. Duygulu, P., Barnard, K., Freitas, J.F.G.D., Forsyth, D.A.: Object recognition as machine translation: learning a Lexicon for a fixed image vocabulary. In: European Conference on Computer Vision, pp. 97–112 (2002)
13. Hong, R., Tang, J., Tan, H.K., Ngo, C.W., Yan, S., Chua, T.S.: Beyond search: event-driven summarization for web videos. ACM Trans. Multimed. Comput. Commun. Appl. **7**(4), 35:1–35:18 (2011)
14. Hong, R., Wang, M., Li, G., Nie, L., Zha, Z.J., Chua, T.S.: Multimedia question answering. IEEE Trans. Multimed. **19**(4), 72–78 (2012)
15. Hotelling, H.: Relations between two sets of variates. Biometrika **28**, 321–377 (1936)
16. Kumar, A., Rai, P., Daume, III H.: Co-regularized multi-view spectral clustering. In: Neural Information Processing Systems, pp. 1413–1421 (2011)
17. Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. Nature (1999)
18. Lee, D.D., Seung, H.S.: Algorithms for nonnegative matrix factorization. In: Neural Information Processing Systems, Vol. 13, pp. 556–562 (2000)
19. Li, Z., Liu, J., Lu, H.: Structure preserving non-negative matrix factorization for dimensionality reduction. Comput. Vis. Image Underst. **117**(9), 1175–1189 (2013)
20. Li, Z., Liu, J., Zhu, X., Liu, T., Lu, H.: Image annotation using multi-correlation probabilistic matrix factorization. In: ACM Multimedia, pp. 1187–1190 (2010)
21. Li, Z., Yang, Y., Liu, J., Zhou, X., Lu, H.: Unsupervised feature selection using nonnegative spectral analysis. In: AAAI (2012)
22. Liu, H., Wu, Z., Cai, D., Huang, T.S.: Constrained nonnegative matrix factorization for image representation. IEEE Trans. Pattern Anal. Mach. Intell. **34**, 1299–1311 (2012)
23. Nigam, K., McCallum, A.K., Thrun, S., Mitchell, T.M.: Text classification from labeled and unlabeled documents using EM. Mach. Learn. **39**, 103–134 (2000)
24. van de Sande, K.E.A., Gevers, T., Snoek, C.G.M.: Evaluating color descriptors for object and scene recognition. IEEE Trans. Pattern Anal. Mach. Intell. **32**, 1582–1596 (2010)
25. Sun, T., Chen, S., yu Yang, J., Shi, P.: A novel method of combined feature extraction for recognition. In: IEEE International Conference on Data Mining, pp. 1043–1048 (2008)
26. Wang, M., Hong, R., Li, G., Zha, Z.J., Yan, S., Chua, T.S.: Event driven web video summarization by tag localization and key-shot identification. IEEE Trans. Multimed. **14**(4), 975–985 (2012)

Author Biographies



Yu Jiang received his B.E. degree in Electrical Engineering from Beihang University, Beijing, China, in 2009. He is currently a Ph.D. Candidate in the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences. His research interests include subspace learning, multiview learning, and recommender system, etc.

Jing Liu received her B.E. degree in 2001 and M.E. degree in 2004 from Shandong University, and her Ph.D. degree from Institute of Automation, Chinese Academy of Sciences in 2008. Currently she is an associate professor in National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences. Her research interests include machine learning, image content analysis and classification, multimedia information indexing and retrieval, etc.



Zechao Li received the Ph.D. degree in Pattern Recognition and Intelligent System from National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, in 2013. He received the B.E. degree from University of Science and Technology of China (USTC), Anhui Province, China, in 2008. He is currently an assistant professor in School of Computer Science, Nanjing University of Science and Technology, China. His research interests include machine learning, subspace learning, multimedia understanding, etc.

Hanqing Lu received his B.E. degree in 1982 and his M.E. degree in 1985 from Harbin Institute of Technology, and Ph.D. degree from Huazhong University of Sciences and Technology, Wuhan, China in 1992. Currently he is a professor at National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences. His research interests include computer vision, object tracking, recognition and image retrieval, etc. He has published more than 300 papers in those areas. He is a senior member of the IEEE.