Yang W, Shen GW, Wang W *et al.* Anomaly detection in microblogging via co-clustering. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 30(5): 1097–1108 Sept. 2015. DOI 10.1007/s11390-015-1585-3

Anomaly Detection in Microblogging via Co-Clustering

Wu Yang (杨 武), *Senior Member, CCF, Member, ACM*
Guo-Wei Shen* (申国伟), *Student Member, CCF, Member, ACM*, Wei Wang (王 巍), Liang-Yi Gong (宫良一)
Miao Yu (于 淼), and Guo-Zhong Dong (董国忠)

Information Security Research Center, Harbin Engineering University, Harbin 150001, China

E-mail: {yangwu, shenguowei, w_wei, gongliangyi, yumiao, dongguozhong}@hrbeu.edu.cn

Received November 17, 2014; revised July 12, 2015.

Abstract Traditional anomaly detection on microblogging mostly focuses on individual anomalous users or messages. Since anomalous users employ advanced intelligent means, the anomaly detection is greatly poor in performance. In this paper, we propose an innovative framework of anomaly detection based on bipartite graph and co-clustering. A bipartite graph between users and messages is built to model the homogeneous and heterogeneous interactions. The proposed co-clustering algorithm based on nonnegative matrix tri-factorization can detect anomalous users and messages simultaneously. The homogeneous relations modeled by the bipartite graph are used as constraints to improve the accuracy of the co-clustering algorithm. Experimental results show that the proposed scheme can detect individual and group anomalies with high accuracy on a Sina Weibo dataset.

Keywords microblogging, anomaly detection, nonnegative matrix tri-factorization, user interaction behavior

1 Introduction

As an emerging social media, microblogging has been a convenient service platform for people to share and communicate. In China, various microblogging platforms have been developed and attracted plenty of users. While people immerse themselves in the convenience and freshment of microblogging, a larger number of business and malicious behaviors, such as adlet, sweepstake, sales promotion, and Internet mercenaries, are widespread on the microblogging platforms. Billions of messages are posted on the microblogging platforms every day and propagated quickly through users' interaction behaviors. Anomalous interaction behaviors or anomalous messages seriously affect the confidence and security of microblogging platforms. Hence, detecting these abnormal activities and messages plays an important role in purifying microblogging platforms[1-2].

Previous researches on anomaly detection focus on individual anomalous users or message detections. Therefore, anomalous users can escape the anomaly detection of microblogging systems using advanced intelligent means. For example, some anomalous users mostly post normal messages, while occasionally posting promotional activities, advertising, spam messages and so on. It is difficult to detect an abnormal event or user based on a single element, such as user behaviors or messages. In this paper, anomalies are typically defined in terms of deviation from some expected behaviors[3], such as some anomalous messages posted by users. Simultaneously, anomalous users evolve into swarm intelligence. The individual users in a collective anomaly may not be anomalies by themselves, but they will show anomaly as a group. For instance, a group of users collude to post some false reviews or threat campaigns in microblogging; in large organizations, malfunctioning teams or insider groups closely coordinate with each other to achieve a malicious goal.

Regular Paper

Special Section on Social Media Processing

This work was supported by the National Natural Science Foundation of China under Grant No. 61170242, the National High Technology Research and Development 863 Program of China under Grant No. 2012AA012802, and the Fundamental Research Funds for the Central Universities of China under Grant No. HEUCF100605.

A preliminary version of the paper was published in the Proceedings of SMP 2014.

*Corresponding Author

To solve the problems above, there are some challenges. Firstly, since users and messages are interrelated, they should be considered simultaneously, but they are heterogeneous. The current algorithms are not suitable to fuse them. Secondly, users can generate multi-typed interactions, such as posting, retweeting, commenting, mentioning, and following. Retweets and follows are generated in the same entities named as homogeneous interactions. Other interactions generated in the different entities are named as heterogeneous interactions. Heterogeneous and homogeneous interactions need to be considered simultaneously for abnormal detections. Thirdly, it is insufficient to only consider interactions or users and messages. They are correlative and compositive in microblogging. However, their characteristics are variant and difficult to coalesce.

In the paper, we propose a co-clustering algorithm based on nonnegative matrix tri-factorization to detect anomalous users and messages simultaneously. The algorithm mainly includes three steps. Firstly, a bipartite graph between users and messages is built to model homogeneous and heterogeneous interactions. Secondly, we integrate homogeneous interactions into heterogeneous interaction matrix based on distance metric learning. Finally, the co-clustering algorithm based on nonnegative matrix tri-factorization co-clusters users and messages.

The main contributions of our work are summarized as follows.

• We firstly propose an innovative framework of anomaly detection based on co-clustering and bipartite graph. A bipartite graph between users and messages is creatively used to model homogeneous and heterogeneous interaction relations, which are extracted based on the interaction behaviors.

• We innovatively provide a co-clustering algorithm based on non-negative matrix tri-factorization to detect anomalous users and messages simultaneously. In this algorithm, we consider not only the attribute of users and content of messages, but also the homogeneous and heterogeneous interactions.

• Extensive empirical experiments are constructed on real-world Sina Weibo data. The performance evaluation reveals that the proposed methods are effective for either individual anomaly detections or group anomaly detections.

The rest of the paper is organized as follows. Related work is described in Section 2. Anomalous analysis in Sina Weibo is provided in Section 3. The framework and the algorithm based on co-clustering for anomaly detecting are detailed in Section 4. Experiment and evaluation are presented in Section 5. Section 6 draws conclusions.

2 Related Work

In this section, we review the related work on anomaly detection in social networks and illustrate our motivations of the paper.

2.1 Individual Anomaly Detection

Anomaly detections have been studied on various social networking platforms[4-5]. Microbloggings, such as Twitter and Sina Weibo, have become popular social network platforms for information dissemination. With the availability of microblogging growing, social spammings have become rampant[6].

Numerous studies are made on Twitter[7-8] and Sina Weibo[9-10]. A large number of individual anomaly detection methods have been proposed, such as machine learning[11] and data mining[12]. The recent researches focus on detecting anomalous users or messages individually[13-15]. The content of a message or the profile of a user is employed by anomaly detection algorithms. In order to evade the Sina Weibo's own anomaly detection system, anomalous users leverage more intelligent means. For example, anomalous users post some normal messages or normal users post some anomalous messages. It is difficult to detect them[16-17].

To address these new challenges, we use a co-clustering algorithm[18] to analyze users and messages simultaneously. We consider not only the attributes of users and the contents of messages, but also the heterogeneous interaction relationships between users and messages[19]. SSDM is the most similar to our work[20-21], which models social networks and content information in a unified framework. However, it does not consider heterogeneous interaction relationships between users and messages.

