Exploring and Inferring User-User Pseudo-Friendship for Sentiment Analysis with Heterogeneous Networks

Hongbo Deng[1], Jiawei Han[1], Hao Li[2], Heng Ji[2], Hongning Wang[1], and Yue Lu[3]

[1]University of Illinois at Urbana-Champaign
[2]City University of New York
[3]Twitter Inc.

Abstract

With the development of social media and social networks, user-generated content, like forums, blogs and comments, are not only getting richer, but also ubiquitously interconnected with many other objects and entities, forming a heterogeneous information network between them. Sentiment analysis on such kinds of data can no longer ignore the information network, since it carries a lot of rich and valuable information, explicitly or implicitly, where some of them can be observed while others are not. In this paper, we propose a novel information network-based framework which can infer hidden similarity and dissimilarity between users by exploring similar and opposite opinions, so as to improve post-level and user-level sentiment classification in the same time. More specifically, we develop a new *meta path*-based measure for inferring pseudo-friendship as well as dissimilarity between users, and propose a semi-supervised refining model by encoding similarity and dissimilarity from both user-level and post-level relations. We extensively evaluate the proposed approach and compare with several state-of-the-art techniques on two real-world forum datasets. Experimental results show that our proposed model with 10.5% labeled samples can achieve better performance than a traditional supervised model trained on 61.7% data samples.

1   Introduction

Recently the rise of social media and social networks, such as blogs, forums, and Twitter, has fueled the online space with lots of reviews for products, ratings for movies, and comments on current events or politics. Over the years, sentiment has been a widely used measure of how customers view a company's products and services, and how people think about current events and politics. Sentiment analysis refers to the task of determining opinions, judgments, and other information related to the attitudes of a speaker or a writer with respect to some topics or the overall contextual polarity
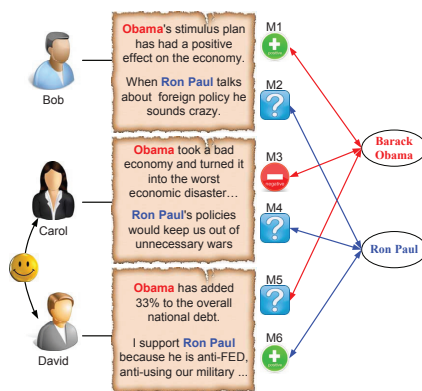

Figure 1: An example data with User-Posts-Entity heterogeneous networks and initial sentiment scores.

of a document. Based on such information, companies have the opportunity to examine what current and potential customers are saying about their products and services without costly and time-consuming surveys. Similarly, political organizations and candidates might be able to determine what issues the public is most interested in, as well as where they stand on those issues. Therefore, it is very important and highly desirable to conduct sentiment analysis automatically, which pertains to products, companies, and commercial and political entities.

Sentiment analysis has been studied with various approaches, such as lexicon-based methods [1] and learning-based methods [2], among which most of them are based on text content alone. But in social media textual documents are ubiquitously interconnected with many other entities, such as users and topics, in many ways, forming a heterogeneous network between documents and other entities. As shown in Figure 1, three users wrote 6 posts that were associated with 2 presidential candidates. In addition, we observe users Caral and David are friends. Such a heterogeneous network carries

rich semantic and valuable information, which should be utilized to improve sentiment analysis. The principles of homophily [3] and "birds of a feature" [4] show the generation of a link depends on its context, and similar contexts could potentially lead to similar links, which suggests that users that are connected (become friends) may tend to hold similar opinions. For example, in a political forum, social network is formed between users due to homophily or influence, where friends are likely to share similar opinions to a certain topic. Therefore, it is essential to utilize the heterogeneous network, especially user-user relations, to enhance sentiment analysis.

To address the problem, we propose a novel information network-based framework by inferring pseudo-friendship between users and exploring post-post relations, so as to improve sentiment analysis. Given a number of labeled data with sentiment scores and observed friendship, we develop a semi-supervised refining model with user regularization (UserReg) to propagate the sentiment scores from labeled data to unlabeled data. It is worth noting that we not only consider the similarity but also the dissimilarity in the information network. With the sentiment scores, we develop a novel meta path-based measure for estimating the similarity and dissimilarity between entities, such as *pseudo-friends* and *pseudo-foes* between users. Consequently, the inferred pseudo-friendship along with the observed friendship may propagate the sentiment scores on the information network more effectively and consistently. Moreover, we also incorporate the post-post relations along with user-user relations to refine sentiment scores in a unified framework.

The basic idea of our model is that friends are more likely to hold similar opinions, while foes are more likely to have conflicting opinions with respect to a certain entity or topic. On the other hand, based on a user's sentiment scores on different topics, we can infer the similarity and dissimilarity (i.e., *pseudo-friend* and *pseudo-foe* relationship) between users. Furthermore, based on the inferred friendship, we may jointly improve post-level and user-level sentiment analysis by considering the global consistency on the heterogeneous networks. To illustrate our methodology, we apply the proposed model UserReg to sentiment classification with two real-world datasets. For the sentiment classification task, we compare with several different state-of-the-art models, including lexison-based models and supervised models. It is shown that our model with 100 (10.5%) labeled data can perform much better than SVM-based supervised model with 600 (61.7%) training data on the political forum dataset.

