Multiple Social Network Learning and Its Application in Volunteerism Tendency Prediction

Xuemeng Song, Liqiang Nie, Luming Zhang, Mohammad Akbari, Tat-Seng Chua
National University of Singapore
{sxmustc, nieliqiang, zglumg}@gmail.com, akbari@nus.edu.sg, chuats@comp.nus.edu.sg

ABSTRACT

We are living in the era of social networks, where people throughout the world are connected and organized by multiple social networks. The views revealed by different social networks may vary according to the different services they offer. They are complimentary to each other and comprehensively characterize a specific user from different perspectives. As compared to the scare knowledge conveyed by a single source, appropriate aggregation of multiple social networks offers us a better opportunity for deep user understanding. The challenges, however, co-exist with opportunities. The first challenge lies in the existence of block-wise missing data, caused by the fact that some users may be very active in certain social networks while inactive in others. The second challenge is how to collaboratively integrate multiple social networks. Towards this end, we first proposed a novel model for data missing completion by seamlessly exploring the knowledge from multiple sources. We then developed a robust multiple social network learning model, and applied it to the application of volunteerism tendency prediction. Extensive experiments on real world dataset verify the effectiveness of our scheme. The proposed scheme is applicable to many other domains, such as demographic inference and interest prediction.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing

Keywords

Multiple Social Network Learning; Missing Data Completion; Volunteerism Tendency Prediction

1. INTRODUCTION

With the explosion of social network services, more and more people are involved in multiple social networks for various purposes at the same time. It is reported that 52% of online adults concurrently use multiple social media services[1]. Different aspects of users are disclosed on different social networks due to their different emphasis. In fact, these views are complementary to each other and essentially characterize the same user from different perspectives. As compared with single social network, appropriate aggregation of multiple social networks provides us a potential to comprehensively understand the given users [1, 29]. For example, we can learn descriptive user representation, build predictive models for user profiles, and recommend prescriptive actions based on complete historical behaviors. Hence, an effective technique for multiple social network learning (MSNL) is highly desired. Distinguished from multi-view learning which maximizes the agreement between views using unlabeled data, MSNL works towards supervised learning.

However, integration of multiple sources is non-trivial [27]. The first tough challenge lies in how to fuse users' heterogeneous distributed data from multiple social networks effectively. One naive approach is to concatenate the feature spaces generated from different sources into a unified feature space. Thereby, traditional machine learning models can be further applied. However, this method simply treats the confidence of all data sources equally and may also lead to the curse of dimensionality. Moreover, it ignores two important facts: 1) different aspects of users are revealed in different social networks and are thus distributed in different feature spaces; and 2) all these aspects tend to characterize the same users. In particular, data from multi-sources describes the same user and thus the results predicted by different sources should be similar. Therefore, it is expected to take the source confidence and source consistency into consideration. Another challenge we are facing is the data missing problem. Although some users have social accounts on multiple social networks, generally they are active on only a few of them. One simple approach to address this challenge is to discard all incomplete subjects. It is apparent that this method will dramatically reduce the training size, thereby result in overfitting in the model learning stage. Therefore, accurately completing missing data by jointly utilizing multiple sources is a necessity to enhance the learning performance.

To address these problems, we present a scheme for MSNL, which co-regulates the source confidence and source consistency. Figure 1 shows our proposed scheme comprising of three components. Given a set of users, we first crawl

---

[1] According Paw Research Internet Project's Social Media Update 2014: http://www.pewinternet.org/
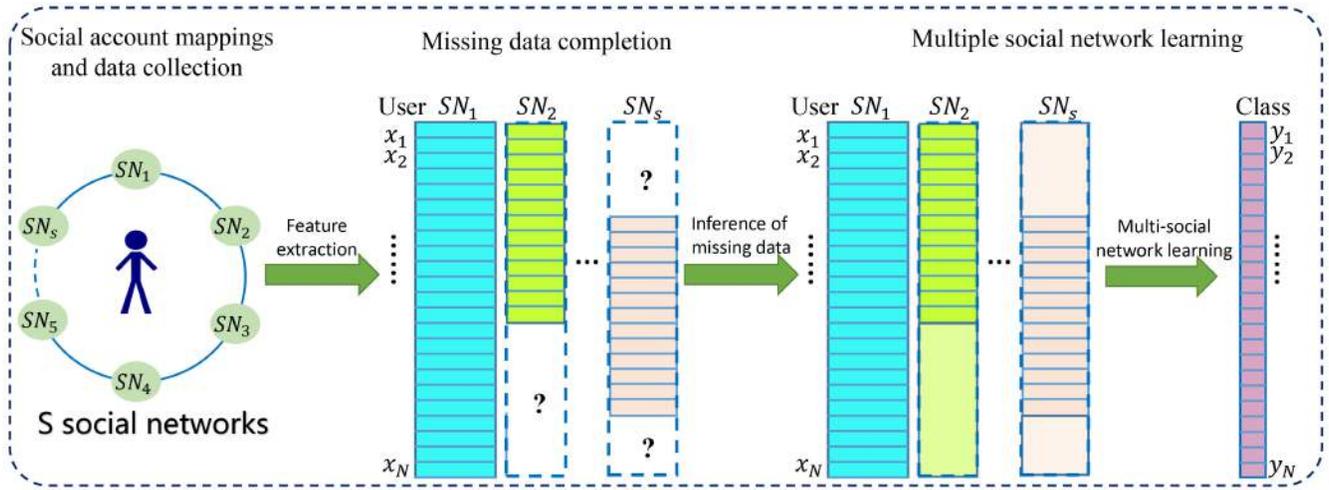
Figure 1: Illustration of our proposed scheme. We first collect and align users' distributed data from multiple social networks. We then jointly infer the block-wise missing data based on the available data. We finally apply MSNL on the complete data. SN_i, x_j, and y_l refer to the i-th social network, j-th user sample, and the l-th corresponding label, respectively.

their historical contents and all social connections. The first component extracts the multi-faceted information cues to describe a given user, including demographic information, practical behaviors, historical posts, and profiles of social connections. To deal with the block-wise missing data, the second component attempts to infer the block-wise missing data by learning a latent space shared by different social networks, achieving a complete input to the next component. We finally use the last component to conduct MSNL on the complete data. Particularly, we model the confidence of different data sources and the consistency among them by unifying two regularization terms into our model.

Based upon our proposed scheme, we introduce one application scenario: volunteerism tendency prediction. Volunteerism was defined in [18] as long-term, planned, prosocial behaviors that occur within organizational settings and can benefit strangers. Persons exhibiting volunteerism are the so-called volunteers, who serve as an important work force in modern society. Traditionally, it is intractable for nonprofit organizations (NPOs) to aimlessly recruit volunteers from the huge crowd. It is thus necessary to develop an automatic volunteerism tendency prediction system to alleviate the dilemma that NPOs are facing. In particular, we take the advantage of users' casually distributed online data, especially from multiple social networks, which can comprehensively reveal users' personal concerns, interests [5, 22] and even personality traits [20, 21]. On the other hand, the real word dataset may contain block-wise missing data due to some users' inactivity in social networks. Moreover, the specific task also brings us another issue in terms of data collection and ground truth construction. Therefore, the proposed scheme naturally fits this application scenario.

Our main contributions can be summarized in threefold:
- We propose a novel MSNL model, which is able to model both the source confidence and source consistency. Specifically, we can obtain a closed-form solution by taking the inverse of a linear system, which has been mathematically proven to be invertible.
- We propose an approach to deal with missing data in multiple social networks, which first learns a common

latent subpace shared by different sources [12] and the original missing data can then be derived in turn.
- We empirically evaluate our proposed scheme on the application of volunteerism tendency prediction. In addition, we develop a set of volunteer-oriented features to characterize users' volunteerism tendency. We have released our compiled dataset[2] to facilitate other researchers to repeat our experiments and verify their proposed approaches.

The remainder of this paper is structured as follows, Section 2 briefly reviews the related work. Section 3 describes the proposed MSNL model. Missing data completion is introduced in Section 4. Section 5 presents the set of volunteer-oriented features we developed. Section 6 details the experimental results and analysis, followed by our concluding remarks in Section 7.