In microblogging, a user can generate many interactions between users and messages, which facilitate the dissemination of information. A bipartite graph is usually used to model the relations between two entities. Some algorithms based on bipartite graphs were proposed to detect anomalies[22-23], but they only consider positive cases and negative cases. By contrast, we consider five user interaction behaviors, including following, retweeting, mentioning, posting, and commenting.

2.2 Group Anomaly Detection

On the basis of individual anomaly users and message detections, we next pay attention to detecting groups that exhibit an anomalous behavior pattern. Group anomaly detection approaches rely on the structure of the groups[24], so the graph is usually used to model users' following relationships or retweeting relationships[25].

Starting with the detection of community structures from user interaction network, community-based features were used to build a classification model for detecting spam nodes in social networks[26]. Another two-stage method was put forward[16], which focuses on detecting spam campaigns that manipulate multiple accounts to spread spam on Twitter.

The two-stage group anomaly detection methods consider only the group structure, but not the mutual influence between the group structure and the content attributes. Yu *et al.* proposed a hierarchical Bayes model GLAD[27] that can accomplish the tasks of group discovery and anomaly detection all at once. However, interaction behaviors are not considered in GLAD. Based on the attributes of users and the content of messages, we analyze the homogeneous and heterogeneous interactions in groups.

3 Anomaly Analysis in Sina Weibo

Sina Weibo is the most popular social network platform in China. Various operational interfaces are provided to generate contents and following relationship, such as posting, retweeting, commenting, mentioning, following and so on. A user can post an original message that is up to 140 characters, called as tweet. A user reposts a tweet to his or her followers, which is called as retweeting. A user comments on a follower's tweet, which is visible to his/her followers. Any tweet can mention some users and contain tags, pictures, videos, etc.

The operating interfaces of posting, following, and mentioning are the same as Twitter's. But the operating interfaces of retweeting and commenting are not exactly the same as Twitter's. The retweeting interface is shown in Fig.1(a), and the commenting interface is shown in Fig.1(b).

In Fig.1(a), when a user retweets the message of user "Shen Guowei", he/she can comment on the message of user "Shen Guowei" and user "Yumiao-miao" at the same time. In Fig.1(b), when a user comments on the message of user "Shen Guowei", he/she can simultaneously retweet the message of user "Yumiao-miao" and comment on the message of "Yumiao-miao". Note that the message of user "Yumiao-miao" is the original message, and user "Shen Guowei" retweeted the original message before. Through the analysis of Sina Weibo interfaces, we can see that a user can produce multiple interactions only by one operation.

Fig.1. (a) Retweeting and (b) commenting produce multiple interactions.

Through operating interfaces, a large number of tweets are posted and propagated. However, some anomalous messages also exist in Sina Weibo, such as spammer, advertising, sweepstake, and promotion.

With the development of techniques, anomalous users are increasingly more intelligent. Fig.2 shows that it is difficult to detect two users' profiles. Anomalous messages are marked by the red boxes, and normal messages are marked by blue boxes. In order to increase the activity and fraudulence, anomalous users post some normal messages shown in Fig.2(a). Another case is shown in Fig.2(b). Normal users often post normal messages, and occasionally post anomalous messages.

Traditional algorithms are based on the assumption: messages posted by anomalous users are anomalous ones, and normal users do not post normal messages. However, the real situation is that anomalous

Fig.2. Anomalous cases in Sina Weibo. (a) An anomalous user's profile. (b) A normal user's profile.

users also post some normal messages, and normal users also post or retweet anomalous messages. Through the above analysis, the assumption no longer holds. In this paper, we focus on the detection of anomalies resulting from interaction behaviors different from normal behaviors in microblogging.

In order to detect anomaly cases shown in Fig.2, the anomaly detection algorithm should consider not only the attributes of users and the contents of messages, but also the heterogeneous interactions between users and messages.

4 Anomaly Detection Based on Co-Clustering

In this section, we firstly provide a framework for anomaly detections based on co-clustering. Then, the details of the framework are introduced.

4.1 Framework of Anomaly Detection

In order to detect anomalous users and messages simultaneously, we propose a framework based on nonnegative matrix tri-factorization (NMTF) to co-cluster users and messages. The framework is shown in Fig.3 and includes three steps.

Firstly, we extract heterogeneous and homogeneous interaction behaviors, the attributes of users, and the contents of messages from microblogging. Heterogeneous and homogeneous interactions are modeled in a bipartite graph. The symbols F, R, M, P and C denote interaction behaviors, which are introduced in Subsection 4.2 in detail. A bipartite graph can be represented by a homogeneous interaction matrix: user relation matrix $\boldsymbol{U}$, tweet relation matrix $\boldsymbol{T}$, and heterogeneous interaction matrix $\boldsymbol{B}$.

Secondly, we build user constraint matrices $\boldsymbol{U}^{\text{sim}}$ and $\boldsymbol{U}^{\text{dis}}$ based on the attributes of users $\boldsymbol{F}_U$ and user relation matrix $\boldsymbol{U}$. Tweet constraint matrices $\boldsymbol{T}^{\text{sim}}$ and $\boldsymbol{T}^{\text{dis}}$ are built similarly. Distance metrics $\boldsymbol{L}_U$ and $\boldsymbol{L}_T$ can be learned from $\boldsymbol{U}^{\text{sim}}$, $\boldsymbol{U}^{\text{dis}}$ and $\boldsymbol{T}^{\text{sim}}$, $\boldsymbol{T}^{\text{dis}}$ respectively. Constraint heterogeneous interaction matrix $\tilde{\boldsymbol{B}}$ can be built based on $\boldsymbol{L}_U$, $\boldsymbol{L}_T$, and $\boldsymbol{B}$.

Finally, the problem of detecting anomalous users and tweets simultaneously is treated as a co-clustering problem. The partition indicator matrices $\boldsymbol{P}_U$, $\boldsymbol{P}_M$ are got from the co-clustering algorithm based on nonnegative matrix tri-factorization (NMTF).

4.2 Features Extraction and Modeling

In Sina Weibo, a user can produce multiple interactions only by one operation, different from Twitter.