The rest of this paper is organized as follows. We first introduce the preliminaries in Section 2. In Section

Figure 2: A simple forum network schema.

3, we systematically present and develop the information network-based framework. In Section 4, we conduct extensive experiments on sentiment classification. Finally, we review some related work in Section 5, and present our conclusions and future work in Section 6.

2 Preliminaries

More formally, a forum consists of a set of threads, along with 1) a set of users $U = \{u_1, u_2, ..., u_m\}$; 2) a set of documents $D = \{d_1, d_2, ..., d_n\}$ written by users, where each d_i is a post with textual content; and 3) a set of topics or issues T associated with them (we assume T is given which can be obtained from the forum). Let $F = \{f_1, f_2, ..., f_n\}$ denote the sentiment scores we want to identify for the set of posts D with respect to different topics. As shown in Figure 2, a heterogeneous information network is formed between these objects. For example, a user links with their written posts, and each post links with some specific topics.

We study sentiment analysis of forum discussions in this paper, and formulate the problem as a semi-supervised sentiment analysis: Given a forum information network with the relations, let the first $l \leq n$ posts be labeled with sentiments $F_l = \{y_1, y_2, ..., y_l\} \in \{+1, -1\}$, the task is to predict the polarity of the remaining posts $F_u = \{f_{l+1}, ..., f_n\}$ which are unlabeled. For each unlabeled post, we can obtain an initial sentiment score y_i^0 ($l + 1 \leq i \leq n$), whose value is the prediction of a separate method, such as lexicon-based method in our experiments. We aim to utilize the heterogeneous information network to better understand post-level sentiment as well as user-level opinions with respect to different topics.

3 Information Network-Enhanced Framework

In this section, we propose a general framework to explore not only the post-level but also user-level relations on a heterogeneous information network.

3.1 Exploring User-User Relations
We discuss how to explore user-user relations so as to enhance post-level sentiment analysis.

Basic Principles: Our approach is based on some basic principles. First, similar users (e.g., friends) are prone to have similar and consistent opinions for a certain topic due to the principle of homophily [3]. In contrast, two users tend to form a friendship if
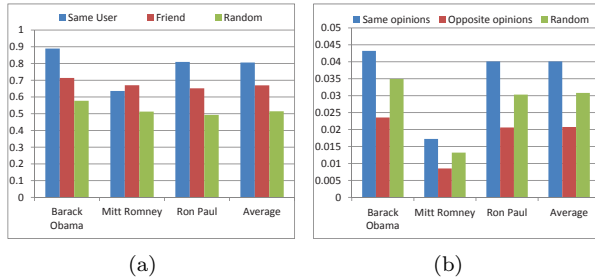
(a)   (b)

Figure 3: Two statistics on a political forum dataset:
(a) Probability of two posts having the same senti-
ment label with respect to different candidates/topics,
conditioned on three types of relationships: two posts
given by the same user, two friends, or two random
users, respectively; (b) Probability that two users are
friends/connected, conditioned on any two posts from
those two users have the same opinions, opposite opin-
ions, or random opinions.

they share a lot of similar opinions. Inspired by the
intuition, we can infer the pseudo-friendship based on
their sentiment scores if the explicit friendship is not
available.

Second, people tend to be foes if they hold conflict-
ing or opposite opinions on many topics. On the other
hand, the opposite opinions and dissimilarities between
posts may indicate the *pseudo-foe* relationship between
users. Such a pseudo-foe relationship, which is usually
hidden, is as important as the friendship for improving
sentiment analysis.

In order to validate these principles, we show two
statistics on a political forum dataset (Details about the
dataset are shown in Section 4.1). Figure 3(a) shows
that the probability of two posts given by the same
user or friends sharing the same sentiment on a topic
is much higher than random. Figure 3(b) shows that it
is more likely for users to be friends/connected if they
share the same opinion on a certain topic than random
(comparing blue and green bars); on the contrary, it is
more unlikely for users to be friends (i.e., users tend
to be foes) if they have opposite opinions on a certain
topic than random (comparing red and green bars).
As a whole, all the observations support our principles
that the sentiment labels and social networks as well as
heterogeneous networks are correlated, at least in the
political domain.

Inferring User-User Relations: Due to privacy and
other issues, it is usually unavailable to obtain the
explicit friendship between users. Therefore, it becomes
very important to infer user-user relations, i.e., pseudo-
friendship. According to previous studies [5], two users
are similar if they are strongly connected. However,

when looking into the sentiment of user's posts, highly
connected users may be very dissimilar in terms of
sentiment context if they hold opposite opinions rather
than similar opinions. Therefore, it is more reasonable
to take the sentimental context into account, as well as
the heterogeneous networks between users and posts.
Intuitively, two users are likely to be similar if they
are connected and hold consistent opinions on different
topics. In contrast, two users are likely to be dissimilar
if they are connected but hold opposite opinions.