2. RELATED WORK

Although our work is distinguished from multi-view learning, we can still benefit from their efforts. Zhang et al. [28] proposed an inductive multi-view multi-task learning model (regMVMT). regMVMT penalizes the disagreement of models learned from different sources over the unlabeled samples. Besides, the authors also studied the structured missing data, which is completely missing for a source in terms of a task. In other words, if a source is available for a task, then all samples will have data from this source. However, they overlooked the source weights and did not pay attention to the partially structured missing data, which are both what we are concerned with. Yuan et al. [25] introduced an incomplete multi-source feature learning method, avoiding the direct inference of block-wise missing data. Particularly, the authors split the incomplete data into disjoint groups, where they conducted feature learning independently. However, such a mechanism constrains us to conduct source level analysis. Later, Xiang et al. [24] investigated multi-source learning with block-wise missing

---

[2]The compiled dataset is currently publicly accessible via: http://multiplesocialnetworklearning.azurewebsites.net/

2

data with an application of Alzheimer's Disease prediction and proposed the iSFS model. Apart from feature-level analysis, the authors also conducted source-level analysis by introducing the weights of models obtained from different sources. However, ignoring the consistency relationships among different models seems inappropriate. In addition, the authors also adapted the model to handle cases where block-wise missing data exists. Different from their work, we infer the missing data by making full use of the available data before applying MSNL, which is more generalizable to other applications.

3. MULTIPLE SOCIAL NETWORK LEARNING

This section details our proposed MSNL model and derives an analytic solution by solving the inverse of a linear system, whose invertibility is proved rigorously.

3.1 Notation

We first declare some notations. In particular, we use bold capital letters (e.g. $\mathbf{X}$) and bold lowercase letters (e.g. $\mathbf{x}$) to denote matrices and vectors, respectively. We employ non-bold letters (e.g. x) to represent scalars, and Greek letters (e.g. λ) as parameters. If not clarified, all vectors are in column forms.

Suppose we have a set of N labeled data samples and $S \geq 2$ social networks. We compile the S social networks with an index set $\mathcal{C} = \{1, 2, \cdots, S\}$. Let D_s and N_s denote the number of features and samples in the s-th social network, $s \in \mathcal{C}$, respectively. Let $\mathbf{X}_s \in \mathbb{R}^{N \times D_s}$ denote the feature matrix extracted from the s-th social network. Each row represents a user sample. Then the dimension of features extracted from all these social networks is $D = \sum_{s=1}^{S} D_s$. The whole feature matrix can be written as $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_S\} \in \mathbb{R}^{N \times D}$ and $\mathbf{y} = \{y_1, y_2, \cdots, y_N\}^T \in \{1, -1\}^{N \times 1}$ is the corresponding label vector.

3.2 Problem Formulations

Based on a set of data samples with S social networks, we can learn S predictive models, where each model is individually and independently trained on a social network. The final predictive model can be strengthened via linear combination of these S models. Mathematically, we learn one linear mapping function $\mathbf{f}_s$ for the s-th social network. In addition, we assume that the mapping functions learned from all social networks agree with one another as much as possible. Particularly, we can formalize this assumption using regularization function. Using the least square loss function, we have the following objective function,

$$\min_{\mathbf{f}_s} \frac{1}{2N} \left\| \mathbf{y} - \mathbf{f}(\mathbf{X}) \right\|^2 + \frac{\mu}{2N} \sum_{s=1}^{S} \sum_{s' \neq s} \left\| \mathbf{f}_s(\mathbf{X}_s) - \mathbf{f}_{s'}(\mathbf{X}_{s'}) \right\|^2$$
$$+ \frac{\lambda}{2} \|\mathbf{f}\|^2, \quad (1)$$

where $\mathbf{f}(\mathbf{X})$ is the final predictive model. $\mathbf{f}_s(\mathbf{X}_s)$ is the prediction results generated from data $\mathbf{X}_s$. λ and μ are the nonnegative regularization parameters that regulate the sparsity of the solution regarding $\mathbf{f}_s$ and the disagreement among models learned from different social networks, respectively. If we just treat the confidence of different social networks equally, the final predictive model can be formalized as follows,

$$\mathbf{f}(\mathbf{X}) = \frac{1}{S} \sum_{s=1}^{S} \mathbf{f}_s(\mathbf{X}_s). \quad (2)$$

However, in reality, different social networks always have different confidence to the final prediction, and we consider modeling the weights of multiple sources instead of treating all sources equally by introducing the weight vector: $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \cdots, \alpha_S]^T \in \mathbb{R}^{S \times 1}$, where α_s controls the weight of model learned from s-th social network. Then the final model is defined as follows,

$$\mathbf{f}(\mathbf{X}) = \sum_{s=1}^{S} \alpha_s \mathbf{f}_s(\mathbf{X}_s)$$
$$\text{subject to} \quad \mathbf{e}^T \boldsymbol{\alpha} = 1, \quad (3)$$

where $\mathbf{e} = [1, 1, \cdots, 1]^T \in \mathbb{R}^{S \times 1}$. It is worth mentioning that we do not impose the constraint of $\alpha_s \geq 0$, as we want to keep both positive and negative weights. Positive weights indicate the positive correlations of social networks with the final results, while negative weights reflect negative correlations between the given task and different sources, which may contain unreliable and noisy data.

For the s-th social network, we learn a linear mapping function indexed by a model $\mathbf{w}_s \in \mathbb{R}^{D_s \times 1}$. Then the objective function can be rewritten as follows,

$$\min_{\mathbf{w}_s, \boldsymbol{\alpha}} \frac{1}{2N} \left\| \mathbf{y} - \sum_{s=1}^{S} \alpha_s \mathbf{X}_s \mathbf{w}_s \right\|^2 + \frac{\mu}{2N} \sum_{s=1}^{S} \sum_{s' \neq s} \left\| \mathbf{X}_s \mathbf{w}_s - \mathbf{X}_{s'} \mathbf{w}_{s'} \right\|^2$$
$$+ \frac{\lambda}{2} \sum_{s=1}^{S} \left\| \mathbf{w}_s \right\|^2 + \frac{\beta}{2} \left\| \boldsymbol{\alpha} \right\|^2, \quad (4)$$

where $\mathbf{e}^T \boldsymbol{\alpha} = 1$ and β is the regularization parameter, controlling the sparsity of the solution regarding $\boldsymbol{\alpha}$.

3.3 Optimization

We adopt the alternating optimization strategy to solve the two variables $\boldsymbol{\alpha}$ and $\mathbf{w}_s$ in Eqn. (4). In particular, we optimize one variable while fixing the other one in each iteration. We keep this iterative procedure until the objective function converges.

3.3.1 Computing $\boldsymbol{\alpha}$ with $\mathbf{w}_s$ fixed

We denote the objective function as $\boldsymbol{\Gamma}$. For simplicity, we replace $\mathbf{y}$ in Eqn. (4) by $\mathbf{y}\mathbf{e}^T \boldsymbol{\alpha}$, as $\mathbf{e}^T \boldsymbol{\alpha} = 1$. With the help of Lagrangian, $\boldsymbol{\Gamma}$ can be rewritten as follows,

$$\min_{\boldsymbol{\alpha}} \frac{1}{2N} \left\| \mathbf{y}\mathbf{e}^T \boldsymbol{\alpha} - \mathbf{X}\mathbf{W}\boldsymbol{\alpha} \right\|^2 + \frac{\beta}{2} \left\| \boldsymbol{\alpha} \right\|^2 + \delta(1 - \mathbf{e}^T \boldsymbol{\alpha}), \quad (5)$$

where δ is the nonnegative Lagrange multiplier and $\mathbf{W} = diag(\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_S) \in \mathbb{R}^{D \times S}$. Taking derivative of $\boldsymbol{\Gamma}$ with respect to $\boldsymbol{\alpha}$, we have,

$$\frac{\partial \boldsymbol{\Gamma}}{\partial \boldsymbol{\alpha}} = \frac{1}{N} (\mathbf{y}\mathbf{e}^T - \mathbf{X}\mathbf{W})^T (\mathbf{y}\mathbf{e}^T - \mathbf{X}\mathbf{W})\boldsymbol{\alpha} + \beta\boldsymbol{\alpha} - \delta\mathbf{e}. \quad (6)$$

