Information Sciences 385–386 (2017) 325–337

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Combining tag correlation and user social relation for microblog recommendation

Huifang Ma*, Meihuizi Jia, Di Zhang, Xianghong Lin

College of Computer Science and Engineering, Northwest Normal University, Lanzhou, Gansu 730070, China

ARTICLE INFO

Article history:
Received 17 October 2015
Revised 28 November 2016
Accepted 31 December 2016
Available online 4 January 2017

Keywords:
Microblog recommendation
Tag retrieval
User-tag matrix
User-tag weight
Tag correlation
User–user social relation

ABSTRACT

With the development of social networking applications, microblog has turned to be an indispensable online communication network in our daily life. For microblog users, recommending high quality information is a demanding service. Some microblog services encourage users to annotate themselves with tags, which are used to describe their interests or attributes. However, few users are willing to create tags and available tags are not fully exploited for microblog recommendation. Besides, following/follower relationship in microblog is asymmetric, which can be used not only for communicating with friends or acquaintances but also for getting information on particular subjects. So far, there is no microblog recommendation algorithm which employs all the above mentioned information. This paper aims to investigate a joint framework to combine tag correlation and user social relation for microblog recommendation. Our approach identifies users' interests via their personal tags and social relations. More specifically, a user tag retrieval strategy is established to add tags for users without or with few tags, and the user-tag matrix is then built and user-tag weights are then obtained. In order to solve the problem of sparsity of the matrix, both inner and outer correlation between tags are investigated to update the user-tag matrix. Considering the significance of user social relation for microblog recommendation, a user–user social relation similarity matrix is constructed. Moreover, an iterative updating scheme is developed to get the final tag-user matrix for computing the similarities between microblogs and users. We illustrate the capability of our algorithm by making experiments on real microblog datasets. Experimental results show that the algorithm is effective for microblog recommendation.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

In recent years, online social networks witness a rapid development growth and become probably the main actor of the Web 2.0. It is universally acknowledged that social media is profoundly affecting not only the global economy but also every aspect of our daily life. Meanwhile, a growing number of users keep up with latest information by utilizing microblog tools, which provide a unique mechanism of information diffusion by allowing each user to receive messages from those he follows. Microblog not only amplifies interpersonal circles in social network, but also serves as a significant media for users to get the latest information. It is a momentary information publishing system based on Web 2.0 technology. According to Wikipedia, Sina Weibo, for example, owns over 500 million users, 46.2 million daily active users, and 100 million forwarding

* Corresponding author.
E-mail addresses: mahuifang@yeah.net, 35882614@qq.com (H. Ma).

http://dx.doi.org/10.1016/j.ins.2016.12.047
0020-0255/© 2017 Elsevier Inc. All rights reserved.

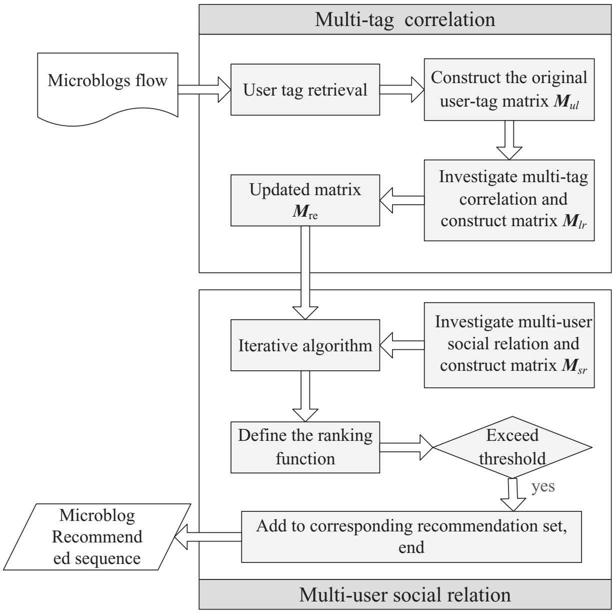

Fig. 1. Overview of the proposed microblog recommendation system.

message each day. However, the richness of online information also brings forth the 'information overload' problem. There is an urgent need to capture, understand and predict user interests for microblog application to provide better product designs, accurate targeted advertising, and personalized services [2,4,5,8]. How to identify users' interests and provide customized recommendation for each user becomes a challenging problem. Thereafter, a large number of recommendation algorithms based on user interest have been proposed [1,11,18,31].

