Fusing Social Networks with Deep Learning for Volunteerism Tendency Prediction

Yongpo Jia[1,2], **Xuemeng Song**[2], **Jingbo Zhou**[3], **Li Liu**[2], **Liqiang Nie**[2], **David S. Rosenblum**[2]
[1]NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore, Singapore
[2]School of Computing, National University of Singapore, Singapore
[3]Big Data Lab, Baidu Research, China
{jiayongpo, dcsliuli}@nus.edu.sg, {sxmustc, nieliqiang}@gmail.com, jzhou@baidu.com, david@comp.nus.edu.sg

Abstract

Social networks contain a wealth of useful information. In this paper, we study a challenging task for integrating users' information from multiple heterogeneous social networks to gain a comprehensive understanding of users' interests and behaviors. Although much effort has been dedicated to study this problem, most existing approaches adopt linear or shallow models to fuse information from multiple sources. Such approaches cannot properly capture the complex nature of and relationships among different social networks. Adopting deep learning approaches to learning a joint representation can better capture the complexity, but this neglects measuring the level of confidence in each source and the consistency among different sources. In this paper, we present a framework for multiple social network learning, whose core is a novel model that fuses social networks using deep learning with source confidence and consistency regularization. To evaluate the model, we apply it to predict individuals' tendency to volunteerism. With extensive experimental evaluations, we demonstrate the effectiveness of our model, which outperforms several state-of-the-art approaches in terms of precision, recall and F_1-score.

1 Introduction

Social networks have revolutionized the way we network, and they provide a wealth of useful information to gain a comprehensive understanding of users' interests and behaviors. Constrained by different designs of social networks, each single source only provides partial information of a user from a certain perspective. Therefore, linking and aggregating information from multiple social networks can enrich a given user's profile and enable us to comprehensively understand the given users (Abel et al. 2013; Zhu et al. 2013; Liu et al. 2014).

However, it is non-trivial to fuse multiple heterogeneous social networks for learning. First of all, it is challenging to deal with the complex nature of heterogeneous social networks (Mislove et al. 2007). Models with non-linear properties and rich modeling capabilities should be considered for tackling this problem. Meanwhile, users use different social networks differently, for example, using Facebook to keep connected with friends, LinkedIn for professional networking, and Twitter to follow information about specific interests. In other words, information from each source will not contribute equally to one given profiling task. Moreover, since such multiple sources are naturally presumed to characterize the same given user, the consistency of data from difference sources must be taken into consideration during information fusion. Therefore, it is desirable to devise an effective approach for multiple social network learning that not only can better capture this complexity, but also can measure the level of confidence in each source as well as the consistency among different sources.

Great effort has been spent on this challenging multi-source learning task. Concatenating all sources into a unified representation is a widely-used method. However, this treats all sources equally without measuring the level of confidence in each source and the consistency among different sources. State-of-the-art approaches (Song et al. 2015; Xiang et al. 2013; Zhang and Huan 2012) employ source confidence or consistency regularization to address this problem. However, such approaches adopt linear models to fuse information from multiple sources for learning, which cannot properly capture the complex nature of social networks. Another method is to use deep learning models to learn a joint representation (Srivastava and Salakhutdinov 2012). These deep learning models may better capture the complexity, but fail to explicitly model the source confidence and consistency to better resolve the problem.

Based on above insights, we present a framework for multiple social network learning, whose core is a novel model that *Fuses sociAl netwoRks uSing dEep lEarnING* (FARSEEING) with source confidence and consistency regularization. Figure 1 presents the framework, which consists of three parts. First, given a set of users, the same users are aligned by linking their multiple social accounts and then their publicly available historical data are crawled from all sources. Second, multi-faceted features, such as demographic, linguistic and behavioral features, are extracted to characterize the given users. Before feeding the extracted features into FARSEEING, missing data, which is caused by users' unbalanced activities in different social networks, are inferred from the learned shared feature spaces by non-negative matrix factorization.

Finally, we use FARSEEING for multiple social network learning. The model consists of two main stages. First, in the feature transformation stage, low-level features are mapped

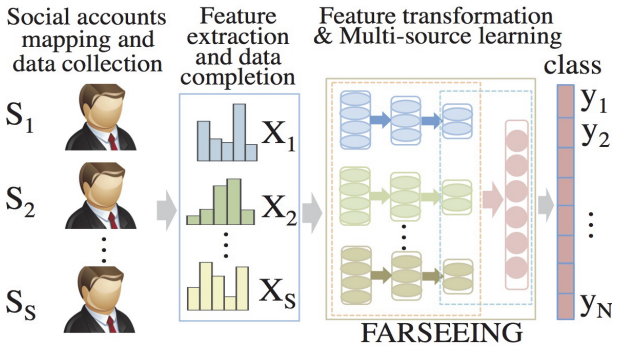

Figure 1: Illustration of our framework which comprises of three components.

into high-level features using deep learning. Second, in the multi-source learning stage, the high-level features from all sources are fused together for learning. Specifically, we integrate source confidence and consistency regularization into the deep neural networks' optimization process to regulate the level of confidence in each source and measure the consistency among different sources. This leads to an enhanced multi-source deep learning model for multi-source learning.

To evaluate the model, we apply it to predict individuals' tendency to volunteerism, which is desirable for nonprofit organizations (NPOs) to identify candidate volunteers from the crowd. We make the prediction by leveraging information from social networks to gain a comprehensive understanding of a user's interests, behaviors and personal traits.