Setting Eqn. (6) to zero, it can be derived that,

$$\boldsymbol{\alpha} = \delta \mathbf{M}^{-1} \mathbf{e}, \quad (7)$$

where

$$\mathbf{M} = \frac{1}{N} (\mathbf{y}\mathbf{e}^T - \mathbf{X}\mathbf{W})^T (\mathbf{y}\mathbf{e}^T - \mathbf{X}\mathbf{W}) + \beta\mathbf{I}. \quad (8)$$

Since $\mathbf{e}^T\boldsymbol{\alpha} = 1$, we can obtain that,

$$\delta = \frac{1}{\mathbf{e}^T\mathbf{M}^{-1}\mathbf{e}}, \qquad \boldsymbol{\alpha} = \frac{\mathbf{M}^{-1}\mathbf{e}}{\mathbf{e}^T\mathbf{M}^{-1}\mathbf{e}}. \qquad (9)$$

Obviously, $\mathbf{M} \in \mathbb{R}^{S \times S}$ is positive definite and invertible, according to the definition. We thus can obtain the analytic solution of $\boldsymbol{\alpha}$ as Eqn. (9). Moreover, we note that when the prediction results learned from all social networks are equal, where $\mathbf{X}_1\mathbf{w}_1 = \mathbf{X}_2\mathbf{w}_2 = \cdots = \mathbf{X}_S\mathbf{w}_S$, then same weights will be assigned, i.e., $\alpha_1 = \alpha_2 = \cdots = \alpha_S$. In addition, Eqn. (9) tends to assign higher weight α_s, if smaller difference exists between $\mathbf{y}$ and $\mathbf{X}_s\mathbf{w}_s$.

3.3.2 Computing $\mathbf{w}_s$ with $\boldsymbol{\alpha}$ fixed

When $\boldsymbol{\alpha}$ is fixed, we compute the derivative of $\boldsymbol{\Gamma}$ regarding $\mathbf{w}_s$ as follows,

$$\frac{\partial \boldsymbol{\Gamma}}{\partial \mathbf{w}_s} = \frac{1}{N}\alpha_s\mathbf{X}_s^T(\sum_{s=1}^{S}\alpha_s\mathbf{X}_s\mathbf{w}_s - \mathbf{y})$$
$$+ \frac{\mu}{N}\mathbf{X}_s^T\sum_{s=1}^{S}\sum_{s'\neq s}(\mathbf{X}_s\mathbf{w}_s - \mathbf{X}_{s'}\mathbf{w}_{s'}) + \lambda\mathbf{w}_s$$
$$= \Big[\lambda\mathbf{I} + \frac{\alpha_s^2}{N}\mathbf{X}_s^T\mathbf{X}_s + \frac{\mu(S-1)}{N}\mathbf{X}_s^T\mathbf{X}_s\Big]\mathbf{w}_s$$
$$+ \sum_{s=1}^{S}\sum_{s'\neq s}\frac{1}{N}(\alpha_s\alpha_{s'} - \mu)\mathbf{X}_s^T\mathbf{X}_{s'}\mathbf{w}_{s'} - \frac{\alpha_s}{N}\mathbf{X}_s^T\mathbf{y}, \quad (10)$$

where $\mathbf{I}$ is a $D_s \times D_s$ identity matrix. Setting Eqn. (10) to zero and rearranging the terms, all $\mathbf{w}_s$'s can be learned jointly by the following linear system,

$$\mathbf{L}\mathbf{w} = \mathbf{t}$$
$$\begin{bmatrix} \mathbf{L}_{11} & \mathbf{L}_{12} & \mathbf{L}_{13} & \cdots & \mathbf{L}_{1S} \\ \mathbf{L}_{21} & \mathbf{L}_{22} & \mathbf{L}_{23} & \cdots & \mathbf{L}_{2S} \\ \mathbf{L}_{31} & \mathbf{L}_{32} & \mathbf{L}_{33} & \cdots & \mathbf{L}_{3S} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{L}_{S1} & \mathbf{L}_{S2} & \mathbf{L}_{S3} & \cdots & \mathbf{L}_{SS} \end{bmatrix} \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \mathbf{w}_3 \\ \vdots \\ \mathbf{w}_S \end{bmatrix} = \begin{bmatrix} \mathbf{t}_1 \\ \mathbf{t}_2 \\ \mathbf{t}_3 \\ \vdots \\ \mathbf{t}_S \end{bmatrix}, \quad (11)$$

where $\mathbf{L} \in \mathbb{R}^{D \times D}$ is a sparse block matrix with $S \times S$ blocks, $\mathbf{w} = [\mathbf{w}_1^T, \mathbf{w}_2^T, \cdots, \mathbf{w}_S^T]^T \in \mathbb{R}^{D \times 1}$ and $\mathbf{t} = [\mathbf{t}_1^T, \mathbf{t}_2^T, \cdots, \mathbf{t}_S^T]^T \in \mathbb{R}^{D \times 1}$ are both sparse block vectors with $S \times 1$ blocks. $\mathbf{t}_s$, $\mathbf{L}_{ss}$ and $\mathbf{L}_{ss'}$ are defined as follows,

$$\begin{cases} \mathbf{t}_s &= \frac{\alpha_s}{N}\mathbf{X}_s^T\mathbf{y}, \\ \mathbf{L}_{ss} &= \lambda\mathbf{I} + \frac{\alpha_s^2 - \mu}{N}\mathbf{X}_s^T\mathbf{X}_s + \frac{\mu S}{N}\mathbf{X}_s^T\mathbf{X}_s, \\ \mathbf{L}_{ss'} &= \frac{\alpha_s\alpha_{s'} - \mu}{N}\mathbf{X}_s^T\mathbf{X}_{s'}. \end{cases} \quad (12)$$

Technically, $\mathbf{t}$ can be treated as a constant matrix as $\boldsymbol{\alpha}$ is fixed. It is worth noting that $\mathbf{L}$ is symmetric as $\mathbf{L}_{ss'} = \mathbf{L}_{s's}^T$. If we can prove that $\mathbf{L}$ is invertible, then we can derive the closed-form solution of $\mathbf{w}$ as follows,

$$\mathbf{w} = \mathbf{L}^{-1}\mathbf{t}. \qquad (13)$$

We now show $\mathbf{L}$ is invertible by proving that $\mathbf{L}$ is a positive-definite matrix. Let $\mathbf{h} = [\mathbf{h}_1^T, \mathbf{h}_2^T, \cdots, \mathbf{h}_S^T]^T \in \mathbb{R}^{D \times 1} \neq \mathbf{0}$ be an arbitrary block vector, where $\mathbf{h}_i \in \mathbb{R}^{D_i \times 1}, i \in \mathcal{C}$. Then

Algorithm 1 Alternative optimization for solving Eqn. (4)