Microblog contents can effectively reflect the public opinions, events in personal life, and personal interests [54]. Compared with the traditional blogs, the sparsity of content and the lack of contextual information in microblogs bring new challenges for recommendation. Most researchers try to extend text features and enrich semantic information for microblogs [33,46]. They usually take advantage of an external database to enrich the semantic information. Some researchers try to trim a short text representation to get a few most representative words for topical classification [44]. Besides, the distinctive features of microblog allow users to post their messages and to follow others, it is possible to utilize these features for better recommendations than other traditional applications.

In this paper, we propose a microblog recommendation algorithm based on tag correlation and user social relation. On the one hand, a user tag retrieval strategy is developed to assign tags for users and a user-tag matrix is created to provide the initial weights for users' tags; then the correlation matrix of multi-tags is constructed by investigating both inner and outer correlation between tags; finally the original user-tag matrix can be updated through multiplication of the above two matrices. On the other hand, a social relation matrix of multi-users is constructed by analysing following/follower relationship between users; then the final user-tag matrix can be obtained via iteratively multiplying social relation matrix with the updated user-tag matrix; finally the similarities between microblogs and users can be defined by the user-tag matrix for microblog recommendation. Our extensive experimental study shows the scalability and efficiency of our approach. Fig. 1 summarizes an overview of the proposed microblog recommendation system.

The contributions of this paper are summarized as follows:

1. We propose a novel tag retrieval method which is suitable for identifying the most representative words from a user's microblogs. Our model selects terms as tags for users by utilizing their significance reflected by importance and topical indication.
2. We investigate the semantic correlation of pairwise tags by involving both the inner correlation, reflecting the explicit correlation within tag pairs for each user and the outer correlation, capturing the implicit correlation between tags by considering the relation strength of their interactions with tags of other users. And we then update the original user-tag matrix, thus better representing users' interests.
3. We design a user–user similarity calculation method based on user–user social relation and provide an iterative approach for updating user-tag matrix to reveal the user's interests more accurately.

The rest of this paper is organized as follows: In Section 2, we discuss several related works. Section 3 presents user tag retrieval strategy, user-tag matrix construction method and multi-tags correlation calculation mechanism. Multi-users social relation matrix construction method, the iterative calculation mechanism and the recommendation algorithm are given in Section 4. The experimental results are demonstrated in Section 5. Lastly, we conclude our paper in Section 6.

2. Related works

Existing works that are most relevant to ours fall into three categories: multi-label learning, microblog user relationship analysis and recommendation on microblogging system. In this section, we briefly summarize the works in each category.

2.1. Multi-label learning

In traditional supervised learning, a sample is represented as a feature vector, and it is always associated with a single label. However, in a broad range of real-world applications [10,14,51], one sample is usually related to multiple class labels simultaneously. Examples include text categorization, image classification, video classification, music categorization, medical diagnosis, chemical analysis and social network mining et al. Generally speaking, Multi-Label Learning (MLL) problems can be dealt with from two perspectives [7,26]. On the one hand, some studies have proposed transformation methods, which transform an original multi-label problem into a multi-class single-label problem by treating an output vector as a scalar value. Then, the problem can be directly solved using any single-label learning approach. Label Power set (LP) approaches and Binary Relevance (BR) approaches are two representative approaches [27]. But they inevitably have some limitations: high computational complexity and assumption of label independence. Through taking the correlations of labels into account, some variations of LP and BR are proposed [29,38,43]. On the other hand, there are studies considering the extension of classical classification paradigms to cope with multi-label data directly, and these methods are called algorithm adaptation methods [23]. There are several lazy learning based approaches proposed that use algorithm adaptation. There also exists neural networks and multi-layer perceptron based algorithms that have been extended for multi-label data [40,47]. These approaches mainly focus on exploiting correlations among multi-labels to facilitate the learning process [16,19]. Recently, multi-label feature selection serves as an effective data pre-processing technique [6,9,26,30].