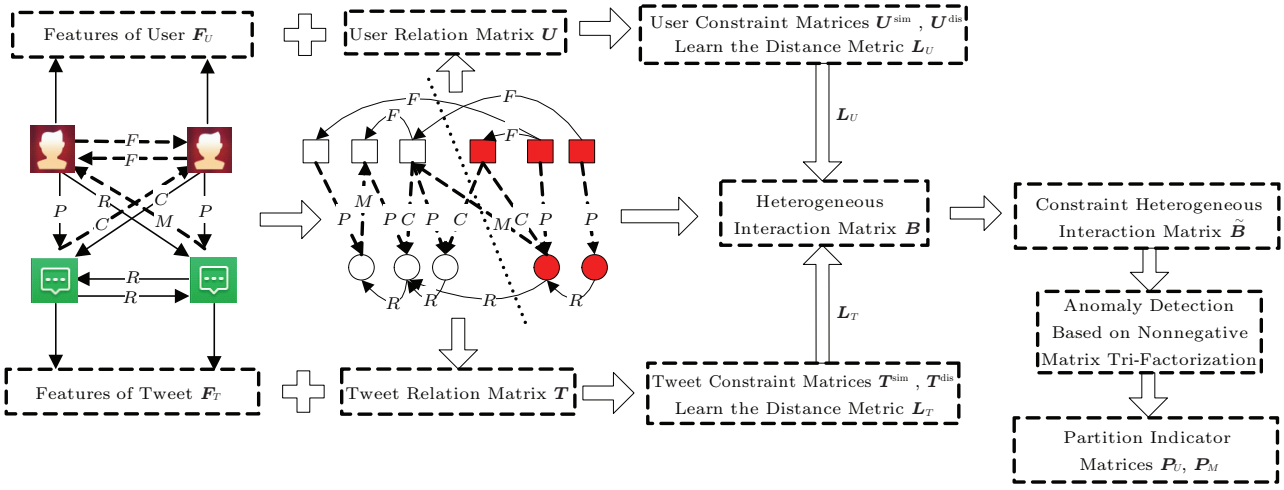


Fig.3. Framework of anomaly detection based on co-clustering.

In order to analyze the user behaviors in microblogging, we only consider two types of entities, users and tweets. Fig.4 shows all interaction behaviors between two entities. In this paper, we consider five types of behaviors: following, retweeting, mentioning, posting, commenting. For any users U_a and U_b, a user interaction relation set is defined by $I = \{F, R, M, P, C\}$. In Fig.4, dotted line arrows represent the interactions between the same types of entities, while solid line arrows represent the interactions between different types of entities.

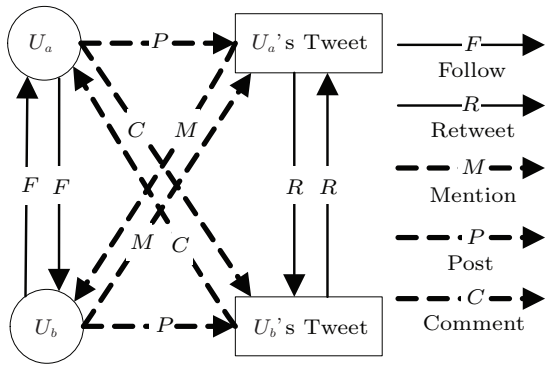


Fig.4. User interaction behaviors in microblogging.

Table 1 lists some symbols used in this paper.

In order to analyze users and messages simultaneously, a bipartite graph model is proposed to model homogeneous and heterogeneous relations, as shown in Fig.5. Heterogeneous interaction behaviors between users and tweets are represented by matrix $\boldsymbol{B}$. Table 2 shows the relation matrix based on user interaction behaviors. Based on matrix $\boldsymbol{B}$, we propose a co-clustering algorithm to process tweets and users simultaneously in matrix $\boldsymbol{B}$.

Table 1. Symbols Used in the Paper

Symbol	Description
$\boldsymbol{U}$	User relation matrix
$\boldsymbol{T}$	Tweet relation matrix
$\boldsymbol{B}$	Heterogeneous interaction matrix
$\tilde{\boldsymbol{B}}$	Constraint heterogeneous interaction matrix
$\boldsymbol{F}_U$	User's feature value vector
$\boldsymbol{F}_T$	Tweet's feature value vector
$\boldsymbol{U}^{\text{sim}}$	User's similar matrix
$\boldsymbol{U}^{\text{dis}}$	User's dissimilar matrix
$\boldsymbol{T}^{\text{sim}}$	Tweet's similar matrix
$\boldsymbol{T}^{\text{dis}}$	Tweet's dissimilar matrix
$\boldsymbol{L}_U$	User distance metric
$\boldsymbol{L}_T$	Tweet distance metric

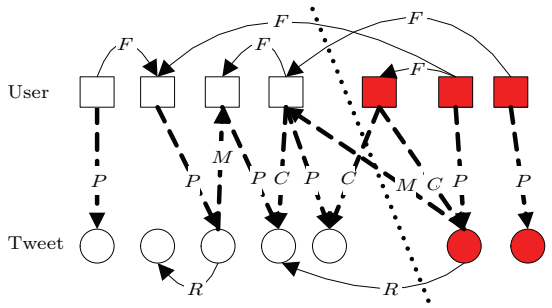


Fig.5. User-message interaction model.

Table 2. Heterogeneous Relations Based on User Behaviors

	U_a's Tweet	U_b's Tweet
U_a	P	C/M
U_b	C/M	P

The value of matrix $\boldsymbol{B}$ is calculated by (1). It is difficult to get all interactions between users and messages, so we only consider the types of interactions, and do not consider the times of interactions.

$$B_{i,j}=\begin{cases}1, \text{if } \ i\leqslant j \ \& \text{ the interaction behavior}\\ \quad \text{is } P \text{ or } C,\\ 1, \text{if } \ i>j \ \& \text{ the interaction behavior}\\ \quad \text{is } M,\\ 0, \text{otherwise.}\end{cases} \tag{1}$$

According to the description in Subsection 3.1, anomalous users and tweets have their own unique characteristics. In this paper, we consider not only users' attributes and the contents of messages, but also the interaction relations between users and tweets. Hence, user U_a's feature value F_{U_a} is calculated by (2), where $N^{U_a}_{\text{follower}}$ is the number of followers, and $N^{U_a}_{\text{following}}$ is the number of followings. The tweet T_x's feature value F_{T_x} is calculated by (3), where $N^{T_x}_{\text{link}}$, $N^{T_x}_{\text{mention}}$, $N^{T_x}_{\text{picture}}$, and $N^{T_x}_{\text{hashtag}}$ are the number of links, mentions, pictures, and hashtags, respectively. $|T_x|$ is the length of the tweet.

$$F_{U_a}=\frac{N^{U_a}_{\text{follower}}}{N^{U_a}_{\text{following}}}, \tag{2}$$

$$F_{T_x}=0.5\left(\frac{N^{T_x}_{\text{link}}+N^{T_x}_{\text{mention}}+N^{T_x}_{\text{picture}}+N^{T_x}_{\text{hashtag}}}{4}\right)+0.5\left(1-\frac{|T_x|}{140}\right). \tag{3}$$

Matrix $\boldsymbol{B}$ is extremely sparse, so homogenous interaction relations are used to build the constraint matrix. There are two types of entities in Sina Weibo. User and tweet constraint matrices are built respectively based on the following assumptions:

1) *User similarity relation*: if two users are anomalous or normal, they are in a cluster.