Our main contributions are summarized as follows:

- We present a novel multi-source deep learning model that fuses social networks using deep learning with source confidence and consistency regularization. This model is the core component of our introduced framework for multiple social network learning.
- We apply the framework to predict individuals' tendency to volunteerism. Extensive experiments validate the effectiveness of our model, which outperforms state-of-the-art competitors in terms of precision, recall and F_1-score.

The rest of the paper is organized as follows: Section 2 formalizes the problem and describes the FARSEEING model. Section 3 presents the details of our framework and data preparation for experiments. Section 4 details the experimental results and analysis. Section 5 presents a review of related work, followed by the conclusion in Section 6.

2 FARSEEING Model

We use deep learning to transform low-level features into high-level feature spaces and then employ ridge regression to fit a final prediction model. The novelty lies in integrating the source confidence and consistency regularization into a unified multi-source deep learning model.

2.1 Notation

We use bold capital letters (e.g., $\mathbf{X}$) and bold lower case letters (e.g., $\mathbf{x}$) to denote matrices and vectors, respectively. We employ nonbold letters (e.g., x) to represent scalars, and Greek letters (e.g., λ) as parameters.

Suppose we have a set of N labeled data samples from S social networks (with $S \geq 2$). We denote D_s, N_s as the number of features and samples in the s-th social network, respectively. We use $\mathbf{X}_s \in \mathbb{R}^{N \times D_s}$ to denote the feature matrix extracted from the s-th social network, where each row represents a user sample. The dimensionality of features extracted from all these social networks is $D = \sum_{s=1}^{S} D_s$. Then 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 $\mathbf{y} = [y_1, y_2, \cdots, y_N]^T \in \{1, -1\}^N$ is the corresponding label vector.

We denote $\hat{\mathbf{X}}_s = \mathbf{f}_s(\mathbf{X}_s) \in \mathbb{R}^{N \times F_s}$ as the high-level feature matrix extracted by a deep learning model $\mathbf{f}_s(\cdot)$ from the s-th social network, and F_s is the number of corresponding high-level features. We use $\hat{\mathbf{X}} = [\hat{\mathbf{X}}_1, \hat{\mathbf{X}}_2, \cdots, \hat{\mathbf{X}}_S] \in \mathbb{R}^{N \times F}$ to denote the output of $\mathbf{f}(\mathbf{X})$, where $F = \sum_{s=1}^{S} F_s$ is the total output dimension.

2.2 Deep Neural Networks

Deep learning approaches can model complex non-linear relationships. In this paper, we adopt *Deep Neural Networks* (DNNs) to extract high-level features, since the DNN is a typical model for deep learning and can be generalized to others as well. In general, a DNN is an *artificial neural network* (ANN), which is inspired by biological neural networks, with multiple hidden layers.

In a typical DNN, the output vector $\mathbf{h}^k$ of the k-th layer is computed by using the output $\mathbf{h}^{k-1}$ of the previous layer as input, starting with an input $\mathbf{h}^0 = \mathbf{x}$, which can be represented as follows (Rumelhart, Hintont, and Williams 1986; Bengio 2009):

$$\mathbf{h}^k = \mathbf{g}(\boldsymbol{b}^k + \mathbf{W}^k \mathbf{h}^{k-1}), \tag{1}$$

where $\boldsymbol{b}^k$ and $\mathbf{W}^k$ are the variables to learn, and $\mathbf{g}(\cdot)$ is an activation function. Typical choices for $\mathbf{g}(\cdot)$ include the $tanh$ and $sigmoid$ functions. In this paper, we employ the $tanh$ function, as it usually yields to faster training.

Typically, the $softmax$ function is applied to the top layer output $\mathbf{h}^K$ to make a prediction. A key innovation in our model is that, instead of using $softmax$ to predict the labels, we fuse the high-level features (i.e., the top-layer outputs) of multiple per-source DNNs for multi-source learning-based prediction. We expound on this in the following sections.

2.3 Problem Formulation

Given data $\mathbf{X}$ from S social networks, our objective is to learn a predictive model $\mathbf{p}(\mathbf{X})$ to minimize the following objective function $\boldsymbol{\Gamma}$,

$$\boldsymbol{\Gamma} = \min_{\mathbf{p}} \frac{1}{2N} ||\mathbf{y} - \mathbf{p}(\mathbf{X})||^2 + \frac{\lambda}{2} ||\mathbf{p}||^2, \tag{2}$$

where λ is a nonnegative regularization parameter.

Since $\mathbf{X}$ is from S social networks, we can learn S predictive models $\mathbf{p}_s(\mathbf{X}_s)$ independently and then combine these models to form the final predictive model $\mathbf{p}(\mathbf{X})$. As discussed before, different social networks always contribute differently to the final prediction, so we introduce a weight vector $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \cdots, \alpha_S]^T \in \mathbb{R}^S$ to regulate the level of

confidence in each social network. Then the final predictive model can be defined as follows:

$$\mathbf{p}(\mathbf{X}) = \sum_{s=1}^{S} \alpha_s \mathbf{p}_s(\mathbf{X}_s), \tag{3}$$

where the value of α_s can be either positive or negative. Positive α_s rewards and negative one dampens the correlation among different sources which may contain unreliable or noisy data.