2.2. Microblog user relationship analysis

Microblog media perfectly imitates the social connection of human beings and organizes its vast users into a social network, thus satisfying the needs of users to post customized information, broadcast socially and establish social contacts. In microblog user behaviour analysis, user relationship analysis is an important part. Usually, the microblog can be considered as a weighted graph. Each edge indicates the relationship between two users and the weighted value of each edge indicates the relationship strength between these two users [52]. Researchers have conducted a lot of research on user relationship analysis in social networks [15,28]. They take advantage of various kinds of information, such as user similarity, network topology, and other characteristics to calculate the strength of user relationships. Some noticeable works include: considering additional transactional events among entities (e.g., communication, file transfers) to infer the true underlying social network, Kahanda et al. [24] develop a supervised learning approach to predict link strength from transactional information and measure the strength of user relationship. By formulating a link-based latent variable model, along with a coordinate ascent optimization procedure for the inference, Xiang et al. [52] develop an unsupervised model to estimate relationship strength from interaction activity (e.g., communication, tagging) and user similarity. Besides, user relationship analysis is also used for friend recommendations [20,22]. By exploiting the “social” features of social folksonomies, Meo et al. [36,37] provide an “enhanced” user with recommendations of similar users and potentially interesting resources based on involved users and resources. Xu et al. [53] define user relationship strength as user similarity, and propose several user similarity estimation approaches by taking advantage of various attribute information of users such as background information, tweets and social information.

2.3. Recommendation on microblogging system

As microblogging systems emerge as one of the most significant social network applications, researchers have started to exploit them to provide various kinds of recommendation services. These include recommending documents, tags, friends and many others in social networks [17,55,57]. Among these recommendation systems, content recommendation play in a key role in social network analysis. Customized recommendation has been well researched in microblog systems. One item is recommended by the collaborative filtering methods to a designated user, in view of the rating of others who are similar to him [50]. Thus, in order to recommend microblogs to a user *u*, some users having similar taste with *u* are found by the collaborative filtering approach, and then relevant microblogs can be recommended for user *u*. There are two types of approaches constantly employed, i.e., the content-based approach [32] and the link-based approach [42]. Content-based approach recommends microblogs that are similar to user’s preference in the past, adopting the state-of-the-art information retrieval approaches to discover the candidate documents. Link-based approach cast recommendation as a link prediction problem by finding the most probable links among existing nodes. Some hybrid approaches [21,25,57] are presented to combine these two techniques to improve the quality of recommendations.

Most of the existing friend recommendation approaches are based on the similarity of user profiles, or the geographical vicinity or the number of common friends [12]. Recently, several matrix factorization methods [13,39,56] have been proposed for friend microblog recommendation. These methods find latent features for users and items by factorizing the observed user-item rating matrix and make latent features for further predictions. As a personalized recommendation task,

Table 1
Symbols used in our analysis.

Notation	Definition
$U=\{u_1, u_2, \ldots, u_i, \ldots u_N\}$	The microblog user dataset
N	The number of users
$D_i = \{d_{i1}, d_{i2}, \ldots, d_{iM_i}\}$	The microblog collection for user u_i
M_i, $i=(1,2,\ldots,N)$	The number of microblog for u_i
$D = \bigcup_{i=1}^{N} D_i$	The microblog dataset of all users
m_i, $m_i \ll M_i$	The number of terms
$L_i = \{l_{i1}, l_{i2}, \ldots, l_{in_i}\}$	The tag set for u_i
n_i	The number of tags for u_i
$L = \bigcup_{i=1}^{N} L_i$	The collection for all tags
n	The total number of tags in tag collection, $\|L\|$

some successful techniques in recommender systems are introduced to address the task of social tag suggestion, these methods include user/item based collaborative filtering [34,48], matrix and tensor decomposition [41], and topic model based methods et al [45].

3. User tag acquisition and multi-tags correlation

3.1. User tag acquisition

In our recommendation algorithm, a set of tags are extracted from microblog content for each user to indicate his/her interests. If the tagging service is provided by microblog system, the built-in tags can be directly used. Otherwise, we will use a tag retrieval method to acquire the personal tags from the microblogs posted by that user. In this section, we introduce our tag retrieval and weighting approach in detail.