To infer the pseudo-friendship between users, we
introduce a similarity function $Sim(u_i, u_j)$ with a real
value ranging from -1 to +1. A score $Sim(u_i, u_j)$
close to 1 indicates these two users are *pseudo-friends*
who have similar and consistent opinions, while a score
$Sim(u_i, u_j)$ close to -1 indicates these two users are
pseudo-foes who hold conflicting and opposite opinions
on many topics. Moreover, a score close to 0 indicates
that two users do not associate with each other, or they
share both consistent and conflicting opinions.

Before we introduce the definition of the dissim-
ilarity, let us define the *signed connectivity* of two
posts, which will be used to estimate the similarity
and dissimilarity between users. Two posts d_i and d_j
on the same topic may be consistent or conflict with
each other. A function $Sc(d_i, d_j)$ is defined to indicate
the degree of consistency or confliction between them
$(-1 \leq Sc(d_i, d_j) \leq 1)$: $Sc(d_i, d_j) = f(d_i) \cdot f(d_j)$, where
$f(d_i)$ and $f(d_j)$ are current sentiment scores of posts.
In the semi-supervised setting, we set $f(d_i) = y_i$ for la-
beled data $(1 \leq i \leq l)$, and $f(d_i) = y_i^0$ for unlabeled
data $(l+1 \leq i \leq n)$. For reliable inference, only labeled
data and highly confident data are used, while setting
some untrustworthy data $f(d_i) = 0$ if $-0.1 < y_i^0 < 0.1$.
Thus we obtain that $Sc(d_i, d_j) > 0$ if d_i and d_j are both
positive or negative, and $Sc(d_i, d_j) < 0$ if d_i and d_j hold
conflicting or opposite opinions.

In order to capture both sentimental context and
heterogeneous networks, we define the *meta path*-based
similarity between two users. Here *meta path* [5]
captures a sequence of relations defined between users,
which depends on the heterogeneous networks. For
example, two users u_i and u_j can be connected via
the path "user-post-topic-post-user" (i.e., U-P-T-P-U)
if both u_i and u_j have written at least one post about
a same topic T. Given a meta path (e.g., $\mathcal{P}$ =U-P-T-P-
U), we can infer the similarity and dissimilarity between
users with the following definition:

$$Sim(u_i, u_j) = \frac{2 \sum_{d_k \in D_{u_i}, d_l \in D_{u_j}} \left(1_{(p_{u_i \rightsquigarrow u_j} \in \mathcal{P})} \cdot Sc(d_k, d_l) \right)}{|\{p_{u_i \rightsquigarrow u_i} \in \mathcal{P}\}| + |\{p_{u_j \rightsquigarrow u_j} \in \mathcal{P}\}|},$$

where $p_{u_i \rightsquigarrow u_j}$ is a path instance between u_i and u_j that
follows the defined meta path $\mathcal{P}$, and D_{u_i} and D_{u_j} are
the set of posts written by user u_i and u_j, respectively.

This formulation is defined in terms of two parts: (1) (numerator) - the *signed connectivity* which is captured by the number of paths connecting two users along with their consistent/conflicting opinions on the topics, and (2) (denominator) - the *visibility* which is defined by the number of path instances between themselves.

The underlying intuition of this formulation is that two users are more likely to be similar, i.e., $Sim(u_i, u_j)$ close to 1, if they share similar opinions (both positive or negative) with respect to different topics, and we denote them as *pseudo-friends*. On the contrary, two users are more likely to be dissimilar, i.e., $Sim(u_i, u_j)$ close to -1, if they have conflicting or opposite opinions on many topics, and we denote them as *pseudo-foes*. It is obvious that the similarity is symmetric, i.e., $Sim(u_i, u_j) = Sim(u_j, u_i)$, and we set $Sim(u_i, u_i) = 1$ for all users.

User Regularization: Since different posts may be associated with different topics, we denote two posts as on-topic posts if they discuss the same topic, and off-topic if they are about different topics. Based on the basic principles, two *on-topic* posts written by *pseudo-friends* are more likely to be consistent. To capture this assumption, we define to minimize the following user regularization term so as to constrain sentiment scores between posts,

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \delta_{d_i,d_j}^T U_{d_i,d_j} (f_i - f_j)^2, \text{ if } U_{d_i,d_j} > 0,$$

where U is a matrix with entry $U_{d_i,d_j} = Sim(u_{d_i}, u_{d_j})$, u_{d_i} is the user that wrote d_i, and δ_{d_i,d_j}^T is the delta function that equals 1 if d_i and d_j are on-topic posts, and 0 otherwise. On the other hand, two *on-topic* posts written by two *pseudo-foes* are more likely to be dissimilar. Correspondingly, we define to minimize the following user regularization term

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \delta_{d_i,d_j}^T (-U_{d_i,d_j})(f_i + f_j)^2, \text{ if } U_{d_i,d_j} < 0,$$

which indicates f_i and f_j should have opposite sentiment scores or both close to zero. By combining the previous two terms together, we would like to minimize a new penalty term

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \delta_{d_i,d_j}^T |U_{d_i,d_j}| (f_i - s_{ij}^U f_j)^2,$$

where $s_{ij}^U = 1$ if $U_{d_i,d_j} \geq 0$, and $s_{ij}^U = -1$ if $U_{d_i,d_j} < 0$. If the explicit friendship is available, we can simply aggregate them together $U = U^{Infer} + U^{Explicit}$.