Input: $\mathbf{X}$, $\mathbf{y}$, λ, β, μ
Output: $\boldsymbol{\alpha}$, $\mathbf{w}$
1: Initialize $(\mathbf{w})^0$ by fitting each source individually on the available data. Initialize $(\boldsymbol{\alpha})^0 = [\frac{1}{S}, \frac{1}{S}, \cdots, \frac{1}{S}]$.
2: **for** $k = 1, 2, \cdots$ **do**
3:     Compute each $(\boldsymbol{\alpha})^k$ according to Eqn. (9).
4:     Update $(\mathbf{w})^k$ according to Eqn. (13).
5:     **if** the objective value stops decreasing **then**
6:         return $\boldsymbol{\alpha} = (\boldsymbol{\alpha})^k$ and $\mathbf{w} = (\mathbf{w})^k$
7:     **end if**
8: **end for**

we need to prove that $\mathbf{h}^T\mathbf{L}\mathbf{h}$

$$= \sum_{i=1}^{S}\sum_{j=1}^{S}\mathbf{h}_i^T\mathbf{L}_{ij}\mathbf{h}_j$$
$$= \lambda\|\mathbf{h}\|^2 + \frac{1}{N}\Big[\sum_{i=1}^{S}\|\alpha_i\mathbf{X}_i\mathbf{h}_i\|^2 + \mu(S-1)\sum_{i=1}^{S}\|\mathbf{X}_i\mathbf{h}_i\|^2$$
$$+ \sum_{i=1}^{S}\sum_{j\neq i}\alpha_i\mathbf{h}_i^T\mathbf{X}_i^T\alpha_j\mathbf{X}_j\mathbf{h}_j - \mu\sum_{i=1}^{S}\sum_{j\neq i}\mathbf{h}_i^T\mathbf{X}_i^T\mathbf{X}_j\mathbf{h}_j\Big], \quad (14)$$

is always larger than zero. In fact, given an arbitrary vector $\mathbf{b}_i$, we have,

$$\|\mathbf{b}_1 - \mathbf{b}_2\|^2 + \cdots + \|\mathbf{b}_{(S-1)} - \mathbf{b}_S\|^2 + \|\mathbf{b}_S - \mathbf{b}_1\|^2 \geq 0$$
$$\sum_{i=1}^{S}\|\mathbf{b}_i\|^2 \geq \sum_{i=1}^{S}\sum_{j\neq i}\mathbf{b}_i^T\mathbf{b}_j. \qquad (15)$$

Therefore, as $S \geq 2$, we have the following inequality,

$$\mu(S-1)\sum_{i=1}^{S}\|\mathbf{X}_i\mathbf{h}_i\|^2 \geq \mu\sum_{i=1}^{S}\|\mathbf{X}_i\mathbf{h}_i\|^2$$
$$\geq \mu\sum_{i=1}^{S}\sum_{j\neq i}(\mathbf{X}_i\mathbf{h}_i)^T\mathbf{X}_j\mathbf{h}_j. \qquad (16)$$

Besides, we know that,

$$\sum_{i=1}^{S}\|\alpha_i\mathbf{X}_i\mathbf{h}_i\|^2 + \sum_{i=1}^{S}\sum_{j\neq i}\alpha_i\mathbf{h}_i^T\mathbf{X}_i^T\alpha_j\mathbf{X}_j\mathbf{h}_j$$
$$= \frac{1}{2}\sum_{i=1}^{S}\|\alpha_i\mathbf{X}_i\mathbf{h}_i\|^2 + \frac{1}{2}\Big\|\sum_{i=1}^{S}\alpha_i\mathbf{X}_i\mathbf{h}_i\Big\|^2 \geq 0. \qquad (17)$$

Based upon Eqn. (16) and Eqn. (17), we have that,

$$\mathbf{h}^T\mathbf{L}\mathbf{h} \geq \lambda\|\mathbf{h}\|^2. \qquad (18)$$

As $\mathbf{h} \neq \mathbf{0}$, $\mathbf{h}^T\mathbf{L}\mathbf{h}$ is always larger than zero. Consequently, $\mathbf{L}$ is invertible. The overall procedures for alternating optimization are summarized in Algorithm 1. As each iteration can decrease $\boldsymbol{\Gamma}$, whose lower bound is zero, we can guarantee the convergence of Algorithm 1 [7, 16].

4. MISSING DATA COMPLETION

In this section, we deal with a more challenging and realistic situation, where block-wise missing data exists, and propose an approach for multiple social network data completion (MSNDC). In such situations, user samples may

not be active in all social networks, which leads to the block-wise data missing.

Suppose we have S data sources in total and each sample has at least one data source available. We employ the subset $\mathcal{C}_i \subseteq \mathcal{C}$ to indicate the presence of each source and the signature of a specific social network combination. Based on these combinations, all the data samples can be split into multiple exclusive sets, where each set corresponds to a combination. Figure 2 illustrates the incomplete data in our dataset. As can be seen, all users have complete features from SN_1, while some users miss data in SN_2 or SN_3. Therefore, our dataset can be split by four exclusive social network combinations: $\mathcal{C}_1 = \{1, 2\}$, $\mathcal{C}_2 = \{1, 2, 3\}$, $\mathcal{C}_3 = \{1, 3\}$, $\mathcal{C}_4 = \{1\}$.

Inspired by [10], we use Non-negative Matrix Factorization (NMF) to explore the latent spaces that are shared by different social networks, and further infer the missing data based upon these latent spaces. It is reasonable to assume that the data from different social networks about the same user shares certain latent features. We employ $\mathbf{X}_s^{\mathcal{C}_i} \in \mathbb{R}^{N_{\mathcal{C}_i} \times D_s}$ to denote the samples generated from the s-th social network. It only contains samples that are available in the set of social networks $\mathcal{C}_i$, where $N_{\mathcal{C}_i}$ stands for the number of these samples. We use $\mathbf{U}_s \in \mathbb{R}^{z \times D_s}$ to represent the latent basis matrix for the s-th social network, and $\mathbf{P}_s^{\mathcal{C}_i} \in \mathbb{R}^{N_{\mathcal{C}_i} \times z}$ to denote the corresponding latent representation of feature matrix $\mathbf{X}_s^{\mathcal{C}_i}$. z is the dimension of the shared latent space of different social networks. The intuitive assumption is that for the samples available in both the s-th and s'-th social networks, their corresponding latent representations should also be quite similar. In particular, we impose this constraint to NMF as follows,

$$\mathbf{P}_s^{\mathcal{C}_i} = \mathbf{P}_{s'}^{\mathcal{C}_i} = \mathbf{P}^{\mathcal{C}_i}, \tag{19}$$

where $s \neq s'$, $s \in \mathcal{C}_i$, and $s' \in \mathcal{C}_i$. We thus learn the shared subspaces by the following objective function,

$$
\begin{aligned}
\min_{\substack{\mathbf{U}_s \geq \mathbf{0} \\ \mathbf{P}_s \geq \mathbf{0}}} & \left\| \begin{bmatrix} \mathbf{X}_1^{\{1\}} \\ \mathbf{X}_1^{\{1,2\}} \\ \mathbf{X}_1^{\{1,3\}} \\ \mathbf{X}_1^{\{1,2,3\}} \end{bmatrix} - \begin{bmatrix} \mathbf{P}^{\{1\}} \\ \mathbf{P}^{\{1,2\}} \\ \mathbf{P}^{\{1,3\}} \\ \mathbf{P}^{\{1,2,3\}} \end{bmatrix} \mathbf{U}_1 \right\|^2 + \nu \left\| \mathbf{P}_1 \right\|_1 + \eta \left\| \mathbf{U}_1 \right\|_1 \\
& + \left\| \begin{bmatrix} \mathbf{X}_2^{\{1,2\}} \\ \mathbf{X}_2^{\{1,2,3\}} \end{bmatrix} - \begin{bmatrix} \mathbf{P}^{\{1,2\}} \\ \mathbf{P}^{\{1,2,3\}} \end{bmatrix} \mathbf{U}_2 \right\|^2 + \nu \left\| \mathbf{P}_2 \right\|_1 + \eta \left\| \mathbf{U}_2 \right\|_1 \\
& + \left\| \begin{bmatrix} \mathbf{X}_3^{\{1,3\}} \\ \mathbf{X}_3^{\{1,2,3\}} \end{bmatrix} - \begin{bmatrix} \mathbf{P}^{\{1,3\}} \\ \mathbf{P}^{\{1,2,3\}} \end{bmatrix} \mathbf{U}_3 \right\|^2 + \nu \left\| \mathbf{P}_3 \right\|_1 + \eta \left\| \mathbf{U}_3 \right\|_1,
\end{aligned}
\tag{20}
$$

where ν and η are the nonnegative tradeoff parameters for the regularizations. Similarly, we employ the alternating optimization strategy to solve the optimization in Eqn. (20). To be more specific, we first initialize $\mathbf{U}_s$ and compute the optimal $\mathbf{P}_s$. Afterwards, $\mathbf{P}_s$ is updated based on the computed $\mathbf{U}_s$. We keep this iterative procedure until the objective function converges.