2) *User dissimilarity relation*: if one is an anomalous user, and the other is a normal user, they are in different clusters.

3) *Tweet similarity relation*: when two users have a following relation, and two tweets are anomalous ones or normal ones, they are in a cluster.

4) *Tweet dissimilarity relation*: when two users do not have a following relation, they are not in a cluster.

User constraint matrices $\boldsymbol{U}^{\text{sim}}$ and $\boldsymbol{U}^{\text{dis}}$ are based on following behaviors. The values of matrices $\boldsymbol{U}^{\text{sim}}$ and $\boldsymbol{U}^{\text{dis}}$ are calculated by (4) and (5) respectively. α is the threshold of anomalous users. $U^{\text{sim}}_{a,b}=1$ shows that users U_a and U_b are similar in the same cluster. $U^{\text{dis}}_{a,b}=1$ shows that users U_a and U_b are dissimilar, and cannot be in the same cluster.

$$U^{\text{sim}}_{a,b}=\begin{cases}1, \text{if } (F_{U_a}>\alpha \ \& \ F_{U_b}>\alpha)|(F_{U_a}\leqslant\alpha\\ \quad \& \ F_{U_b}\leqslant\alpha),\\ 0, \text{otherswise,}\end{cases} \tag{4}$$

$$U^{\text{dis}}_{a,b}=\begin{cases}1, \text{if } \ F_{U_a}\leqslant\alpha \ \& \ F_{U_b}>\alpha,\\ 0, \text{otherwise.}\end{cases} \tag{5}$$

Tweet constraint matrices $\boldsymbol{T}^{\text{sim}}$ and $\boldsymbol{T}^{\text{dis}}$ are based on retweet behaviors. The values of $\boldsymbol{T}^{\text{sim}}$, $\boldsymbol{T}^{\text{dis}}$ are calculated by (6) and (7) respectively where β is the threshold of anomalous tweets, $T^{\text{sim}}_{x,y}$ means the component of $\boldsymbol{T}^{\text{sim}}$ corresponding to tweets T_x and T_y, U_{T_x} and U_{T_y} denote the users of T_x and T_y respectively, F_{T_y} denotes the value of the feature vector of tweet T_y, and $F_{u_{T_x}}$ means the value of the feature vector of user U_{T_x} who posted T_x.

$$T^{\text{sim}}_{x,y}=\begin{cases}1, \text{if } \ U_{T_x} \text{ following } U_{T_y} \ \& \ (F_{T_y}\geqslant\beta\\ \quad \& \ F_{U_{T_x}}\leqslant\alpha),\\ 1, \text{if } \ U_{T_x} \text{ following } U_{T_y} \ \& \ (F_{T_y}<\beta\\ \quad \& \ F_{U_{T_x}}\geqslant\alpha),\\ 0, \text{otherwise,}\end{cases} \tag{6}$$

$$T^{\text{dis}}_{x,y}=\begin{cases}1, \text{if } \ U_{T_x} \text{not following } U_{T_y}\\ \quad \& \ (F_{T_y}<\beta \ \ \& \ F_{U_{T_x}}\leqslant\alpha),\\ 0, \text{otherwise.}\end{cases} \tag{7}$$

4.3 Homogeneous Relations Integration Based on Distance Metric Learning

In order to improve the performance of the co-clustering algorithm, homogenous interaction relations are integrated into heterogeneous relational matrix through the distance metric learning[28].

Given any two data points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, the Mahalanobis distance between them can be formulated by

$$\|\boldsymbol{x}_i-\boldsymbol{x}_j\|_{\boldsymbol{L}}=\sqrt{(\boldsymbol{x}_i-\boldsymbol{x}_j)^{\mathrm{T}}\boldsymbol{L}(\boldsymbol{x}_i-\boldsymbol{x}_j)},$$

where $\boldsymbol{L}$ is the Mahalanobis distance metric. Because $\boldsymbol{L}$ is a positive semi-definite matrix, we can reasonably write $\boldsymbol{L}=\boldsymbol{W}\boldsymbol{W}^{\mathrm{T}}$ by eigen-decomposition. The Mahalanobis distance metric can be formulated by

$$\|\boldsymbol{x}_i-\boldsymbol{x}_j\|_{\boldsymbol{L}}=\sqrt{(\boldsymbol{x}_i-\boldsymbol{x}_j)^{\mathrm{T}}\boldsymbol{W}\boldsymbol{W}^{\mathrm{T}}(\boldsymbol{x}_i-\boldsymbol{x}_j)}.$$

The transformation matrices $\boldsymbol{W}$ can be learned by solving the following objective functions Q_U and Q_T,

which can be solved via simultaneous ℓ_1-norm minimization and maximization[29].

$$Q_U = \min_{\boldsymbol{W}_U^{\mathrm{T}}\boldsymbol{W}_U=1} \frac{trace(\boldsymbol{W}_U^{\mathrm{T}}\boldsymbol{S}_{U^{\mathrm{sim}}}\boldsymbol{W}_U)}{trace(\boldsymbol{W}_U^{\mathrm{T}}\boldsymbol{S}_{U^{\mathrm{dis}}}\boldsymbol{W}_U)}, \tag{8}$$

$$Q_T = \min_{\boldsymbol{W}_T^{\mathrm{T}}\boldsymbol{W}_T=1} \frac{trace(\boldsymbol{W}_T^{\mathrm{T}}\boldsymbol{S}_{T^{\mathrm{sim}}}\boldsymbol{W}_T)}{trace(\boldsymbol{W}_T^{\mathrm{T}}\boldsymbol{S}_{T^{\mathrm{dis}}}\boldsymbol{W}_T)}, \tag{9}$$

where $\boldsymbol{S}_{U^{\mathrm{sim}}}$ and $\boldsymbol{S}_{U^{\mathrm{dis}}}$ are the covariance matrices, built from user constraint matrices $\boldsymbol{U}^{\mathrm{sim}}$ and $\boldsymbol{U}^{\mathrm{dis}}$ respectively. $\boldsymbol{S}_{T^{\mathrm{sim}}}$ and $\boldsymbol{S}_{T^{\mathrm{dis}}}$ are the covariance matrices, built from tweet constraint matrices $\boldsymbol{T}^{\mathrm{sim}}$ and $\boldsymbol{T}^{\mathrm{dis}}$ respectively. In (8) and (9), $trace()$ is the trace of matrix.