Moreover, since the data from multiple social networks characterize the same given user, we measure the consistency among different sources using source consistency regularization, which is defined as follows:

$$\sum_{s=1}^{S} \sum_{s' \neq s} ||\mathbf{p}_s(\mathbf{X}_s) - \mathbf{p}_{s'}(\mathbf{X}_{s'})||^2. \tag{4}$$

To better capture the complexity, we adopt a DNN $\mathbf{f}_s(\cdot)$ to iteratively learn and optimize high level features $\hat{\mathbf{X}}_s$ from $\mathbf{X}_s$ and then use $\hat{\mathbf{X}}_s$ to learn the final predictive model. The predictive model $\mathbf{p}_s(\cdot)$ is defined as follows:

$$\mathbf{p}_s(\mathbf{X}_s) = \hat{\mathbf{X}}_s \mathbf{w}_s + \boldsymbol{b}_s = \mathbf{f}_s(\mathbf{X}_s)\mathbf{w}_s + \boldsymbol{b}_s, \tag{5}$$

where $\mathbf{w}_s \in \mathbb{R}^{F_s}$ is the mapping model, and $\boldsymbol{b}_s = b_s \mathbf{v} \in \mathbb{R}^N$, with $\mathbf{v}$ an identity vector, and b_s the bias term.

Then the objective function $\mathbf{\Gamma}$ can be rewritten as follows:

$$\begin{aligned}\mathbf{\Gamma} = & \min_{\{\boldsymbol{\theta}_s, \mathbf{w}_s, \alpha_s, \boldsymbol{b}_s\}} \frac{1}{2N}||\mathbf{y} - \sum_{s=1}^{S} \alpha_s(\mathbf{f}_s(\mathbf{X}_s)\mathbf{w}_s + \boldsymbol{b}_s)||^2 \\ & + \frac{\mu}{4N} \sum_{s=1}^{S} \sum_{s' \neq s} ||(\mathbf{f}_s(\mathbf{X}_s)\mathbf{w}_s + \boldsymbol{b}_s) - (\mathbf{f}_{s'}(\mathbf{X}_{s'})\mathbf{w}_{s'} + \boldsymbol{b}_{s'})||^2 \\ & + \frac{\lambda}{2} \sum_{s=1}^{S} ||\mathbf{w}_s||^2 + \frac{\gamma}{2} \sum_{s=1}^{S} ||\boldsymbol{\theta}_s||^2 + \frac{\beta}{2}||\boldsymbol{\alpha}||^2,\end{aligned} \tag{6}$$

where $\boldsymbol{\theta}_s$ denotes the variables in $\mathbf{f}_s(\cdot)$ and μ, λ, γ, and β are the regularization parameters.

2.4 Optimization

We adopt the alternating optimization strategy to solve the variables $\boldsymbol{\theta}_s, \mathbf{w}_s, \alpha_s$ and $\boldsymbol{b}_s$ in Eqn. (6). In particular, we iteratively optimize one variable with others fixed and maintain this procedure until meeting a predefined stop condition.

Compute $\boldsymbol{\theta}_s$ with others fixed When other variables are fixed, $\boldsymbol{\theta}_s$ can be updated as follows:

$$\boldsymbol{\theta}_s^{(t+1)} = \boldsymbol{\theta}_s^{(t)} - \eta \frac{\partial \mathbf{\Gamma}}{\partial \boldsymbol{\theta}_s^{(t)}}. \tag{7}$$

This equation leads to the well-known back-propagation algorithm for DNN. Here we employ Theano[1] to train the DNN model using Stochastic Gradient Descent with mini-batches. It is not necessary to provide an analytic solution for $\frac{\partial \mathbf{\Gamma}}{\partial \boldsymbol{\theta}_s^{(t)}}$ as Theano can perform automatic differentiation.

[1]It is a Python library. http://deeplearning.net/software/theano/

Compute $\mathbf{w}_s$ with others fixed When other variables are fixed, the derivative of $\mathbf{\Gamma}$ with respect to $\mathbf{w}_s$ is as follows:

$$\begin{aligned}\frac{\partial \mathbf{\Gamma}}{\partial \mathbf{w}_s} = & \frac{1}{N}\alpha_s \hat{\mathbf{X}}_s^T (\sum_{s=1}^{S} \alpha_s(\hat{\mathbf{X}}_s \mathbf{w}_s + \boldsymbol{b}_s) - \mathbf{y}) + \lambda \mathbf{w}_s \\ & + \frac{\mu}{N} \hat{\mathbf{X}}_s^T \sum_{s' \neq s} ((\hat{\mathbf{X}}_s \mathbf{w}_s + \boldsymbol{b}_s) - (\hat{\mathbf{X}}_{s'} \mathbf{w}_{s'} + \boldsymbol{b}_{s'})) \\ = & [\lambda \mathbf{I} + \frac{\alpha_s^2}{N} \hat{\mathbf{X}}_s^T \hat{\mathbf{X}}_s + \frac{\mu(S-1)}{N} \hat{\mathbf{X}}_s^T \hat{\mathbf{X}}_s] \mathbf{w}_s + \frac{\alpha_s^2}{N} \hat{\mathbf{X}}_s^T \boldsymbol{b}_s \\ & + \frac{\mu(S-1)}{N} \hat{\mathbf{X}}_s^T \boldsymbol{b}_s + \sum_{s' \neq s} \frac{1}{N}(\alpha_s \alpha_{s'} - \mu) \hat{\mathbf{X}}_s^T \hat{\mathbf{X}}_{s'} \mathbf{w}_{s'} \\ & + \sum_{s' \neq s} \frac{1}{N}(\alpha_s \alpha_{s'} - \mu) \hat{\mathbf{X}}_s^T \boldsymbol{b}_{s'} - \frac{\alpha_s}{N} \hat{\mathbf{X}}_s^T \mathbf{y},\end{aligned} \tag{8}$$