3.2 Incorporating Post-Post Relations Now we discuss how to explore relations between posts based on multiple evidences. To explore the most useful relations between posts, we will only explore relations between on-topic posts including cross-thread posts.

Similarity Relation: Each post can be represented as a feature vector, e.g., bag of words, and a similarity ($Sim(d_i, d_j) \geq 0$) between any two posts d_i and d_j can be calculated given a similarity measure. Generally, a large similarity implies that the two posts tend to express the same sentiment [6, 7]. Although counterexamples are easy to construct, it has been found that exploiting the similarity between posts can boost the performance of sentiment classification [6, 7].

Given a similarity measure, we can construct a kNN graph, where each post is connected to its k nearest on-topic posts. Let M^{Sim} be a similarity matrix with $M_{ij}^{Sim} = \begin{cases} Sim(d_i, d_j), & j \in kNN(i) \\ 0, & \text{Otherwise} \end{cases}$. We experiment with word-vector cosine similarity. To encode the assumption that two posts with a large similarity share similar sentiment, we define to minimize the following loss function

$$\sum_{i=1}^{n} \sum_{j \in kNN(i)} M_{ij}^{Sim} (f_i - f_j)^2.$$

Reply-to relation: A forum discussion on a topic typically consists of many seed posts, and a large number of posts that are responses to a seed post or responses to responses. Typically there are explicit quotations from earlier posts for responses. Based on such quotations, we can extract "Reply-to" relations between posts, and create a link between these two posts. An interesting characteristic of many forum discussions, especially political discussion board, is that users tend to quote posts by users with different views [8, 9]. For example, users often debate a controversial topic, quoting and disputing each others' previous claims. To treat the "Reply-to" relation between two posts, we can simply assume these two posts have opposite opinions with respect to a topic. Recently, some advanced methods [10, 11] were proposed to discover supporting and opposite "Reply-to" relations between posts.

Let R be a matrix to denote "Reply-to" graph with entries $R_{ij} = 1$ if d_i replies to d_j with supporting opinions, $R_{ij} = -1$ if d_i replies to d_j with opposite opinions, and $R_{ij} = 0$ otherwise. In order to handle both positive and negative relations, we choose a suitable penalty term as follows:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} |R_{ij}| (f_i - s_{ij}^R f_j)^2,$$

where $s_{ij}^R = 1$ if $R_{ij} \geq 0$, and $s_{ij}^R = -1$ if $R_{ij} < 0$. In order to minimize the term, f_i and f_j should have similar sentiment scores when $R_{ij} > 0$ (i.e., d_i and d_j are mutual supportive). When $R_{ij} < 0$ (i.e., d_i and d_j are mutual exclusive/unsupportive), d_i and d_j should have opposite scores or both close to zero.

User Consistency: Generally users may publish many posts on a topic in different threads. Different on-topic posts from the same user tend to express consistent opinion. More specifically, suppose we know one post from u_i shows strong positive opinion for a given topic, then all the other posts written by u_i on that topic would be likely to share a similar opinion. Following this assumption, we can encode it as a matrix A with $A_{ij} = 1$ iff on-topic posts d_i and d_j are written by the same user (i.e., $user(d_i) = user(d_j)$). To force consistent sentiment scores among on-topic posts from the same user, we define to minimize the following cost function

$$\sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} (f_i - f_j)^2 .$$

3.3 Regularization Framework To put all the information together, we have the following objective function:

$$
\begin{aligned}
\mathbf{f}^* \;=\; \arg\min_{\mathbf{f}} \Bigg\{ & \mu \sum_{i=l+1}^{n} (f_i - y_i^0)^2 \\
+\; & \frac{\beta}{2} \sum_{i=l+1}^{n} \sum_{j=1}^{n} \delta_{d_i,d_j}^T |U_{d_i,d_j}| (f_i - s_{ij}^U f_j)^2 \\
+\; & \frac{(1-\beta)}{2} \sum_{i=l+1}^{n} \sum_{j=1}^{n} |M_{ij}^{Sim} + R_{ij} + A_{ij}| (f_i - s_{ij}^R f_j)^2 \Bigg\}
\end{aligned}
$$

subject to
$$f_i = y_i, \; 1 \leq i \leq l,$$

where μ and β are the weights to trade off three different components: The first component enforces the refined sentiment scores to fit the initial sentiment scores; the second component enables the user-level constraints based on inferred pseudo-friendship; and the third component incorporates the post-level supporting and opposite relations. The key idea is to refine sentiment scores by encoding similarity and dissimilarity from both user-level and post-level relations. The above optimization problem can be solved directly as the objective function is convex [9], and a closed-form solution can be derived.