Distance metrics $\boldsymbol{L}_U$ and $\boldsymbol{L}_T$ can be learned from $\boldsymbol{L}_U = \boldsymbol{W}_U\boldsymbol{W}_U^{\mathrm{T}}$ and $\boldsymbol{L}_T = \boldsymbol{W}_T\boldsymbol{W}_T^{\mathrm{T}}$. Through learning the distance metrics $\boldsymbol{L}_U$ and $\boldsymbol{L}_T$, the homogenous interactions are embedded into heterogeneous relation matrix $\boldsymbol{B}$. Through (10), the original relation matrix $\boldsymbol{B}$ is projected into a new space. The new heterogeneous relation matrix $\tilde{\boldsymbol{B}}$ is provided for co-clustering algorithm.

$$\tilde{\boldsymbol{B}} = \sqrt{\boldsymbol{L}_U}\boldsymbol{B}\sqrt{\boldsymbol{L}_T}. \tag{10}$$

4.4 Anomaly Detection Algorithm

Following the distance metric learning, the task of co-clustering is formulated as an optimization problem with nonnegative matrix tri-factorization for $\tilde{\boldsymbol{B}}$. Optimization objective function Q is provided to partition the user and the tweet simultaneously.

$$Q = \min_{\boldsymbol{P}_U\geqslant 0,\boldsymbol{P}_M\geqslant 0} \left\|\tilde{\boldsymbol{B}} - \boldsymbol{P}_U\boldsymbol{S}\boldsymbol{P}_M\right\|_F^2, \tag{11}$$

where $\boldsymbol{P}_U$ is the user partition indicator matrix, $\boldsymbol{P}_M$ is the tweet partition indicator matrix, and $\boldsymbol{S}$ is the cluster association matrix, which provides the relation between users and tweets.

The overall anomaly detection algorithm is shown in Algorithm 1. In steps 2 and 3, homogenous interactions are employed as constraint conditions, which are embedded into heterogeneous relation matrix $\boldsymbol{B}$ by distance metric learning. In order to obtain the local optimal solution for objective function (11), cluster structures for users and tweets are updated iteratively. In steps 5∼7, we derive an EM (expectation maximization) style approach that iteratively performs the matrix decomposition using a set of multiplicative updating rules.

In Algorithm 1, when detecting individual anomaly, we set K to 2. There are two clusters: an anomalous cluster and a normal cluster. The partition indicator matrices $\boldsymbol{P}_U$ and $\boldsymbol{P}_M$ can easily be distinguished between normal and anomaly. The element of indicator matrices $\boldsymbol{P}_U$ and $\boldsymbol{P}_M$ is 1 or 0, which indicates a normal user and message or an anomalous user and message respectively. In group anomaly detection experiments, we set K based on priori knowledge. Through the detailed analysis of groups' roles, we can easily detect anomalous groups.

Algorithm 1. Anomaly Detection Based on Nonnegative Matrix Tri-Factorization (NMTF)

Input: matrices: $\boldsymbol{B}$, $\boldsymbol{U}^{\mathrm{sim}}$, $\boldsymbol{U}^{\mathrm{dis}}$, $\boldsymbol{T}^{\mathrm{sim}}$, $\boldsymbol{T}^{\mathrm{dis}}$
Output: user and tweet partition indicator matrices: $\boldsymbol{P}_U$ and $\boldsymbol{P}_M$

1: Initialize $\boldsymbol{P}_U, \boldsymbol{P}_M, \boldsymbol{S}$, and the number of clusters K;
2: Learn distance metrics $\boldsymbol{L}_U$ and $\boldsymbol{L}_T$ based on $\boldsymbol{U}^{\mathrm{sim}}, \boldsymbol{U}^{\mathrm{dis}}, \boldsymbol{T}^{\mathrm{sim}}, \boldsymbol{T}^{\mathrm{dis}}$;
3: Calculate relation matrix $\tilde{\boldsymbol{B}}$ based on $\boldsymbol{B}$, $\boldsymbol{L}_U$ and $\boldsymbol{L}_T$;
4: Take (11) as the objective function, iteratively update $\boldsymbol{P}_U$, $\boldsymbol{P}_M$ and $\boldsymbol{S}$;
5: $\boldsymbol{P}_U \leftarrow \boldsymbol{P}_U \frac{\boldsymbol{S}^{\mathrm{T}}\boldsymbol{P}_M^{\mathrm{T}}\tilde{\boldsymbol{B}}}{\boldsymbol{S}^{\mathrm{T}}\boldsymbol{P}_M^{\mathrm{T}}\boldsymbol{P}_M\boldsymbol{S}\boldsymbol{P}_U}$;
6: $\boldsymbol{P}_M \leftarrow \boldsymbol{P}_M \frac{\tilde{\boldsymbol{B}}\boldsymbol{P}_U^{\mathrm{T}}\boldsymbol{S}^{\mathrm{T}}}{\boldsymbol{P}_M\boldsymbol{S}\boldsymbol{P}_U\boldsymbol{P}_U^{\mathrm{T}}\boldsymbol{S}^{\mathrm{T}}}$;
7: $\boldsymbol{S} \leftarrow \boldsymbol{S}\frac{\boldsymbol{P}_M^{\mathrm{T}}\tilde{\boldsymbol{B}}\boldsymbol{P}_U^{\mathrm{T}}}{\boldsymbol{P}_M^{\mathrm{T}}\boldsymbol{P}_M\boldsymbol{S}\boldsymbol{P}_U\boldsymbol{P}_U^{\mathrm{T}}}$;
8: Until convergence;
9: Return partition indicator matrices $\boldsymbol{P}_U$ and $\boldsymbol{P}_M$;

5 Experiments

In this section, we demonstrate the performance of the proposed scheme based on a Sina Weibo dataset in detail.

5.1 Dataset

In the field of anomaly detection, labeling is often done manually by a human expert and hence requires a substantial effort to obtain the labeled dataset. Typically, it is difficult to get a labeled set of anomalous users and messages which covers all possible type of anomalous behavior.