The proposed approach differs from [10] in the following three aspects. First, MSNDC is generalized to handle the more challenging scenario where data samples are extracted from more than two social networks. Second, apart from regulating the latent representation matrix, we also incorporate the regularization on the latent basis matrix. Third, we further derive the original missing data from
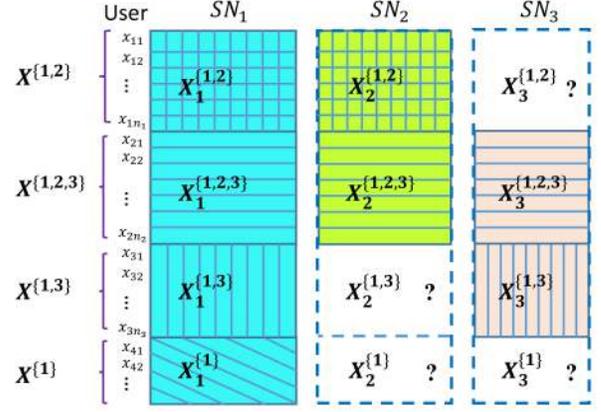
Figure 2: Illustration of the incomplete data from three sources. $\mathbf{X}_s^{\mathcal{C}_i}$ denotes the samples generated from social network s that are only available in the social network combination of $\mathcal{C}_i$.

the latent representation, where the authors in [10] just apply cluster algorithms directly to the latent representation of data instead of the original data. This is due to two considerations. One is that we believe the value of original known data is higher than the latent representation. The other one is that we need to preserve the heterogeneity among data from different sources to fit the MSNL model.

4.1 Optimization

In order to increase the efficiency of the iterative procedure, we initialize $\mathbf{U}_s$ by optimizing the following objective function,

$$
\begin{aligned}
\min_{\mathbf{U}_s \geq \mathbf{0}} & \left\| \mathbf{X}_1^{\{1,2,3\}} - \mathbf{P}^{\{1,2,3\}} \mathbf{U}_1 \right\|^2 + \nu \left\| \mathbf{P}^{\{1,2,3\}} \right\|_1 + \eta \left\| \mathbf{U}_1 \right\|_1 \\
& + \left\| \mathbf{X}_2^{\{1,2,3\}} - \mathbf{P}^{\{1,2,3\}} \mathbf{U}_2 \right\|^2 + \eta \left\| \mathbf{U}_2 \right\|_1 \\
& + \left\| \mathbf{X}_3^{\{1,2,3\}} - \mathbf{P}^{\{1,2,3\}} \mathbf{U}_3 \right\|^2 + \eta \left\| \mathbf{U}_3 \right\|_1.
\end{aligned}
\tag{21}
$$

We then alternatively optimize $\mathbf{U}_s$ and $\mathbf{P}_s$ until the objective function converges. Specifically, we employ the greedy coordinate descent (GCD) approach [9], which has been proven to be tremendously fast to solve NMF decomposition with L1-norm regularization. Finally, we obtain $\mathbf{P}_s, \mathbf{U}_s, s \in \mathcal{C}$, based on which we can infer the missing data as follows,

$$\hat{\mathbf{X}}_s^{\mathcal{C}_i} = \mathbf{P}^{\mathcal{C}_i} \mathbf{U}_s, \quad \forall s \notin \mathcal{C}_i. \tag{22}$$

Algorithm 2 summarizes the overall procedures for alternating optimization.

5. APPLICATION: VOLUNTEERISM TENDENCY PREDICTION

In this work, we apply the proposed scheme to an application scenario: volunteerism tendency prediction. In modern society, volunteers are extremely crucial to NPOs to sustain their continuing operations. The discovery of users' volunteerism tendency can significantly facilitate the recruitment of volunteers for NPOs, which can save considerable cost to find the potential volunteers. In particular, we cast the problem of volunteerism tendency prediction as a user binary classification. If the predicted tendency score of a given user is larger than a pre-defined

Algorithm 2 Alternative optimization for solving Eqn. (20)

Input: $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \nu, \eta$
Output: $\hat{\mathbf{X}}$

1: Initialize $\mathbf{U}_s^{(0)}$ according to Eqn. (21).
2: **for** $k = 1, 2, \cdots$ **do**
3:   **for** $s = 1, 2, \cdots, S$ **do**
4:     Compute each $\mathbf{P}_s^{(k)}$ according to Eqn. (20) via GCD approach.
5:     Update $\mathbf{U}_s^{(k)}$ according to Eqn. (20) via GCD approach.
6:     **if** the objective value stops decreasing **then**
7:       return $\mathbf{U}_s = \mathbf{U}_s^{(k)}$ and $\mathbf{P}_s = \mathbf{P}_s^{(k)}$
8:     **end if**
9:   **end for**
10: **end for**
11: **for** $j = 1, 2, \cdots, S$ **do**
12:   **for** $\mathcal{C}_q \subseteq \mathcal{C}$ **do**
13:     **if** $j \in \mathcal{C}_q$ **then**
14:       $\hat{\mathbf{X}}_j^{\mathcal{C}_q} = \mathbf{X}_j^{\mathcal{C}_q}$.
15:     **else**
16:       Infer $\hat{\mathbf{X}}_j^{\mathcal{C}_q}$ according to Eqn. (22).
17:     **end if**
18:   **end for**
19: **end for**

---

threshold γ, we regard this user as a volunteer. In this work, we explore three popular social networks: Twitter, Facebook and LinkedIn, as they are representative of a public, private, and professional social network, respectively. Besides, it is known that users exhibit different aspects on different social networks [1], and the combination of these three social networks would help to better characterize user behaviors on social platforms.

5.1   Multiple Social Accounts Mappings

To represent the same users with multiple sources, we need to first tackle the problem of "Social Account Mapping", which aims to align the same users across different social networks by linking their multiple social accounts [1]. To accurately establish this mapping, we employ the emerging social services such as About.me[3] and Quora[4], where they encourage users to explicitly list their multiple social accounts on one profile.

We proposed two strategies to collect data from About.me.

- **Keyword search**: We searched About.me with the keyword "volunteer" and obtained $4,151$ volunteer candidates.
- **Random select**: We employed Random API[5], provided by About.me, to collect non-volunteers. This API returns a specified number of random user profiles. Finally, we harvested $1,867$ non-volunteer candidates. It is worth mentioning that volunteers may be present in these random users.

To enlarge our dataset, we also collected candidates from Quora by the breadth-first-search method. Particularly, we took advantage of both the follower and followee[6] relations

---

[3]https://about.me/
[4]http://quora.com/
[5]http://about.me/developer/api/docs/
[6]If A follows B, then A is B's follower and B is A's followee.

---

provided by Quora. Initially, we selected two popular users as the seed users and then explored all their neighboring connected users. We applied similar exploration approach to all other non-seed users. In the end, we collected $172,235$ users' profiles and only retained those who have accounts in Facebook, Twitter and LinkedIn.

5.2   Ground Truth Construction

Based on these candidates, we launched a crawler to collect their historical social contents, including their basic profiles, social posts and relations. However, the traditional web-based crawler is not applicable to Facebook due to its dynamic loading mechanism. We thus resorted to the Selenium[7] to simulate users' click and scroll operations on a FireFox browser and load users' publicly available information. We limited the access rate to one request per second to avoid being blocked by the robot checkers. It is worth mentioning that the data we collected is all publicly available. On the other hand, due to the privacy constraint, we could not access uses' social relations in Facebook and LinkedIn. We hence only collected users' followee relations in Twitter.