4 Experimental Evaluation

4.1 Data Collection We created our data sets from two online forums. The first data set is a political

Table 1: Basic statistics of Data sets

PF1901: Political Forum (with explicit user-user friendship)

Candidates	#Post	#Reply	#User	#Pos/#Neg
Barack Obama	681	188	167	93/220
Mitt Romney	335	118	92	50/75
Ron Paul	885	258	153	276/258

MF1560: Military Forum (without explicit user-user friendship)

Topics	#Post	#Reply	#User	#Pos/#Neg
abortion	297	41	93	110/70
healthcare reform	323	62	84	90/121
illegal immigrants	307	49	105	54/194
iraq war	324	51	98	124/136
president obama	309	42	96	59/97

forum[1]. We crawled the most recent posts from March 2011 to April 2012, containing 608 threads and 31,991 posts. In order to make it easier for the human judges to annotate, we further narrowed down to three popular US presidential candidates, and applied information extraction method to extract relevant posts. The basic statistics are shown in Table 1. There are totally 1,901 labeled posts written by 232 unique users. Moreover, we manually labeled 419 positive and 553 negative posts with respect to the associated candidates, the rest posts are either neutral or not sure about their polarity. *For these 232 users, we crawled their profiles and observed a total of 782 friendship edges, which indicates that each user has 3.37 friends on average.*

The second data set is crawled from a military forum[2], containing 43,483 threads and 1,343,427 posts. In order to make it easier for the human judges to annotate[3], we further narrowed down to five popular and controversial topics, and applied information retrieval method to retrieve the top five most relevant threads for each topics. The basic statics are shown in Table 1. There are totally 1,560 labeled posts written by 320 unique users. We manually labeled 437 positive and 618 negative posts with respect to the topics, the rest posts are either neutral or not sure about their polarity. Unlike the above political forum, there is no explicit friendship between users listed in this forum.

4.2 Experimental Setup We evaluate our proposed model and compare with several state-of-the-art methods as follows:

SentiWordNet We tag each post by taking the average SentiWordNet [1] score of words. It represents an unsupervised sentiment analysis method which

---

[1] http://www.politicalforum.com/elections-campaigns/
[2] http://forums.military.com/
[3] Please refer to [11] for more details about human annotation.

only relies on the text. Note the output is used as initial sentiment scores for test (unlabeled) data in the semi-supervised learning methods.

SSL+WV This method is proposed for semi-supervised sentiment classification in [7]. They create a similarity graph on both labeled and unlabeled posts, and the graph is formed based on the word- or sentence-level similarity. Here we choose the same similarity measure as our proposed method.

SSL+Dissim The authors [9] further improved the above method by considering both similarity and dissimilarity on labeled and unlabeled posts. The dissimilarity edge is created between two posts if they have exhibited the "ReplyTo" relation.

LP LP is proposed in [11] for analyzing agree/disagree relations between posts. Then the post-post relations are used in a linear programming framework for sentiment analysis (while they did not consider user-user relations). The method is defined in the unsupervised setting, and cannot be directly applied in semi-supervised setting. We develop LP+ for comparison in the semi-supervised setting. Note this method uses all the three post-post relations, but does not use user-user friendship.

UserReg This is our proposed method by exploring and utilizing both post-post relations and user-user relations. It is the first method to explore the inferred user-user relations[4] for enhancing sentiment classification.

In the semi-supervised setting, we randomly choose a small number of labeled data with equal size of positive and negative points. In order to randomize the experiments and make the comparison fair, we conduct the evaluation with the size of labeled data ranging from 50 to 250, choose the same random initializations for different models, and use the same similarity measure and "ReplyTo" relation. For each method, 10 test runs were conducted, and the final performance score represents the average result across 10 trials. To quantitatively compare with these methods, we use several popular metrics to evaluate the sentiment classification, including accuracy, precision, recall[5].

4.3 Evaluation of Post-Level Classification We first evaluate the accuracy of the post-level sentiment classification. The ground truth is derived from the positive and negative posts based on post level judgment, and ignore the neutral and "Not Sure" cases.

Table 2: Accuracy of Post-Level Classification -*PF1901*

Method	Accuracy	Precision	Recall
SentiWordNet	0.4444	0.4573	0.6134
SVM	0.6124	0.556	0.6548
Semi-supervised setting with 100 (10.5%) labels			
SSL+WV	0.5325	0.4619	0.6144
SSL+Dissim	0.5431	0.4711	0.6084
LP+	0.6607	0.5879	0.6707
UserReg	**0.6903**	**0.6125**	**0.7344**
UserReg*	**0.7275**	**0.6578**	**0.7493**

UserReg* uses both the inferred user-user relation and the explicit friendship, while UserReg only use the inferred user-user relation.