In the experiments, the dataset was collected from Sina Weibo. In order to collect real anomalous users, we purchased 1 000 anomalous accounts from Taobao, which are usually used for a special purpose. In order to verify the validity of the anomalous accounts, we added 1 000 anomalous accounts as our fans, and 222 anomalous accounts are detected by the Sina Weibo's own anomalous detection system. We cannot get any information of the 222 anomalous accounts from Sina Weibo.

Therefore, 778 anomalous accounts are included in our dataset. We collected 66 283 normal users randomly.

The first page's tweets of each user were collected in the experiments. In order to collect interaction behaviors as many as possible, we collected interaction behavior data separately. The details of the dataset are shown in Table 3.

Table 3. Dataset Description

	Users		Tweets	
	Normal	Anomaly	Normal	Anomaly
Number	66 283	778	1 819 568	942 325

During the preprocessing, we sorted the messages of the users according to post time, and then extracted hashtags, links, pictures and mentioning users in each message. The structural data were prepared for anomaly detection algorithm.

In order to evaluate the effectiveness of our anomaly detection method based on nonnegative matrix trifactorization, we employ the standard information retrieval metrics, viz. precision, recall, and $F1$-score.

5.2 Results and Discussion

5.2.1 *Selection of Message Number*

For anomalous users detection, if the number of messages is very large, the efficiency of the algorithm may be affected; otherwise, the accuracy of the algorithm could be affected. We need to verify how many messages are needed for anomalous user detection. For every user, we extracted the messages on the first page, which are ordered by time. In Fig.6, the result of varying the number of messages shows that $F1$-score is higher than 0.9 when we extracted ten messages.

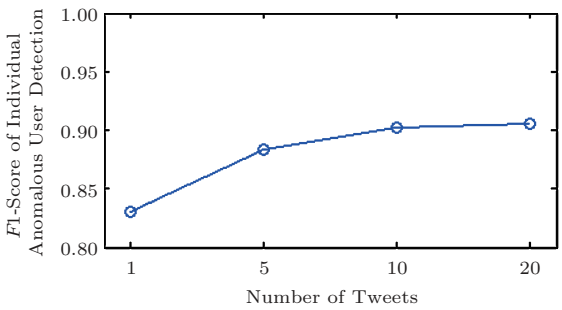


Fig.6. Result of varying the number of messages.

5.2.2 *Individual Anomaly Detection*

To empirically study the effectiveness of our NMTF-based method, we compared the accuracy of the NMTF-based method, NMF-based method, SVM-based method, and SSDM-based method[20]. The NMF-based method is a nonnegative matrix factorization without integrating constraint conditions. The SVM-based method is a classical classification method to detect spammers. We employed LibSVM[30] as the baseline classifier method, and trained two SVM models, using the user and the tweet as the feature vector. The SSDM-based method is designed to only detect social spammers. In the experiment of anomalous message detection, the messages posted by anomalous users are considered as anomalous messages.

In comparative experiments, for each user, we extracted ten messages. The average results of individual anomalous user detection are shown in Fig.7. Compared with the SSDM-based method, the NMTF-based method increases the precision by 1.7%. The average results of individual anomalous message detection are shown in Fig.8. The NMTF-based method can increase the precision by more than 5%. Experimental results show that the NMTF-based method has the highest accuracy in both anomalous user and anomalous message detection. The reason is that both homogeneous and heterogeneous interactions are considered in the NMTF-based method. However, as a classic classification method, the accuracy of the SVM-based method is sensitive to the characteristics of data.

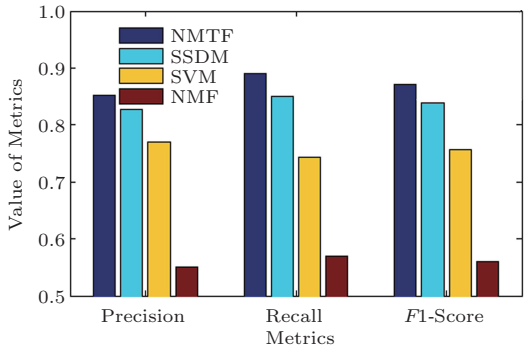


Fig.7. Results of individual anomalous user detection.

In the anomalous user detection experiments, although the characteristics of anomalous users are more obvious, the accuracy of anomalous user detection is lower than that of normal user detection. Through analyses, the reason is that anomalous users post plenty of normal messages but fewer anomalous messages.

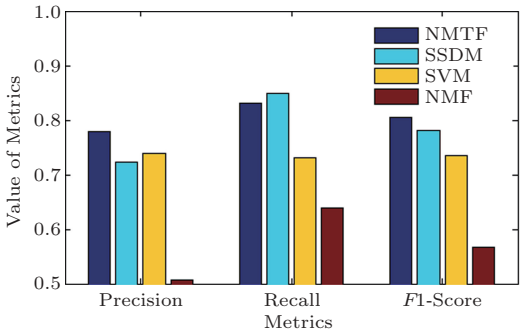


Fig.8. Results of individual anomalous message detection.

In experiments of anomalous message detection, the accuracy of anomaly detection is not very high. Since plenty of messages have few features, it is hard to judge whether an anomalous message is posted by a normal user or not. The result of the SSDM-based method with high recall shows that the anomalous messages posted by anomalous users have high quality.

For the in-depth analysis of the accuracy of the NMTF-based method, we analyze the messages and users in detail. Fig.9 shows four characteristics of messages. Compared with normal messages, anomalous messages contain more links, pictures, and hashtags. But the number of mentions is roughly equal. Anomalous messages have more obvious features, so the accuracy of anomalous detection is very high correspondingly.

The ratio of users' follower number (Nfollower) to following number (Nfollowing) is depicted in Fig.10. For a normal user, the user's follower number is more than his/her following number, as shown in Fig.10. The number of the most anomalous users' followers is smaller than the number of their followings. But there are some exceptions shown in Fig.8 above. Users named "@Love-Constellation Shopping Fashion" and "@Miko Sweater Channel" are detected as anomalous ones marked in Fig.10. The number of the user's followers is higher than the number of his/her followings. This feature is the same as that of normal users, so the traditional method cannot detect such users. Because the user occasionally posts some promotional messages, it is difficult to detect it.