3.1.1. User tag retrieval

Table 1 summarizes the notations that will be used in our analysis. Whether a particular term from users previously posted microblog is chosen as query word or not depends on its significance reflected by importance and topical indication. As for importance, the most widely used *TF* (Term Frequency) or *TF-IDF* (Term Frequency-Inverse Document Frequency) weighting scheme can be adopted [49]. As for topical indication, the clarity score [44] is used to measure the topical-specificity of a certain term, which means that if one term is pertained to a specific topic, it should be topically indicative. A set of microblogs best matching a given query is obtained as the query model and the entire microblog collection $D_i = \{d_{i1}, d_{i2}, \ldots, d_{iM_i}\}$is considered as collection model. Let the j-th term l_{ij} be a candidate word for selection, l_{ij} is taken as a single-term query to retrieve its top-g_i most relevant microblogs, denoted as $Q_{l_{ij}}$. Therefore the clarity score of term l_{ij} can be defined as:

$$Clarity\left(l_{ij}\right) = \sum_{l_{ij}' \in \theta} P\left(l_{ij}'|Q_{l_{ij}}\right) \log \frac{P\left(l_{ij}'|Q_{l_{ij}}\right)}{P\left(l_{ij}'|D_i\right)} \tag{1}$$

If l_{ij} is specific to a topic, then the microblogs matching l_{ij} share a common topic indicated by a few words with very high probabilities of occurrences against their probabilities in the entire collection.

Then the score of the *j-th* term is computed as the equation below:

$$s_j = tf_j \times clarity\left(l_{ij}\right) \tag{2}$$

n_i words with the highest weight are chosen as user u_i*'s* tags, and each tag is assigned a normalized weight:

$$normalized\left(s_j\right) = \frac{s_j}{\sum_{x=1}^{n_i} s_x} \tag{3}$$

3.1.2. User-tag matrix construction

The tag weight vector $\overrightarrow{V_i} = (w_{i1}, w_{i2}, \ldots w_{in})$is created for user u_i to represent the initial weights of tags. If the tags are obtained from the above tag retrieval scheme, Eq. (3) is used to generate the initial weights for these tags. Otherwise, if the tags are provided by the tagging service, each tag is treated as of equal importance. Assuming that u_i has Z_i tags, the initial tag weight w_{ij} is defined as follows:

$$w_{ij} = \begin{cases} 1/Z_i, & \text{if tag } l_j \text{assigned to } u_i, \\ normalized\left(s_j\right), & \text{tag retrieval}, \\ 0, & \text{otherwise}, \end{cases} \tag{4}$$

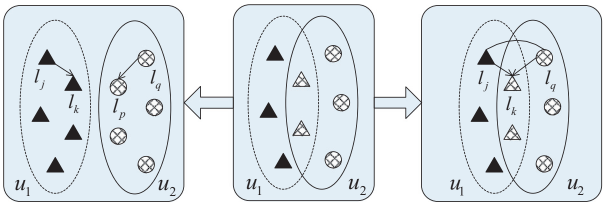

Fig. 2. An overview of multi-tags correlation analysis.

Based on users' weight vectors, we create a $N \times n$ matrix M_{ul}.

$$M_{ul} = \begin{bmatrix} \vec{V_1} \\ \vec{V_2} \\ \cdots \\ \vec{V_N} \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ w_{N1} & w_{N2} & \cdots & w_{Nn} \end{bmatrix} \tag{5}$$

where N denotes the total number of users, n is the number of tags and w_{ij} is the weight of the *j-th* tag for the *i-th* user.

3.2. Multi-tags correlation

In this section, we investigate correlation between tags for updating the original user-tag matrix to better represent users' interests.

Associations between tags can be divided into two categories, which can be shown as below:

In Fig. 2, u_1 and u_2 denotes user 1 and user 2 respectively. White triangle is a jointly marked tag by user 1 and user 2. The solid boxes pointed by the left and right arrow represent two tag correlation patterns respectively. The left solid box represents the tag correlation within a same user, while the right one represents the tag correlation within different users.

3.2.1. Multi-tags inner correlation

The inner correlation is to explore the explicit semantic correlation between tags, which considers the statistical analysis of tag co-occurrence patterns [35]. Specifically, it assumes that tags are regarded relational if they co-occur within the same user. Besides, the more frequently they co-occur, the stronger relation they have. Hence, the explicit relation between tags can be estimated based on the tag co-occurrence frequency across the entire users.