where $\mathbf{I}$ is a $F_s \times F_s$ identity matrix. Setting Eqn. (8) to zero and rearranging the terms, all $\mathbf{w}_s$ can be learned jointly by a linear system $\mathbf{L}\mathbf{w} = \mathbf{t}$, where $\mathbf{L} \in \mathbb{R}^{F \times F}$ is a sparse block matrix with $S \times S$ blocks, $\mathbf{w} = [\mathbf{w}_1^T, \mathbf{w}_2^T, \cdots, \mathbf{w}_S^T]^T \in \mathbb{R}^F$ and $\mathbf{t} = [\mathbf{t}_1^T, \mathbf{t}_2^T, \cdots, \mathbf{t}_S^T]^T \in \mathbb{R}^F$ are vectors with S blocks. $\mathbf{t}_s, \mathbf{L}_{ss}$ and $\mathbf{L}_{ss'}$ are defined as follows:

$$\begin{cases} \mathbf{t}_s = \frac{\alpha_s}{N} \hat{\mathbf{X}}_s^T \mathbf{y} - \frac{\alpha_s^2 + \mu(S-1)}{N} \hat{\mathbf{X}}_s^T \boldsymbol{b}_s - \sum_{s' \neq s} \frac{\alpha_s \alpha_{s'} - \mu}{N} \hat{\mathbf{X}}_s^T \boldsymbol{b}_{s'}, \\ \mathbf{L}_{ss} = \lambda \mathbf{I} + \frac{\alpha_s^2}{N} \hat{\mathbf{X}}_s^T \hat{\mathbf{X}}_s + \frac{\mu(S-1)}{N} \hat{\mathbf{X}}_s^T \hat{\mathbf{X}}_s, \\ \mathbf{L}_{ss'} = \frac{1}{N}(\alpha_s \alpha_{s'} - \mu) \hat{\mathbf{X}}_s^T \hat{\mathbf{X}}_{s'}. \end{cases} \tag{9}$$

When other variables are fixed, $\mathbf{t}$ can be treated as a constant vector. $\mathbf{L}$ is symmetric as $\mathbf{L}_{ss'} = \mathbf{L}_{s's}^T$. Since $\mathbf{L}$ is a positive-definite matrix and thus invertible (Song et al. 2015), we can derive the solution of $\mathbf{w}$ as follows:

$$\mathbf{w} = \mathbf{L}^{-1}\mathbf{t}. \tag{10}$$

Compute $\boldsymbol{\alpha}$ with others fixed When other variables are fixed, the objective function $\mathbf{\Gamma}$ can be rewritten as follows:

$$\min_{\boldsymbol{\alpha}} \frac{1}{2N}||\mathbf{y} - (\hat{\mathbf{X}}\mathbf{W} + \mathbf{B})\boldsymbol{\alpha}||^2 + \frac{\beta}{2}||\boldsymbol{\alpha}||^2, \tag{11}$$

where $\mathbf{W} = diag(\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_S) \in \mathbb{R}^{F \times S}$, and $\mathbf{B} = diag(\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_S) \in \mathbb{R}^{N \times S}$. Taking the derivative of $\mathbf{\Gamma}$ with respect to $\boldsymbol{\alpha}$, we have,

$$\frac{\partial \mathbf{\Gamma}}{\partial \boldsymbol{\alpha}} = \frac{1}{N}(\hat{\mathbf{X}}\mathbf{W} + \mathbf{B})^T(\hat{\mathbf{X}}\mathbf{W} + \mathbf{B})\boldsymbol{\alpha} - (\hat{\mathbf{X}}\mathbf{W} + \mathbf{B})^T\mathbf{y} + \beta\boldsymbol{\alpha}. \tag{12}$$

Setting Eqn. (12) to zero, it can be derived that,

$$\boldsymbol{\alpha} = (\frac{1}{N}(\hat{\mathbf{X}}\mathbf{W} + \mathbf{B})^T(\hat{\mathbf{X}}\mathbf{W} + \mathbf{B}) + \beta\mathbf{I})^{-1}(\hat{\mathbf{X}}\mathbf{W} + \mathbf{B})^T\mathbf{y}. \tag{13}$$

Since $(\frac{1}{N}(\hat{\mathbf{X}}\mathbf{W}+\mathbf{B})^T(\hat{\mathbf{X}}\mathbf{W}+\mathbf{B})+\beta\mathbf{I}) \in \mathbb{R}^{S \times S}$ is positive-definite and invertible according to the definition, we can obtain the analytic solution of $\boldsymbol{\alpha}$ as Eqn. (13).

Compute $\boldsymbol{b}_s$ with others fixed When other variables are fixed, the derivative of $\mathbf{\Gamma}$ with respect to $\boldsymbol{b}_s$ is as follows:

$$\begin{aligned}\frac{\partial \mathbf{\Gamma}}{\partial \boldsymbol{b}_s} = & \frac{1}{N}\alpha_s(\sum_{s=1}^{S} \alpha_s(\hat{\mathbf{X}}_s \mathbf{w}_s + \boldsymbol{b}_s) - \mathbf{y}) \\ & + \frac{\mu}{N} \sum_{s' \neq s} ((\hat{\mathbf{X}}_s \mathbf{w}_s + \boldsymbol{b}_s) - (\hat{\mathbf{X}}_{s'} \mathbf{w}_{s'} + \boldsymbol{b}_{s'})) \\ = & \frac{\alpha_s^2 + \mu(S-1)}{N} \boldsymbol{b}_s + \sum_{s' \neq s} \frac{1}{N}(\alpha_s \alpha_{s'} - \mu) \boldsymbol{b}_{s'} - \frac{\alpha_s}{N} \mathbf{y} \\ & + \frac{\alpha_s^2 + \mu(S-1)}{N} \hat{\mathbf{X}}_s \mathbf{w}_s + \sum_{s' \neq s} \frac{1}{N}(\alpha_s \alpha_{s'} - \mu) \hat{\mathbf{X}}_{s'} \mathbf{w}_{s'}.\end{aligned} \tag{14}$$