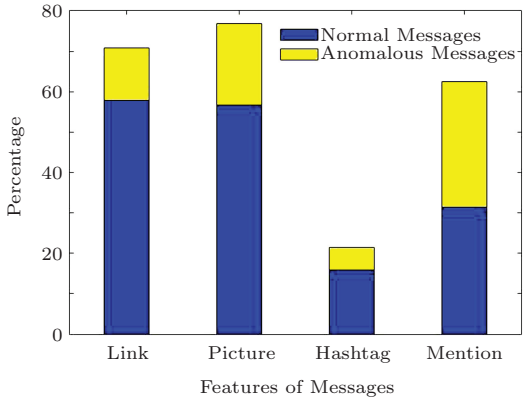


Fig.9. Statistical characteristics of messages.

5.2.3 *Group Anomaly Detection*

In the group anomalous user detection experiment, we chose the GLAD-based method[16] as the baseline method. Because the Sina Weibo dataset does not possess ground truth, we illustrate the effectiveness of our method by comparing it with GLAD.

In the Sina Weibo dataset, our method successfully detected some anomaly groups. Through the de-

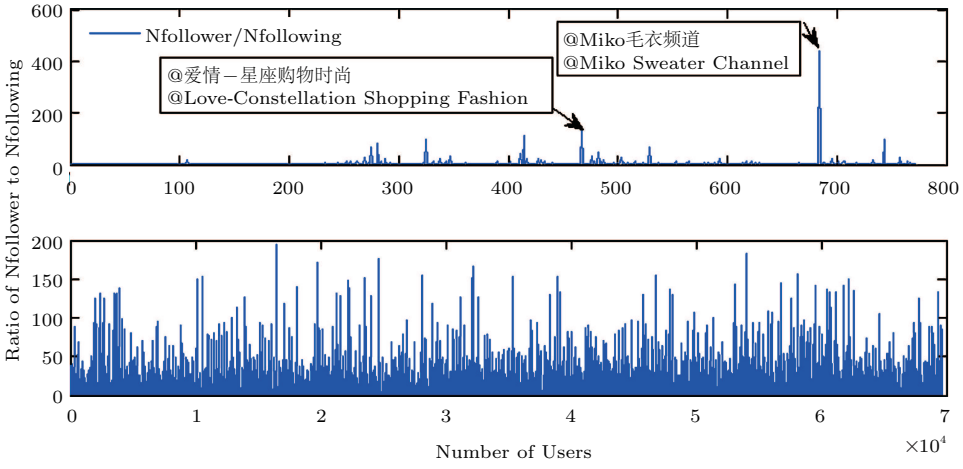


Fig.10. Ratio of users' follower number to following number.

tailed analysis of the groups' roles, we find that the groups closely coordinate with each other to achieve a special goal, such as sale promotion groups and vote groups. The most representative words of two anomalous groups, sales promotion and vote, are shown as in Table 4, which illustrates the roles of anomaly groups. Representative words detected by the NMTF-based method and the GLAD-based method are similar. It is proved that our method can detect not only individual anomalous users and messages, but also group anomalies.

Table 4. Most Representative Words Used in Two Anomalous Groups Detected by NMTF and GLAD

Sales Promotion		Vote	
NMTF	GLAD	NMTF	GLAD
Taobao	Taobao	Help	Thanks
Favorable	Presence	Follow	Help
Special	Price	Thanks	Vote
Price	Gift	Support	Follow
Gift	Haha	Vote	Retweet
New	Sale	Retweet	Microblog
Style	Style	Microblog	Support
Haha	Trousers	Fans	Fans
Woman	New	Group	Popularize
Sale	Favorable	Popularize	Game

6 Conclusions

With the increase of the intelligence of anomalous users, the performance of traditional abnormal detection suffers from serious falloff. In this paper, we proposed an innovative framework of anomaly detection based on co-clustering and bipartite graph. A bipartite graph model was proposed to depict homogeneous and heterogeneous interaction relations of users and messages in Sina Weibo. Then a co-clustering algorithm based on nonnegative matrix tri-factorization was put forward to detect anomalous users and messages simultaneously. Further, our method can detect group anomalous users and give the roles of groups. The homogeneous interactions were integrated into a co-clustering algorithm and improved the accuracy of the algorithm. The experimental results showed that the accuracy of our method is very high.

In the future work, we will employ more information for our anomaly detection algorithm, such as external media information[31] and sentiment information[32]. And we will extend our method to online mode and build a dynamic bipartite graph for online anomaly detection based on real-time message streams.

Acknowledgements Authors would like to thank the anonymous reviewers and the editors for their valuable comments and suggestions to improve the paper's quality.

References

[1] Takahashi T, Tomioka R, Yamanishi K. Discovering emerging topics in social streams via link-anomaly detection. *IEEE Trans. Knowledge and Data Engineering*, 2014, 26(1): 120-130.

[2] Guille A, Favre C. Mention-anomaly-based event detection and tracking in Twitter. In *Proc. the IEEE International Conference on Advances in Social Network Analysis and Mining*, August 2014, pp.375-382

[3] Savage D, Zhang X, Yu X *et al.* Anomaly detection in online social networks. *Social Networks*, 2014, 39: 62-70.

[4] O'Callaghan D, Harrigan M, Carthy J *et al.* Network analysis of recurring YouTube spam campaigns. In *Proc. the 6th AAAI Conference on Weblogs and Social Media*, June 2012, pp.531-534.

[5] Gao H, Hu J, Huang T *et al.* Security issues in online social networks. *IEEE Internet Computing*, 2011, 15(4): 56-63.

[6] Zhu Y, Wang X, Zhong E *et al.* Discovering spammers in social networks. In *Proc. the 26th AAAI Conference on Artificial Intelligence*, July 2012, pp.171-177.

[7] Kwak H, Lee C, Park H *et al.* What is Twitter, a social network or a news media? In *Proc. the 19th WWW*, April 2010, pp.591–600.

[8] Wu S, Hofman J M, Mason W A *et al.* Who says what to whom on Twitter. In *Proc. the 20th WWW*, Match 28-April 1, 2011, pp.705–714.

[9] Yu L, Asur S, Huberman B A. What trends in Chinese social media. In *Proc. the 5th SNA-KDD Workshop*, August 2011.

[10] Gao Q, Abel F, Houben G *et al.* A comparative study of users' microblogging behavior on Sina Weibo and Twitter. In *Lecture Notes in Computer Science 7379*, Masthoff J, Mobasher B, Desmarais M C *et al.* (eds.), Springer Berlin Heidelberg, 2012, pp.88-101.

[11] McCord M, Chuah M. Spam detection on Twitter using traditional classifiers. In *Lecture Notes in Computer Science 6906*, Alcaraz Calero J M, Yang L T, Mármol F G *et al.* (eds.), Springer Berlin Heidelberg, 2011, pp.175–186.