4.3.1 Comparison with Different Models In Table 2, we compare with different models on the data set *PF1901*. Top part of the table shows the performance of a Lexicon-based unsupervised method SentiWordNet [1] and a SVM-based supervised method [12]. The accuracy of SentiWordNet [1] without using any supervision is 44.4%, indicating the poor performance of unsupervised method. Support vector machines (SVMs) [12] have been shown to be highly effective at traditional text categorization as well as sentiment analysis. We used LIBSVM [13] for training and testing. The model was trained on 300 positive and 300 negative samples, and tested on the remaining 372 samples. We tried different features including N-gram feature, lexicon feature, POS feature, punctuation feature, and so no. The best classification accuracy we obtained is 61.2%. From the results, although supervised model achieves better performance than Lexicon-based unsupervised model, it is still much lower than the results reported in previous studies on movie reviews [12][6]. The reason is that the sentiment analysis task on forum data is fairly difficult and complicated. For example, "NONE of the GOP candidates have a significant advantage on national polls against Obama". This post is supporting "Obama" by arguing with GOP supporters. Within the context, "significant" and "advantage" are positive words while "against" and "NONE" are negative words, so it is not a trivial work to identify the correct sentiment label of the sentence by using lexicon words. If the user of this post wrote some other posts to support "Obama", it becomes easier to classify it correctly by combining such information with previous posts. Our proposed model can connect complicated posts with straightforward ones through heterogeneous networks, which can enhance the performance significantly.

---

[4]Previous studies only consider explicit friendship.

[5]http://en.wikipedia.org/wiki/Precision_and_recall

---

[6]Note that the accuracy of SentiWordNet is 72% for movie reviews, but it is only 44.4% for this forum data

Table 3: Accuracy of Post-Level Classification - *MF1560*

Method	Accuracy	Precision	Recall
SentiWordNet	0.4664	0.4275	0.5469
LP	0.4844	0.4271	0.5629
Semi-supervised setting with 100 (9.5%) labels			
SSL+WV	0.5086	0.4272	0.6191
SSL+Dissim	0.5044	0.4217	0.5956
LP+	0.5353	0.4466	0.6121
UserReg	**0.556**	**0.4659**	**0.6496**

In the low part of Table 2, we compare three base-lines with our UserReg method in the semi-supervised setting. For each method, we use 100 (10.5%) posts as labeled data and test on the rest posts. The matrix M^{Sim} is constructed by building a kNN graph ($k = 20$) between posts based on word-vector (TF-IDF) cosine similarity. For "Reply-to" relation, we follow [9] by simply setting $R_{ij} = -1$ if d_i replies to d_j. As shown in the table, it is obvious that 1) SSL+WV and SSL+Dissim perform slightly better than SentiWordNet, indicating that only considering similarity and reply-to relations among posts is not effective enough; 2) LP+ provides better accuracy than the previous two baselines, suggesting that user consistency is a reasonable relation; 3) our proposed methods, both UserReg and UserReg*, perform much better than all the other methods with a statistically significant improvement ($p < 0.05$) in all measures. This is because we further explore and infer pseudo-friends and pseudo-foes among users, and our model can benefit from the inferred pseudo-friendship as well as the explicit observed friendship.

In Table 3, we compare with different methods on another dataset *MF1560*. In order to fairly compare with the unsupervised method LP proposed in [11], we use the same topic-based similarity measure and post-post relations discovered in [11] (i.e., set M^{Sim} and R to be T^{agr} and $(R^{agr} - R^{dis})$, respectively). The top part of this table shows the performance of two unsupervised methods SentiWordNet and LP [11], and LP performs better than SentiWordNet because it considers three relations between posts. The low part of the table shows the comparison of different methods in the semi-supervised setting with 100 (9.5%) labels. Similarly, SSL+WV and SSL+Dissim perform slightly better than SentiWordNet, and LP+ achieves better results by considering user consistency. As expected, our proposed method UserReg achieves the best performance. The overall results on the dataset *MF1560* is worse than the results on *PF1901*, and one major reason is the heterogeneous networks between users, posts and topics on *MF1560* are more sparse which somehow limits the advantage of our method.
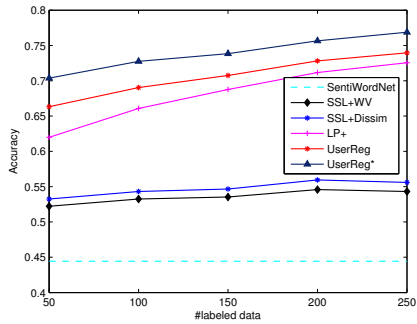
(a) PF1901
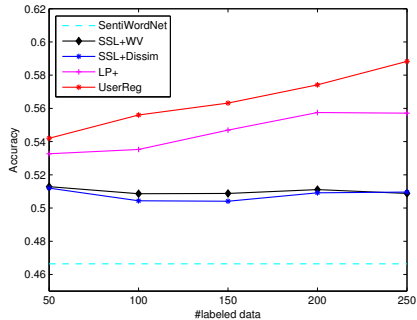
(b) MF1560

Figure 4: Varying the number of labeled data.