Similar to the process of computing $\mathbf{w}_s$ with the others fixed, all $\boldsymbol{b}_s$ can be learned jointly by a linear system. We have $\boldsymbol{b}_s = b_s\mathbf{v} \in \mathbb{R}^N$ where $\mathbf{v}$ is an identity vector, and we define $\boldsymbol{b} = [\boldsymbol{b}_1^T, \boldsymbol{b}_2^T, \cdots, \boldsymbol{b}_S^T]^T \in \mathbb{R}^{(N\cdot S)}$. We construct a linear system with the form $\mathbf{A}\boldsymbol{b} = \boldsymbol{c}$, where $\mathbf{A} \in \mathbb{R}^{(N\cdot S)\times(N\cdot S)}$ is a sparse block matrix with $S \times S$ blocks. Each block in $\mathbf{A}$ is an $N \times N$ diagonal matrix. For $\mathbf{c}$ we have $\mathbf{c}_s \in \mathbb{R}^N$ and $\mathbf{c} = [\mathbf{c}_1^T, \mathbf{c}_2^T, \cdots, \mathbf{c}_S^T]^T \in \mathbb{R}^{(N\cdot S)}$. $\mathbf{c}_s$, $\mathbf{A}_{ss}$ and $\mathbf{A}_{ss'}$ are defined as follows:

$$\begin{cases} \mathbf{c}_s = \frac{\alpha_s}{N}\mathbf{y} - \frac{\alpha_s^2+\mu(S-1)}{N}\hat{\mathbf{X}}_s\mathbf{w}_s - \sum\limits_{s'\neq s}\frac{\alpha_s\alpha_{s'}-\mu}{N}\hat{\mathbf{X}}_{s'}\mathbf{w}_{s'}, \\ \mathbf{A}_{ss} = \frac{\alpha_s^2+\mu(S-1)}{N}\mathbf{I}, \\ \mathbf{A}_{ss'} = \frac{1}{N}(\alpha_s\alpha_{s'}-\mu)\mathbf{I}, \end{cases} \tag{15}$$

where $\mathbf{I}$ is an identity matrix with $\mathbf{I} \in \mathbb{R}^{N\times N}$. Similar to $\mathbf{L}$, $\mathbf{A}$ is also a positive-definite matrix and thus invertible. Then we can derive the solution of $\boldsymbol{b}$ as follows:

$$\boldsymbol{b} = \mathbf{A}^{-1}\mathbf{c}. \tag{16}$$

3 The Framework

To evaluate our approach, we apply it to predict individuals' tendency to volunteerism. We cast the problem of volunteerism tendency prediction as a binary classification task. If the predicted score of a user is larger than a threshold value, we regard this user as a prospective volunteer.

3.1 Data Preparation

We collected data[2] from three popular social networks, namely Facebook, LinkedIn and Twitter. They are representatives of private, professional and public social networks, respectively. The data preparation process can be found in detail in our previous work (Song et al. 2015). We briefly summarize this process next.

Social Accounts Mapping Quora and About.me[3], where a user's multiple social accounts are explicitly listed, are used to align the same user across social networks (Abel et al. 2013). Only the users who have accounts in Facebook, LinkedIn and Twitter were kept, and then their publicly available historical data from all sources were crawled.

Ground Truth Construction Information on LinkedIn is presumed to be more reliable and complete. Therefore, a user is regarded as a volunteer if and only if this user lists volunteer experiences in the section "Volunteer Experience & Causes" or the section "Experience" in the profile. Finally, we obtained 1,284 volunteers and 1,215 non-volunteers.

Low-level Feature Extraction Generally, three kinds of features were extracted. The first kind is demographic features (Penner 2002), such as *Gender, Relationship Status, Education Level*, and *Number of Social Connections*. The second kind is linguistic features, including *Linguistic Inquiry and Word Count (LIWC)* features (Bazelli, Hindle, and Stroulia 2013; Markovikj et al. 2013), *User Topics* and *Contextual Topics* (Blei, Ng, and Jordan 2003; Wei and Croft 2006; Guo et al. 2009). The third kind is behavioral features, including *Posting Patterns* and *Egocentric Network Patterns* (Pennacchiotti and Popescu 2011).

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

[3]Their links are: https://www.quora.com/ and https://about.me/

3.2 Missing Data Completion

Unbalanced user activities among different social networks cause a problem of data missing. For example, we consider a Facebook user to be inactive if the user has fewer than 10 historical posts, and such users' Facebook data are treated as missing. Before feeding the extracted features into our FARSEEING model, we use Non-negative Matrix Factorization (NMF) (Lee and Seung 2001) to explore the shared latent spaces of different social networks, and further infer the missing data based upon these latent spaces (Li, Jiang, and Zhou 2014; Song et al. 2015).

3.3 Feature Transformation and Fusion

We feed the completed data into our FARSEEING model for multiple social network learning. Since it is non-trivial to fuse these complex data for learning, the FARSEEING model adopts DNNs to iteratively transform the completed low-level features into high-level feature spaces. Then we use the transformed high-level features to build the final predictive model, which has been shown in Section 2.

4 Experiment

We implemented both FARSEEING and a set of competitors in Python for comparative evaluation, and we conducted extensive experiments to evaluate our model.