[12] Martinez-Romo J, Araujo L. Detecting malicious tweets in trending topics using a statistical analysis of language. *Expert Systems with Applications: An International Journal*, 2013, 40(8): 2992-3000.

[13] Bosma M, Meij E, Weerkamp W. A framework for unsupervised spam detection in social networking sites. In *Lecture Notes in Computer Science 7224*, Baeza-Yates R, de Vries A P, Zaragoza H *et al.* (eds.), Springer Berlin Heidelberg, 2012, pp.364–375.

[14] Altshuler Y, Fire M, Shmueli E *et al*. Detecting anomalous behaviors using structural properties of social networks. In *Proc. the 6th International Conference on Social Computing, Behavioral Cultural Modeling and Prediction*, April 2013, pp.433-440.

[15] Zhang Q, Ma H, Qian W *et al*. Duplicate detection for identifying social spam in microblogs. In *Proc. the 2nd IEEE International Congress on Big Data*, June 27-July 2, 2013, pp.141-148.

[16] Chu Z, Widjaja I, Wang H. Detecting social spam campaigns on Twitter. In *Lecture Notes in Computer Science 7341*, Bao F, Samarati P, Zhou J (eds.), Springer Berlin Heidelberg, 2012, pp.455-472.

[17] Jiang J, Wilson C, Wang X *et al*. Understanding latent interactions in online social networks. In *Proc. the 10th ACM SIGCOMM Conference on Internet Measurement*, November 2010, pp.369-382.

[18] Chen Y, Wang L, Dong M. Non-negative matrix factorization for semi-supervised heterogeneous data coclustering. *IEEE Trans. Knowledge and Data Engineering*, 2010, 22(10): 1459-1474.

[19] Tang L, Wang X F, Liu H. Community detection via heterogeneous interaction analysis. *Data Mining and Knowledge Discovery*, 2012, 25(1): 1-33.

[20] Hu X, Tang J L, Zhang Y C *et al*. Social spammer detection in microblogging. In *Proc. the 23rd International Joint Conference on Artificial Intelligence*, August 2013, pp.2633-2639.

[21] Hu X, Tang J L, Liu H. Online social spammer detection. In *Proc. the 28th AAAI Conference on Artificial Intelligence*, July 2014, pp.59-65.

[22] Dai H, Zhu F, Lim E *et al*. Detecting anomalies in bipartite graphs with mutual dependency principles. In *Proc. the 12th ICDM*, December 2012, pp.171-180.

[23] Sun J, Qu H, Chakrabarti D *et al*. Neighborhood formation and anomaly detection in bipartite graphs. In *Proc. the 5th ICDM*, Nov. 2005, pp.418-425.

[24] Akoglu L, Tong H, Koutra D. Graph based anomaly detection and description: A survey. *Data Mining and Knowledge Discovery*, 2014, 29(3): 626-688.

[25] Zhao B, Ji G, Qu W *et al*. Detecting spam community using retweeting relationships — A study on Sina microblog. In *Lecture Notes in Computer Science 8178*, Cao L, Motoda H, Srivastava J *et al*. (eds.), Springer International Publishing, 2013, pp.178-190.

[26] Bhat S Y, Abulaish M. Community-based features for identifying spammers in online social networks. In *Proc. the 2013 IEEE International Conference on Advances in Social Networks Analysis and Mining*, August 2013, pp.100-107.

[27] Yu R, He X R, Liu Y. GLAD: Group anomaly detection in social media analysis. In *Proc. the 20th ACM SIGKDD KDD*, August 2014, pp.372-381.

[28] Xing E P, Ng A Y, Jordan M I *et al*. Distance metric learning, with application to clustering with side-information. In *Proc. the 16th Neural Information Processing Systems*, December 2002, pp.505-512.

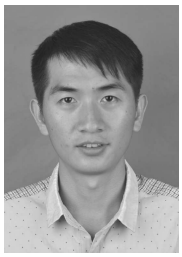
[29] Wang H, Nie F P, Huang H. Robust distance metric learning via simultaneous $l1$-norm minimization and maximization. In *Proc. the 31st International Conference on Machine Learning*, June 2014, pp.1836-1844.

[30] Chang C C, Lin C J. LIBSVM: A library for support vector machines. *ACM Trans. Intelligent Systems and Technology*, 2011, 2(3): 27:1-27:27.

[31] Hu X, Tang J L, Liu H. Leveraging knowledge across media for spammer detection in microblogging. In *Proc. the 37th SIGIR*, July 2014, pp.547-556.

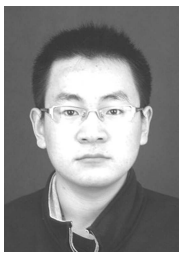
[32] Hu X, Tang J L, Gao H J *et al*. Social spammer detection with sentiment information. In *Proc. the 14th ICDM*, December 2014, pp.180-189.

Wu Yang received his Ph.D. degree in computer system architecture from Harbin Institute of Technology, Harbin, in 2005. He is currently a professor and doctoral supervisor of Harbin Engineering University. His main research interests include data mining, information security and wireless sensor network. He is a senior member of CCF and a member of ACM.



Guo-Wei Shen is currently a Ph.D. candidate in the Department of Computer Science and Technology and Information Security Research Center, Harbin Engineering University. He received his B.E. degree in computer science in 2009 from Harbin Engineering University, Harbin. His main research interests include data mining, social computing and information security. He is a student member of CCF and a member of ACM.

Wei Wang received his Ph.D. degree in computer system architecture from Harbin Institute of Technology in 2005. He is currently an associate professor of Harbin Engineering University. His main research interests include data mining and information security.



Liang-Yi Gong is currently a Ph.D. candidate in the Department of Computer Science and Technology and Information Security Research Center, Harbin Engineering University. He received his B.E. degree in computer science in 2010 from Harbin Engineering University, Harbin. His main research interests include wireless networks, mobile computing, network and information security.

Miao Yu is currently a Ph.D. candidate in the Department of Computer Science and Technology and Information Security Research Center, Harbin Engineering University. He received his B.E. degree in computer science in 2010 from Harbin University of Science and Technology, Harbin. His main research interests include data mining and information security.

Guo-Zhong Dong is currently a Ph.D. candidate in the Department of Computer Science and Technology and Information Security Research Center, Harbin Engineering University. He received his B.E. degree in computer science in 2011 from Harbin Engineering University, Harbin. His main research interests include data mining and information security.