Varying Labeled Set: In previous experiments, we fixed $|L| = 100$. Now we systematically vary labeled set size $|L| \in \{50, 100, 150, 200, 250\}$ to investigate the effect of semi-supervised learning. Figure 4 shows the results on datasets *PF1901* and *MF1560*. Generally, the models perform better with the increase of the size of labeled data, especially for LP+ and our model UserReg. In the dataset *PF1901*, SSL+WV and SSL+Dissim approaches perform slightly better with larger labeled set sizes. However, in the dataset *MF1560*, these two approaches do not perform better with large labeled set sizes. We believe the reason is that topic-based similarity measure is too rough and not accurate enough to capture the semantics. Final, we can observe UserReg and UserReg* achieve much higher performance than other methods in all the settings.

4.3.2 Parameter Analysis There are two parameters, decay factor β and regularization parameter μ, in our method. Previous experimental results were obtained by empirically setting $\beta = 0.5$ and $\mu = 0.1$. The optimal parameters can be obtained by cross validation. In this subsection, we study and evaluate the effect of parameters β and μ based on the dataset *PF1901* using 100 labeled data (The results on *MF1560* illustrate similar results).
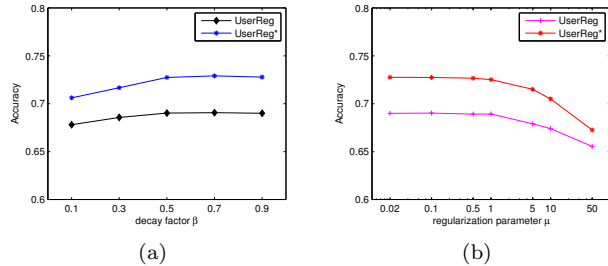
Figure 5: The effect of varying parameters β and μ with 100 labels.

Table 4: Accuracy of User Opinion Prediction

Method	Accuracy (For)	Accuracy (Against)	Accuracy (For+Against)
SentiWordNet	0.6250	0.5250	0.5896
LP	0.6429	0.5513	0.5896
Semi-supervised models with 100 (9.5%) labels			
SSL+WV	0.7143	0.4746	0.5746
SSL+Dissim	0.6786	0.4615	0.5522
LP+	0.7143	0.5385	0.6119
UserReg	**0.7857**	**0.6410**	**0.7015**

Figure 5(a) shows the performance of UserReg and UserReg* by varying the decay factor β from 0.1 to 0.9 (and fixing $\mu = 0.1$). As mentioned before, β is used to control the balance between user-user relations and post-post relations. We can see that the performance is relatively stable, and it is better to set β between 0.5 and 0.75.

Figure 5(b) shows the performance of our models by varying the regularization parameter μ from 0.02 to 50 (and fixing $\beta = 0.5$). As mentioned before, the parameter μ is used to control the trade-off between the consistency of sentiment scores on the information network and the initial scores. When μ is set to 1 or less than 1, the model tends to trust more on the information network (e.g., explicit or inferred friendship and post-post relations). With the increase of μ, the model tends to trust more on the initial scores. As we can see, when μ is set to 5 and 10, the final results start to decrease. It is clear that the performance of our model is very stable by setting μ to be 1 or less than 1, which confirms the effectiveness of our method.

4.4 Evaluation of User-Level Sentiment Beyond evaluating the quality of post-level sentiment polarity, it would be more meaningful to assess the model's capability of identifying user-level sentiment polarity, which can be further used as higher level summarization of user's opinion towards specific topics.

Following [11]'s setting, we only evaluate on a subset of uses, who possess strong opinions. In particular, we selected the users in the MF1560 data set with at least two posts and their aggregated ground-truth opinion score $s > 0.5$. This results in 57 users with strong positive opinions and 78 users with strong negative opinions for the selected 5 topics. To compare the reported results in [11], we also evaluated prediction accuracy for supporting users (i.e., "*for*"), against users (i.e., "*against*") and both of them. The experiment results are listed in Table 4

Comparing to the unsupervised methods list on the top half of Table 4, all the semi-supervised models

achieved promising accuracy improvement in the "*for*" category, especially for our proposed UserReg method (accuracy improvement over 22.2% over the unsupervised LP method). And for the "against" category, our method also achieved the largest improvement over other semi-supervised methods (accuracy improved over 19.0% over the runner-up LP+ method). These improvements confirm the usefulness of prorogating information through the heterogenous network for better classification performance.

5 Related Work

Traditional sentiment analysis has mostly focused on formal genres such as newswire, many levels of granularity such as document level [2], sentence level [14] and phrase level [15]. During recent years informal texts in forum discussions and microblogging have become one of the major forms of online communication, enabling the sharing of real-time updates by both individuals and organizations. Some recent studies have moved to informal genres such as tweets [16, 17] and forum discussions [11]. However, most of them were based on text content alone. The debating nature and informality of forum posts have brought great additional challenges to sentiment analysis. It is thus impractical to use standard supervised machine learning techniques alone which are dependent on annotated training examples.