4.1 Baselines and Experimental Settings

We compared the performance of our FARSEEING model with nine approaches, including the SVM, three state-of-the-art linear approaches, two deep learning approaches, and three variants of our FARSEEING model.

- **SVM**: We chose SVM with radial basis function (RBF) kernel as a baseline approach for comparison.
- **regMVMT**: The *regularized Multi-View Multi-Task* (regMVMT) model uses a linear model to fuse all sources with source consistency only (Zhang and Huan 2012).
- **iSFS**: The *incomplete Source-Feature Selection* (iSFS) model (Xiang et al. 2013) uses a linear model to fuse all sources. It measures the source confidence only.
- **MSNL**: The *Multi-Social Network Learning* (MSNL) (Song et al. 2015) uses a linear model to fuse all sources. It measures the source confidence and consistency.
- **DBN**: The *Deep Belief Networks* (DBN) model (Hinton and Salakhutdinov 2006) fuses all sources by concatenating all input sources together, which does not measure the source confidence or consistency.
- **M-DBM**: The *Multimodal Deep Boltzmann Machine* (M-DBM) model fuses all sources by learning a joint representation (Srivastava and Salakhutdinov 2012). It does not measure the source confidence or consistency.

- **FARSEEING**-no: This is a variant of FARSEEING with $\boldsymbol{\alpha} = [\frac{1}{S}, \cdots, \frac{1}{S}]$ and $\mu = 0$, i.e. without measuring source confidence or consistency.
- **FARSEEING**-na: This is a variant of FARSEEING with $\boldsymbol{\alpha} = [\frac{1}{S}, \cdots, \frac{1}{S}]$, i.e. without measuring source confidence.
- **FARSEEING**-nw: This is a variant of FARSEEING with $\mu = 0$, i.e. without measuring source consistency.

To save the cost of memory and computation, we use three hidden layers to construct our FARSEEING model and its variants as well as the DBN and M-DBM, which is sufficient to achieve good performance. Although we also conducted experiments using FARSEEING with four and more hidden layers, no significant improvement has been observed. We follow established procedures (Bengio 2009) to train the deep learning models to avoid getting stuck in local minima.

To avoid overfitting and achieve the best performance, we selected the optimal parameters for each model based on 10-fold cross validation, and we performed another 9-fold cross validation on the training data with grid search in each round (i.e. nested cross-validation). Hence, in each experiment, for each round of the 10-fold cross validation, 90% of the samples were used for training the model with 9-fold cross validation, and the remaining 10% were reserved for testing. For the grid search, it was conducted between 10^{-2} and 10^{2} with small but adaptive step sizes. The step sizes are 0.01, 0.05, 0.5 and 5 for the range of $[0.01, 0.1]$, $[0.1, 1]$, $[1, 10]$ and $[10, 100]$, respectively (Nie et al. 2015).

4.2 Evaluation Scheme

Our main evaluation metric is the F_1-score, which balances precision and recall as $F_1 = 2\frac{precision \cdot recall}{precision + recall}$, and the average results over the aforementioned 10-fold cross validation are reported. We also applied a pairwise t-test between our model and each of the competitors, and we present the resulting p-values to indicate the statistical significance of the improvements achieved by our model.

4.3 Model Performance

Table 1 presents the performance comparison of different models. We can see from Table 1 that multi-source deep learning models, namely M-DBM, our FARSEEING model and its variants, outperform state-of-the-art linear models, namely regMVMT, iSFS and MSNL, in terms of F_1-score. This demonstrates that deep learning approaches can better capture the complex nature of social networks than linear models. We also observe from this table that multi-source deep learning approaches outperform DBN, which means that simply concatenating all sources together is not an effective way for multiple social network learning.

Approach	Precision	Recall	F_1-score	p-value
SVM	83.177	83.334	83.202	0.0004
regMVMT	83.694	86.205	84.835	0.0029
iSFS	84.533	88.734	86.254	0.0282
MSNL	87.623	85.710	86.641	0.0798
DBN	87.062	86.707	86.635	0.0569
M-DBM	87.585	88.578	88.085	0.0933
FARSEEING-no	87.238	88.373	87.675	0.1949
FARSEEING-na	87.361	88.585	87.867	0.1586
FARSEEING-nw	87.518	91.121	89.101	0.6358
FARSEEING	**88.069**	**92.197**	**89.921**	-

Table 1: Model performance (%) and significance test.

The experimental data in Table 1 also demonstrate that both the source confidence and the source consistency deserve particular attention. First, the experiment demonstrates that source confidence regularization is critical for fusing information from multiple social networks. Our FARSEEING model and its variant with source confidence regularization only, FARSEEING-nw, outperform M-DBM significantly. Though we notice that M-DBM outperforms two variants of our model without source confidence regularization, FARSEEING-no and FARSEEING-na, the difference is not significant. Moreover, by comparing the performance between FARSEEING-na and FARSEEING-no, FARSEEING and FARSEEING-nw, we can see that adopting source consistency regularization can enrich the final performance.

To summarize, by fusing social networks using deep learning with source confidence and consistency regularization, our FARSEEING model significantly outperforms state-of-the-art approaches in terms of precision, recall and F_1-score. This implies that the data on multiple social networks are complementary and characterize users' volunteerism tendency consistently. This also indicates that the correlation of different social networks with the task of volunteerism tendency prediction cannot be treated equally.

In a real-world scenario, volunteers usually constitute a small portion of social network users. Therefore, we changed the percentage of volunteer samples in our dataset to evaluate the usefulness of our model in a real-world scenario. Figure 2 shows the F_1-score with respect to different positive sample percentages. We can see from the figure that our FARSEEING model still outperforms other approaches. Meanwhile, it is generally stable to different percentages of positive samples which can achieve satisfactory performance with a low positive sample proportion.