In order to improve the quality of our dataset, we employed three annotators to finalize our ground truth. As users tend to provide more complete and reliable profiles in LinkedIn, we guided the annotators to study the LinkedIn profiles of candidate users, and determine whether they are "volunteers" by majority votes. To ensure a uniformly labeling procedure, we provided them a piece of guideline. Given a user's LinkedIn profile, we classified the user as a volunteer if and only if this user lists his/her volunteer experiences in the section "Volunteer experience & Causes" or section "Experience". Candidates who do not satisfy the above two criteria were tagged as non-volunteers. We focused on LinkedIn to determine whether users are volunteers because the volunteer experiences in LinkedIn are the most straightforward evidence to identify volunteers. It should be noted that those who do not mention their volunteer experiences in LinkedIn are not necessarily classified as "non-volunteers". However, the absence of these mentions, at least, reveals their limited interests and low enthusiasm in volunteerism. Therefore, in our work, we broadly defined users as "non-volunteers" if they do not mention their relevant volunteerism experiences in LinkedIn.

Table 1 lists the statistics of our dataset. We obtained the data for $1,425$ volunteers and $4,011$ non-volunteers according to the aforementioned strategies. The crawling was conducted between 22nd August to 11th September, 2013. Here we only selected a subset of non-volunteer data and made the dataset balanced to avoid the training bias. To facilitate this line of research, this dataset has been released after certain privacy preservation processing.

However, in reality, not all users are active enough on all social networks. To ensure the data quality, we treated those inactive users as missing with respect to a specific social network. Therefore, there exists block-wise missing data in our dataset. In particular, we treated a user as missing in Twitter or Facebook, if this user has less than 10 historical social posts. In addition, due to the absence of social post

---

[7]http://docs.seleniumhq.org/download/

Table 1: Statistics of our dataset.

Data	Volunteer	Non-volunteer
Twitter profiles	∼1.5k	∼4k
Twitter posts	∼559k	∼1m
Twitter followees' profiles	∼902k	∼3m
Facebook profiles	∼1.5k	∼4k
Facebook posts	∼83k	∼338k
LinkedIn profiles	∼1.5k	∼4k

mechanism in LinkedIn, we treated a user as missing[8] in LinkedIn if the word count of this user's profile is less than 50. Figure 3 shows the statistics of our incomplete data. As can be seen, about 50% of users have complete data from all three social networks. 1% and 47% of users only miss the data either from Facebook and LinkedIn, while 2% of users miss the data from both of them.
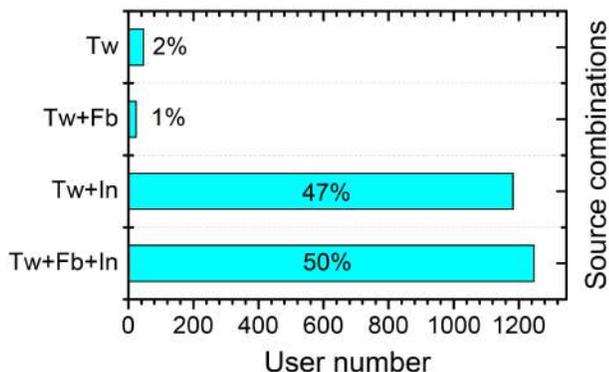
Figure 3: Statistics of the incomplete data. Tw: Users with Twitter data only; Tw+Fb: Users with Twitter and Facebook data only; Tw+In: Users with Twitter and Linkedin data only; Tw+Fb+In: Users without missing data.

5.3 Features

To capture users' volunteerism tendency, we extracted a rich set of volunteer-oriented features.

5.3.1 Demographic Characteristics

The study in [19] reported that some demographic characteristics, such as education and income level, are strong indicators for volunteerism. This study inspires us to extract demographic characteristics from users' profiles, especially the Facebook and LinkedIn profiles. In our work, we explored users' demographic characteristics, including *Gender*, *Relationship status*, *Education level*, and *Number of social connections*.

5.3.2 Linguistic Features

We also extracted linguistic features, including Linguistic Inquiry and Word Count (LIWC) features, user topics and contextual topics.

LIWC features. LIWC is widely-used to analyze the psycho-linguistic transparent lexicon. It plays an important role in predicting users' personality [2, 15]. The main component of LIWC is a directory which contains the mapping from words to 72 categories[9]. Given a document, LIWC computes the percentage of words in each category and represents it as a vector of 72 dimensions.

To capture the key aspects of LIWC features, we selected the top 5 dimensions as the representative LIWC features according to the information gain ratio. Considering that the emotions for individuals may also affect users' volunteerism tendency, we additionally selected two categories from LIWC: positive emotion and negative emotion. Besides, we also utilized the positive-negative emotion ratio to further reflect users' emotional states. Let $L(\bullet)$ represent the percentage of users' words in certain LIWC category. The positive-negative emotion ratio is defined as,

$$PN_{emo} = L(pos)log\frac{L(pos) + \xi_p}{L(neg) + \xi_n}, \qquad (23)$$

where ξ_p and ξ_n are introduced to avoid the situation: individuals have no positive or negative emotional word. They are both set as 0.0001. In total, we have 16 dimension LIWC features, extracted from Twitter and Facebook.

User topics. According to our observation, volunteers may have, on average, a higher probability of talking about topics such as social caring or giving back, while the non-volunteers may mention other topics more often. This motivates us to explore the topic distributions of users' social posts to identify volunteers. We generated topic distributions using Latent Dirichlet Allocation (LDA) [4], which has been widely found to be useful in latent topic modeling [8, 23]. Based on perplexity [14] metric frequently utilized to find the optimal number of hidden topics, we ultimately obtained 52, 26, 42 dimensional topic-level features over users' Twitter, Facebook and LinkedIn data, respectively.

Contextual topics. We define users' contextual topics as the topics of users' connections. We believe that the contextual topics intuitively reflect the contexts of users. "He that lies down with dogs must rise up with fleas" tells us that the context significantly affects a user's tendency. Particularly, we studied followees and retweeting[10] connections on Twitter because of their intuitive reflection of topics that users concern. As the bio descriptions are usually provided by users to briefly introduce themselves and may indicate users' summarized interests, we integrated the bios of a user's followees or those whose tweets are retweeted by this user into two kinds of bio documents, on which we further applied LDA model. We utilized the perplexity to fix the dimensions of topic-level features over followees' bio documents and retweetings' bio documents as 40 and 20, respectively. In this work, we only explored the contextual topics in Twitter, since we were unable to crawl the connections' profiles in LinkedIn and the bio descriptions are usually missing in Facebook.

5.3.3 Behavior-based features

This kind of features is characterized by users' posting behavior patterns and networking behavior patterns. The former focuses on the written style of users' social posts, while the latter captures their egocentric network features.

Posting behavior patterns. Posting behavior patterns have been investigated in many scenarios, spanning from

---

[8]Here we exclude the contents of section "Volunteer experience & Causes" and section "Experience".

[9]http://www.liwc.net/

[10]If A broadcasts a tweet posted by B, then B is A's a retweeting user.

age estimation to social spammers discovery [3, 13]. These patterns can be used to depict users' participation in information diffusion, which correlates with volunteerism tendency much.

On one hand, we employed the fraction of users' posts containing certain behaviors, including emoticons, slang words, acronyms, hashtags, URLs, and user mentions, to intuitively reflect users' engagement in topic discussion and social interaction. On the other hand, we observed that users' posting behaviors in social networks can be classified into a few categories. For example, posts in Twitter can be classified into two categories, $C_{tw} = \{tweets, retweets\}$, while posts in Facebook can roughly be split into eight types: $C_{fb} = \{share\_link, share\_sideo, share\_status, share\_photo, change\_photo, repost, post, tagged\}$. The distributions over users' posts on these categories also reflect their participation in information diffusion, revealing whether a given user tend to share information in social networks. When it comes to Linkedin, we utilized the profile completeness to characterize users' behaviors. Based on our observation, we found that volunteers tend to provide more information for all the sections. This not only reflects volunteers' active participation in LinkedIn but also signals their self-confidence and openness to public. Profile completeness is defined as a boolean vector over six dimensions to denote the presence of the six common sections in LinkedIn profiles: `summary`, `interest`, `language`, `education`, `skill` and `honor`. We excluded the sections on `experience` and `volunteer experience & causes`, because the ground truth is built on these two sections.