Several studies [18, 19] have been proposed to improve sentiment analysis with user-user relationships [18] and other connections [19]. In [19], the authors exploited knowledge about word types encoded in a lexicon, in combination with the Twitter follower graph for label propagation to improve sentiment analysis for tweets. But this model treats different types of objects in a similar way, which does not fully explore the heterogeneous information network. In [20], the authors exploited the retweet networks to classify the users into coarse-grained sentiments (left and right), without analyzing the content of messages. Tan et al. [18] proposed another approach to predict user-level sentiment

with extracted social networks from followers and "@" mentions. But this model cannot handle the sentiment in the post level, and it heavily relies on the explicit observed friendship. Our proposed model is different, which can not only predict both post-level and user-level sentiment in a unified way, but also infer the hidden friendship as well as dissimilarity between users.

This work is also related to graph-based semi-supervised learning [21, 7, 9], which usually assumes label smoothness over the graph. Glodberg and Zhu [7] applied a graph-based semi-supervised learning algorithm to address sentiment analysis, by constructing similarity graphs to ensure that similar reviews receive similar labels. Our work is different from theirs, as we explore not only the relations between posts, but also the relations between users, including explicitly observed and implicitly inferred friendship.

6   Conclusions and Future Work

We have presented a novel information network-based framework for enhancing sentiment analysis. Consequently, we have investigated a new meta path-based measure which can estimate not only the similarity but also the dissimilarity between users. Furthermore, a semi-supervised refining model with user regularization was developed to estimate the sentiment scores by exploring both user-level and post-level relations. Experimental results on the sentiment classification task show the effectiveness and correctness of our proposed approach, and the improvement of our approach is promising. For future work, we plan to extend our framework to explore the correlation between different topics, and it would be interesting to take more information such as time into consideration.

7   Acknowledgments

References

[1] S. Baccianella, A. Esuli, and F. Sebastiani, "Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining," in *LREC*, 2010.

[2] B. Pang and L. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," in *ACL*, 2004, pp. 271–278.

[3] P. Lazarsfeld and R. Merton, "Friendship as a social process: A substantive and methodological analysis," *Freedom and control in modern society*, vol. 18, pp. 18–66, 1954.

[4] M. McPherson, L. Smith-Lovin, and J. Cook, "Birds of a feather: Homophily in social networks," *Annual review of sociology*, pp. 415–444, 2001.

[5] Y. Sun, J. Han, X. Yan, P. S. Yu, and T. Wu, "Pathsim: Meta path-based top-k similarity search in heterogeneous information networks," *PVLDB*, vol. 4, no. 11, pp. 992–1003, 2011.

[6] B. Pang and L. Lee, "Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales," in *ACL*, vol. 43, no. 1, 2005, p. 115.

[7] A. Goldberg and X. Zhu, "Seeing stars when there aren't many stars: graph-based semi-supervised learning for sentiment categorization," in *Proceedings of the First Workshop on Graph Based Methods for Natural Language Processing*, 2006, pp. 45–52.

[8] R. Agrawal, S. Rajagopalan, R. Srikant, and Y. Xu, "Mining newsgroups using networks arising from social behavior," in *WWW*. ACM, 2003, pp. 529–535.

[9] A. B. Goldberg, X. Zhu, and S. Wright, "Dissimilarity in graph-based semi-supervised classification," in *AISTATS*, 2007.

[10] A. Murakami and R. Raymond, "Support or oppose?: classifying positions in online debates from reply activities and opinion expressions," in *COLING*, 2010, pp. 869–875.

[11] Y. Lu, H. Wang, C. Zhai, , and D. Roth, "Unsupervised discovery of opposing opinion networks from forum discussions," in *CIKM*, 2012.

[12] P. D. Turney, "Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews," in *ACL*, 2002, pp. 417–424.

[13] C. Chang and C. Lin, "Libsvm: a library for support vector machines," *TIST*, vol. 2, no. 3, p. 27, 2011.

[14] S.-M. Kim and E. Hovy, "Determining the sentiment of opinions," in *COLING*, 2004.

[15] A. Agarwal, F. Biadsy, and K. Mckeown, "Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams," in *EACL*, 2009.

[16] A. Bermingham and A. Smeaton, "Classifying sentiment in microblogs: Is brevity an advantage?" in *CIKM*, 2010.

[17] A. Tumasjan, T. Sprenger, P. Sandner, and I. Welpe, "Predicting elections with twitter: What 140 characters reveal about political sentiment," in *ICWSM*, 2010.

[18] C. Tan, L. Lee, J. Tang, L. Jiang, M. Zhou, and P. Li, "User-level sentiment analysis incorporating social networks," in *KDD*, 2011, pp. 1397–1405.

[19] M. Speriosu, N. Sudan, S. Upadhyay, and J. Baldridge, "Twitter polarity classification with label propagation over lexical links and the follower graph," in *EMNLP*, 2011, pp. 53–63.

[20] M. D. Conover, J. Ratkiewicz, M. Francisco, B. Goncalves, A. Flammini, and F. Menczer, "Political polarization on twitter," in *ICWSM*, 2011.

[21] X. Zhu, Z. Ghahramani, and J. D. Lafferty, "Semi-supervised learning using gaussian fields and harmonic functions," in *ICML*, 2003, pp. 912–919.