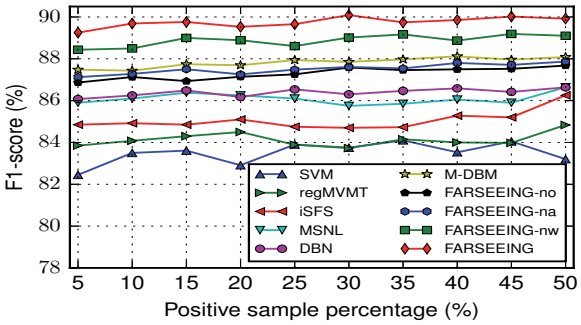

Figure 2: Performance comparison among various models in terms of F_1-score with different positive fractions.

4.4 Source and Feature Performance

To demonstrate the effectiveness of fusing multiple social networks, we conducted experiments over different source combinations and performed significance tests to validate the advantage of fusing multiple social networks.

The experiment on source combinations further demonstrates the benefits of measuring the confidence and consistency of different sources. Table 2 presents the performance

Source	Precision	Recall	F_1-score	p-value
fb	82.365	83.377	82.382	0.0002
in	82.159	83.187	82.367	0.0008
tw	83.612	83.455	83.395	0.0015
fb+in	85.933	88.188	86.619	0.0532
fb+tw	86.864	87.269	86.863	0.0979
tw+in	87.243	85.193	86.941	0.0574
fb+in+tw	**88.069**	**92.197**	**89.921**	-

Table 2: FARSEEING performance (%) of different source combinations. fb: Facebook, in: LinkedIn, tw: Twitter.

of FARSEEING over various source combinations. This table shows that incorporating more sources will achieve better performance, which shows that there is a complementary rather than conflicting relationship among these sources.

Figure 3 presents the performance of all approaches over

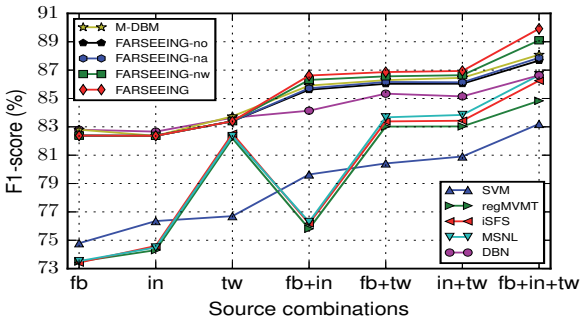

Figure 3: Performance comparison in terms of F_1-score (%) with different social network combinations.

different source combinations. From this figure, we obtain conclusions similar to the ones from Table 2, namely that incorporating multiple sources leads to better performance. Moreover, we observe that multi-source deep learning models outperform linear models over each source combination. Additionally, from single source performance, we observe that linear models achieve significantly better performance on Twitter alone than that on Facebook or LinkedIn (over 8%) and the combination of Facebook and LinkedIn as well, while the performance of deep learning models is relatively stable ($\sim$ 1%). This is mainly because that the discriminative power of data in each source is different, and deep learning models can better capture the complexity of these multi-feature social network data than linear models.

To examine the discriminative power of extracted features, we conducted experiments on different kinds of features using FARSEEING. Table 3 presents the performance

Source & Feature	F_1-score
Facebook	**82.382**
Demographic	68.138
LIWC	78.568
Posting	78.284
User topic	80.228
LinkedIn	**82.367**
Demographic	69.728
Posting	70.995
User topic	81.183
Twitter	**83.395**
Contextual	79.397
LIWC	76.568
Posting	77.559
User topic	80.108
Egocentric	75.304

Feature	F_1-score
Demographic	**76.661**
Linguistic	**88.211**
User topic	86.690
Contextual	79.397
LIWC	79.162
Behavioral	**78.190**
Posting	77.720
Egocentric	75.304

Table 3: FARSEEING performance comparison in terms of F_1-score (%) with different sources and features.

of FARSEEING with different features. From Table 3, we can see that the *user topic* features achieve the best performance in each source, and the *linguistic* features achieve the best performance comparing *demographic* and *behavioral* features. In particular, the *user topic* features achieve significantly better performance against others. This shows that individuals' tendency to volunteerism is reflected in their social contents, especially the topics they discussed on social networks. Meanwhile, the performance of *contextual topic* features means that "birds of a feather flock together", which is to say that contextual information, namely the topics of a given user's followers/followees, reflects this user's tendency to volunteerism as well. The performance of *behavioral* implies that posting patterns and social connections indeed reflect a users personal character and concerns.

5 Related Work

In this paper, we fuse data from multiple social networks with deep learning and cast the problem of volunteerism tendency prediction as a binary classification task. So our work is related to studies of volunteerism, multi-source learning as well as deep learning.

Volunteerism is generally regarded as non-profit activities or services offered by individuals or organizations to benefit the served person or community. A previous work by Penner (Penner 2002) studies the influences of dispositional variables on volunteerism using data from an on-line survey. This research shows that people's personality traits and religiosity have a significant relationship with their volunteer activities, and it presents a theoretical model for the causes of volunteerism and gives explanations from an interactionist perspective.