Egocentric network patterns. We also studied users' social behaviors from their egocentric networks. Intuitively, we believe that users belong to certain class tend to be connected with several class-specific accounts, as it goes for that "birds of a feather flock together". Therefore, volunteers should interact with some typical accounts in social media. The set of typical accounts is denoted as FV. Inspired by [17], we measured the degree of a user's correlation with volunteerism by three features: the frequency and fraction of a user's "friends" that belong to FV as well as the total number of "friends". In particular, we treated both the followees and retweetings as the "friends" of users in Twitter.

To construct the FV, we utilized the Twitter profile repository Wefollow[11], which allows us to find the most prominent people given a particular category. By crawling prominent users falling into categories of *Nonprofit*, *Charity*, *Volunteer*, *NGO*, *Community Service*, *Social Welfare* and *Christian* from Wefollow, we obtained $23,285$ accounts.

6. EXPERIMENTS

We conducted extensive experiments to comparatively verify our proposed scheme from various angles over a system equipped with Intel i72.60 GHz CPU, and 8 GB memory. In particular, we launched 10-fold cross validation for each experiment, and reported the average performance. Each fold involves $2,249$ training and 250 testing samples.

6.1 On Model Comparison

We compared **MSNL** with four baselines. Before that, the data was completed by **MSNDC**. We also performed significant test to validate the effectiveness of **MSNL**.

SVM: We chose the learning formulation with the kernel of radial-basis function. We implemented this method based on LIBSVM [6].

RLS: Regularized least squares model [11] aims to minimize the objective function of $\frac{1}{2N} \|\mathbf{y} - \mathbf{Xw}\|^2 + \frac{\lambda}{2} \|\mathbf{w}\|^2$. In fact, the **RLS** model can be deduced from **MSNL** via the settings of $\alpha = [\frac{1}{S}, \frac{1}{S}, \cdots, \frac{1}{S}]^T$, $\mu = 0$ and $\beta = 0$.

iSFS: The third baseline is the incomplete source-feature selection model proposed in [24]. This model only assigns weights to models learned from different social networks but ignores the relationships among them. We can derive **iSFS** from **MSNL** by making $\mu = 0$.

regMVMT: The fourth baseline is the regularized multi-view multi-task learning model [28]. This model only regulates the relationships among different views but fails to take the source confidence into account. We can derive **regMVMT** from **MSNL** by making $\alpha = [\frac{1}{S}, \frac{1}{S}, \cdots, \frac{1}{S}]^T$.

Table 2: Performance of different models(%).

Approaches	F1-measure	P-value
SVM	83.11	0.038
RLS	82.82	0.025
regMVMT	84.07	0.173
iSFS	84.72	0.281
MSNL	**85.59**	-

Table 2 shows the performance comparison between baselines and our proposed **MSNL**. We noticed that **MSNL** significantly outperforms the **SVM** and **RLS**. This implies that the information on multiple social networks are complementary and characterize users' volunteerism tendency consistently. This also proves that the correlations of different social networks with the task of volunteerism tendency prediction cannot be treated equally. In addition, **MSNL** achieves better performance, as compared with **iSFS** and **regMVMT**, which are the derivations of **MSNL**. This demonstrates that both the source confidence and the source consistency deserve particular attention.

6.2 On Data Completion Comparison

We further evaluated the component for missing data completion with the following three baseline methods.

Remove: This method eliminates all data samples that are not complete.

Average: This method imputes the missing features with the average values of the corresponding feature items.

KNN: The missing data is inferred by averaging its K-nearest neighbors. K is experimentally set as 1.

Table 3: Performance of different models over different data completion strategies.

Approaches	SVM	RLS	MSNL
Remove	74.91	74.66	81.81
Average	82.09	81.99	85.43
KNN	82.60	82.22	85.55
MSNDC	83.11	82.82	**85.59**

Table 3 shows the performance of different models over different data completion strategies. It can be seen that **MSNDC** outperforms the other strategies. Additionally, removing all incomplete data samples achieves the worst performance, which may be caused by the fact that it introduces training bias, making the dataset unbalanced and

reduces the size of training dataset. We found that the percentage of volunteer samples decreases from 50% to 40% after filtering out all incomplete data samples.

6.3 On Feature Comparison

To examine the discriminative features we extracted, we conducted experiments over different kinds of features using **MSNL**. We also performed significant test to validate the advantage of combining multiple social networks. Table 4 comparatively shows the performance of **MSNL** in terms of different feature configurations. It can be seen that the linguistic features achieves the best performance, as compared against demographic characteristics and behavior-based features. This reveals that volunteerism tendency is better reflected by their social contents, including their own social posts and the self-descriptions of their social connections. This also implies that users with volunteerism

Table 4: **Performance of different features(%).**

Features	F1-measure
Demographic characteristics	**68.43**
Linguistic features	**80.06**
User topics	75.04
Contextual topics	78.14
LIWC	68.48
Behavior-based features	**78.52**
Posting behavior patterns	69.83
Egocentric network patterns	75.91

tendency may talk about related topics and follow or retweet related social accounts. In addition, we found that contextual topics are more discriminative as compared to users' own topics. This may be due to the fact that users' self-descriptions are of more value and contain less noise than users' tweets. Some hot topics discussed by volunteers are given in Table 5. Besides, the egocentric network patterns also play a dominant role in our task. This implies that one's social connections indeed reflect the user's personal concerns to a large extent.

6.4 On Source Comparison

To demonstrate the descriptiveness of multiple social network integration, we conducted experiments over various source combinations. Notably, data from Facebook and LinkedIn is incomplete and we need to infer the block-wise missing data first taking advantage of the complete data samples from Twitter.

Table 6 shows the performance of **MSNL** over different social network combinations. We noted that the more sources we incorporate, the better the performance can be achieved. This implies the complementary relationships rather than mutual conflicting relationships among the sources. Moreover, we found that aggregating data from all these three social networks can achieve significantly better performance as compared to each of the single source. Additionally, as the performance obtained from different single social networks are not the same, this validates that incorporating the confidence of different social networks to **MSNL** is reasonable. Interestingly, we observed that **MSNL** over Twitter alone achieves the much better performance, as compared to that over LinkedIn or Facebook alone. This may be caused by the fact that the

Table 5: **Hot topics discussed by volunteers. Followee and retweeting: contextual topics; Self: user topics.**

| Data source	Topic words
Followee	• public, politics, rights, development
	• editor, global, journalist, university
Retweeting	• global, nonprofit, change, community
	• health, education, learning, university
Self	• woman, help, education, child
	• volunteer, nonprofit, support

most discriminative features evaluated by Section 6.3 are all extracted from Twitter.

Table 6: **Performance of different social network combinations(%). Facebook* and LinkedIn* both refer to the complete data, whose missing data is pre-inferred. F1: F1-measure.**

Social network combinations	F1	p-value
Twitter	82.35	$4.2e\text{-}2$
Facebook*	73.53	$5.0e\text{-}7$
LinkedIn*	74.49	$3.1e\text{-}7$
Twitter+Facebook*	83.67	$1.1e\text{-}1$
Twitter+LinkedIn*	83.84	$1.4e\text{-}1$
Facebook*+LinkedIn*	76.29	$6.0e\text{-}6$
Twitter+Facebook*+LinkedIn*	**85.59**	-

6.5 Size Varying of Positive Samples

In order to verify the usefulness of our model on real world dataset, where the volunteers should account for a minority portion of user population, we tuned the fraction of volunteer samples in our dataset. In particular, we fed $x\%$, $x \in [5, 50]$, of volunteer samples to our model with stepsize 5%. Figure 4 shows the F1-measure with respect to different fraction of volunteer samples of different models. As can be seen, our model can achieve satisfactory performance even when volunteer samples only accounts for 5% of the whole samples. This demonstrates that the proposed **MSNL** model is not sensitive to the percentage of positive samples. Whereas, **SVM** and **RLS** are relatively more sensitive to the fraction of volunteer samples in dataset.
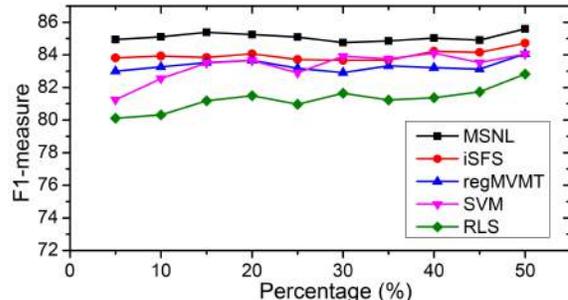
Figure 4: **F1-measure at different fraction of volunteer samples.**

6.6 Complexity Discussion

In order to analyze the complexity of **MSNL**, we need to solve the time complexity in terms of constructing **M**, **L** and **t** as defined in Eqn. (8) and Eqn. (12), and computing the inverse of **M** and **L**. Assume $D \gg S$, the construction of matrix **M** has a time complexity of

$\mathbf{O}(NDS)$, and the construction of matrix $\mathbf{L}$ has a time complexity of $\mathbf{O}(ND^2)$. Due to the fact that the cost of matrix multiplications $(\mathbf{X}_s^T\mathbf{X}_{s'})$ and that of constructing $\mathbf{t}$ involved in Eqn. (12) remain the same for all iterations and $\mathbf{L}$ is symmetric, we can save much practical time cost. Also, using the standard method, computing the inverse of two core matrices, $\mathbf{M}$ and $\mathbf{L}$, has the complexity of $\mathbf{O}(S^3)$ and $\mathbf{O}(D^3)$, respectively. Furthermore, using the method of Coppersmith and Winogard, the time cost can be bounded by $\mathbf{O}(S^{2.376})$ and $\mathbf{O}(D^{2.376})$ [26], respectively. We note that the speed bottleneck lies in the number of features and the number of social networks instead of the number of data samples. As S and D are usually small, especially S, **MSNL** should be efficient in time complexity.

To validate the practical efficiency of the proposed **MSNL** model, we conducted a set of experiments. The comparison of average time consumption of different models is shown in Table 7. As can be seen, **MSNL** shows superiority over **SVM** in terms of the time cost, which takes only 19% of the time that **SVM** uses. By careful observation, we observed that **MSNL** converges very fast, which on average takes about 20 iterations. Even though **MSNL** takes more time than **RLS** and **regMVMT** due to the consideration of source consistency and source confidence, it improves the performance in terms of F1-measure.

Table 7: Time cost of different models (%).

Approach	Total (s)	Train (s)	Test (s)
SVM	2.0550	1.8211	0.2339
RLS	0.0639	0.0631	0.0008
regMVMT	0.0605	0.0595	0.0006
iSFS	0.5565	0.5557	0.0008
MSNL	0.3936	0.3929	0.0007

7. CONCLUSIONS AND FUTURE WORK

This paper presented a novel scheme for multiple social network learning. This scheme takes the source confidence and source consistency into consideration by introducing regularization to the objective function. We further demonstrated that the proposed scheme, designed for complete data, is also able to handle the real and more challenging cases where there exists block-wise missing data. In particular, before feeding the data into the proposed MSNL model, we inferred the missing data via NMF technique. Furthermore, we practically evaluated the proposed scheme in an interesting scenario of volunteerism tendency prediction. We developed a set of volunteer-oriented features to characterize users' volunteerism tendency. Experimental results demonstrated the effectiveness of our proposed scheme and verified the advantages of utilizing multiple social network over a single source. Currently, we only consider solving a single task in the proposed scheme. In the future, we will extend our work to the context of multiple task learning.

Acknowledgments

8. REFERENCES

[1] F. Abel, E. Herder, G.-J. Houben, N. Henze, and D. Krause. Cross-system user modeling and personalization on the social web. *UMUAI*, 2013.

[2] B. Bazelli, A. Hindle, and E. Stroulia. On the personality traits of stackoverflow users. In *ICSM*, 2013.

[3] F. Benevenuto, T. Rodrigues, V. Almeida, J. Almeida, and M. Gonçalves. Detecting spammers and content promoters in online video social networks. In *SIGIR*, 2009.

[4] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *JMLR*, 2003.

[5] D. Carmel, N. Zwerdling, I. Guy, S. Ofek-Koifman, N. Har'El, I. Ronen, E. Uziel, S. Yogev, and S. Chernov. Personalized social search based on the user's social network. In *CIKM*, 2009.

[6] C.-C. Chang and C.-J. Lin. Libsvm: a library for support vector machines. *TIST*, 2011.

[7] Y. Gao, M. Wang, Z.-J. Zha, J. Shen, X. Li, and X. Wu. Visual-textual joint relevance learning for tag-based social image search. *TIP*, 2013.

[8] J. Guo, G. Xu, X. Cheng, and H. Li. Named entity recognition in query. In *SIGIR*, 2009.

[9] C.-J. Hsieh and I. S. Dhillon. Fast coordinate descent methods with variable selection for non-negative matrix factorization. In *SIGKDD*, 2011.

[10] S.-Y. L. Y. Jiang and Z.-H. Zhou. Partial multi-view clustering. In *AAAI*, 2014.

[11] S. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky. A method for large-scale l1-regularized least squares problems with applications in signal processing and statistics. *J-STSP*, 2007.

[12] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*.

[13] K. Lee, J. Caverlee, and S. Webb. Uncovering social spammers: social honeypots+ machine learning. In *SIGIR*, 2010.

[14] D. Li, B. He, Y. Ding, J. Tang, C. Sugimoto, Z. Qin, E. Yan, J. Li, and T. Dong. Community-based topic modeling for social tagging. In *CIKM*, 2010.

[15] D. Markovikj, S. Gievska, M. Kosinski, and D. Stillwell. Mining facebook data for predictive personality modeling. In *ICWSM*, 2013.

[16] L. Nie, Y.-L. Zhao, X. Wang, J. Shen, and T.-S. Chua. Learning to recommend descriptive tags for questions in social forums. *TOIS*, 2014.

[17] M. Pennacchiotti and A.-M. Popescu. Democrats, republicans and starbucks afficionados: user classification in twitter. In *SIGKDD*, 2011.

[18] L. A. Penner. Dispositional and organizational influences on sustained volunteerism: An interactionist perspective. *JSI*, 2002.

[19] L. A. Penner. Volunteerism and social problems: Making things better or worse? *JSI*, 2004.

[20] D. Quercia, M. Kosinski, D. Stillwell, and J. Crowcroft. Our twitter profiles, our selves: predicting personality with twitter. In *SocialCom*, 2011.

[21] D. Quercia, R. Lambiotte, D. Stillwell, M. Kosinski, and J. Crowcroft. The personality of popular facebook users. In *CSCW*, 2012.

[22] X. Shen, B. Tan, and C. Zhai. Implicit user modeling for personalized search. In *CIKM*, 2005.

[23] X. Wei and W. B. Croft. Lda-based document models for ad-hoc retrieval. In *SIGIR*, 2006.

[24] S. Xiang, L. Yuan, W. Fan, Y. Wang, P. M. Thompson, and J. Ye. Multi-source learning with block-wise missing data for alzheimer's disease prediction. In *SIGKDD*, 2013.

[25] L. Yuan, Y. Wang, P. M. Thompson, V. A. Narayan, and J. Ye. Multi-source learning for joint analysis of incomplete multi-modality neuroimaging data. In *SIGKDD*, 2012.

[26] D. Zhai, H. Chang, S. Shan, X. Chen, and W. Gao. Multiview metric learning with global consistency and local smoothness. *TIST*, 2012.

[27] D. Zhang, F. Wang, and L. Si. Composite hashing with multiple information sources. In *SIGIR*, 2011.

[28] J. Zhang and J. Huan. Inductive multi-task learning with multiple view data. In *SIGKDD*, 2012.

[29] X. Zhu, Z.-Y. Ming, X. Zhu, and T.-S. Chua. Topic hierarchy construction for the organization of multi-source user generated contents. In *SIGIR*, 2013.