Multi-source learning has attracted great attention from machine learning community. regMVMT (Zhang and Huan 2012) introduces regularization to measure the consistency of different sources, but neglects source confidence. A later work, iSFS (Xiang et al. 2013) adopts source confidence regularization when fusing data for Alzheimers Disease predic-

tion. This work does not measure the consistency of different sources. The state-of-the-art work, MSNL (Song et al. 2015) was devised for multiple social network learning. This model measures both the confidence and the consistency of different sources, and it also uses volunteerism tendendy prediction as an application. All these approaches adopt linear models to fuse information from multiple sources. However, such linear models may not be powerful enough to capture the complex nature of different social networks.

Deep learning approaches have also been widely used for multi-source learning. Deep denoised auto-encoders were adopted to learn a shared representation of multiple sources (Ngiam et al. 2011). M-DBM (Srivastava and Salakhutdinov 2012) is presented to fuse image and text tag data by learning a joint representation. A hybrid deep learning model, HLDBN (Wang and Wang 2014), is presented to fuse users' rating data and audio features for music recommendation. These deep learning-based approaches can better capture the complex nature of different sources. However, these approaches neither measure the level of confidence in each source nor the consistency among different sources.

6 Conclusion

This paper presents a novel multi-source deep learning model. It fuses social networks using deep learning with source confidence and consistency regularization to better capture the complex nature of social networks. This model is the core component of our framework for multiple social network learning. We evaluate our framework on the application of user volunteerism tendency prediction. By extensive experiments, we demonstrate that our approach is intuitively reasonable and empirically beneficial, compared with other approaches. Currently, we only solve a multi-source mono-task learning problem using the presented framework. In the future, we will extend it to the context of multi-source multi-task learning.

Acknowledgment

This research was supported in part by grants R-252-000-473-133 and R-252-000-473-750 from the National University of Singapore.

References

Abel, F.; Herder, E.; Houben, G.-J.; Henze, N.; and Krause, D. 2013. Cross-system user modeling and personalization on the social web. *UMUAI* 23(2-3):169–209.

Bazelli, B.; Hindle, A.; and Stroulia, E. 2013. On the personality traits of stackoverflow users. In *29th IEEE ICSM*, 460–463.

Bengio, Y. 2009. Learning deep architectures for ai. *Foundations and trends in Machine Learning* 2(1):1–127.

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent dirichlet allocation. *the Journal of machine Learning research* 3:993–1022.

Guo, J.; Xu, G.; Cheng, X.; and Li, H. 2009. Named entity recognition in query. In *32nd ACM SIGIR*, 267–274.

Hinton, G. E., and Salakhutdinov, R. R. 2006. Reducing the dimensionality of data with neural networks. *Science* 313:504–507.

Lee, D. D., and Seung, H. S. 2001. Algorithms for non-negative matrix factorization. In *NIPS*, 556–562.

Li, S.-Y.; Jiang, Y.; and Zhou, Z.-H. 2014. Partial multi-view clustering. In *28th AAAI*, 1968–1974.

Liu, S.; Wang, S.; Zhu, F.; Zhang, J.; and Krishnan, R. 2014. Hydra: Large-scale social identity linkage via heterogeneous behavior modeling. In *2014 ACM SIGMOD*, 51–62.

Markovikj, D.; Gievska, S.; Kosinski, M.; and Stillwell, D. 2013. Mining facebook data for predictive personality modeling. In *7th ICWSM*, 23–26.

Mislove, A.; Marcon, M.; Gummadi, K. P.; Druschel, P.; and Bhattacharjee, B. 2007. Measurement and analysis of online social networks. In *7th ACM SIGCOMM*, 29–42.

Ngiam, J.; Khosla, A.; Kim, M.; Nam, J.; Lee, H.; and Ng, A. Y. 2011. Multimodal deep learning. In *28th ICML*, 689–696.

Nie, L.; Zhang, L.; Yang, Y.; Wang, M.; Hong, R.; and Chua, T.-S. 2015. Beyond doctors: Future health prediction from multimedia and multimodal observations. In *23rd ACM Multimedia*, 591–600.

Pennacchiotti, M., and Popescu, A.-M. 2011. Democrats, republicans and starbucks afficionados: user classification in twitter. In *17th ACM SIGKDD*, 430–438.

Penner, L. A. 2002. Dispositional and organizational influences on sustained volunteerism: An interactionist perspective. *Journal of Social Issues* 58(3):447–467.

Rumelhart, D. E.; Hintont, G. E.; and Williams, R. J. 1986. Learning representations by back-propagating errors. *NATURE* 323:9.

Song, X.; Nie, L.; Zhang, L.; Akbari, M.; and Chua, T.-S. 2015. Multiple social network learning and its application in volunteerism tendency prediction. In *38th ACM SIGIR*, 213–222.

Srivastava, N., and Salakhutdinov, R. R. 2012. Multimodal learning with deep boltzmann machines. In *NIPS*, 2222–2230.

Wang, X., and Wang, Y. 2014. Improving content-based and hybrid music recommendation using deep learning. In *ACM Multimedia*, 627–636.

Wei, X., and Croft, W. B. 2006. Lda-based document models for ad-hoc retrieval. In *29th ACM SIGIR*, 178–185.

Xiang, S.; Yuan, L.; Fan, W.; Wang, Y.; Thompson, P. M.; and Ye, J. 2013. Multi-source learning with block-wise missing data for alzheimer's disease prediction. In *19th ACM SIGKDD*, 185–193.

Zhang, J., and Huan, J. 2012. Inductive multi-task learning with multiple view data. In *18th ACM SIGKDD*, 543–551.

Zhu, X.; Ming, Z.-Y.; Zhu, X.; and Chua, T.-S. 2013. Topic hierarchy construction for the organization of multi-source user generated contents. In *36th ACM SIGIR*, 233–242.