If both tags are marked by one particular user, an **inner correlation** between two tags is established. Based on Jaccard similarity, the inner correlation between l_j and l_k can be defined as:

$$LIR\left(l_j, l_k\right) = \frac{1}{|H|} \times \sum_{i \in H} \frac{w_{ij} w_{ik}}{w_{ij} + w_{ik} - w_{ij} w_{ik}} \tag{6}$$

where w_{ij} and w_{ik} denote the weight of the *j-th* tag l_j and the *k-th* tag l_k for user u, respectively. $H=\{i|(w_{ij} \neq 0)\&(w_{ik} \neq 0)\}$, if $H=\emptyset$, $LIR(l_j, l_k)=0$. We normalize $LIR(l_j, l_k)$ to [0,1], and the normalized inner correlation of tags l_i and l_k is defined as:

$$N - LIR\left(l_j, l_k\right) = \begin{cases} 1, & j = k \\ \frac{LIR\left(l_j, l_k\right)}{\sum_{j=1, j \neq k}^{n} LIR\left(l_j, l_k\right)}, & j \neq k \end{cases} \tag{7}$$

where n denotes the number of tags, For all tags l_j, if l_j and l_k are not the same, $\sum_{j=1, j \neq k}^{n} N - LIR(l_j, l_k) = 1$.

3.2.2. Multi-tags outer correlation

The outer correlation can capture the semantic correlation of tag pairs by taking the interactions of other tags in the user set into account. It is clear that the correlation between tags can be further enriched by exploring the outer correlation, due to that it is not based on tags within the same user, but on interactions with all other tags from all users.

If two users u_1 and u_2 are simultaneously marked by the same tag, an **outer correlation** between two tags is established. To be more specific, given two tags l_j and l_k, if there is a tag l_q, satisfying $N - LIR(l_j, l_q) > 0$ and $N - LIR(l_k, l_q) > 0$, there is an outer correlation between tags l_j and l_k, where tag l_q is a linking tag which links l_j and l_k. The outer correlation between two tags l_j and l_k linked by term l_q can be defined as:

$$LOR\left(l_j, l_k | l_q\right) = \min\left(N - LIR\left(l_j, l_q\right), N - LIR(l_k, l_q)\right) \tag{8}$$

We then define the outer correlation between two tags l_j and l_k with all the linked terms and normalize the values to [0,1] as:

$$N-LOR\left(l_j,l_k\right)=\begin{cases}0, & j=k\\ \frac{\sum_{\forall t_q\in E}LOR\left(l_j,l_k|l_q\right)}{|E|}, & j\neq k\end{cases} \tag{9}$$

where $E=\{l_q|(N-LIR(l_j,l_q)>0)\&(N-LIR(l_k,l_q)>0)\}$is the set of all linking tags. If $E=\emptyset$, there is no linking tags between two tags.

Given a pair of tags l_j and l_k, the tag correlation between l_j and l_k can be defined as:

$$LR\left(l_j,l_k\right)=\begin{cases}1 & j=k\\ \alpha\times N-LOR\left(l_j,l_k\right)+(1-\alpha)\times N-LIR\left(l_j,l_k\right) & otherwise\end{cases} \tag{10}$$

where $\alpha(\alpha\in[0,1])$ determines the relative importance of inner correlation and outer correlation between multi-tags.

3.2.3. Update the user-tag matrix

A $n\times n$ multi-tags relationship matrix $\boldsymbol{M}_{lr}$ can be created, each entry $LR(l_j,\ l_k)$reflects the semantic correlation between tag l_j and l_k, and $LR(l_j,\ l_k)$are elements in this matrix. The final user-tag matrix $\boldsymbol{M}_{re}$ can be defined:

$$\mathbf{M}_{re}=\mathbf{M}_{ul}\times\mathbf{M}_{lr} \tag{11}$$

The diagonal entry must be greater than zero because each tag is similar to itself and others entries are almost greater than zero. Hence, after mapping, each user is represented by a less sparse vector that has non-zero entries for all tags that are semantically correlated to those that appear for users. Accordingly, $\boldsymbol{M}_{re}$ is denser than the original matrix $\boldsymbol{M}_{ul}$ and can better represent user's interest.

4. Multi-users social relation and the recommendation algorithm

It is observed from Eq. (11) that the accuracy of the recommendation model is only determined by correlation of multi-tags. However, social relations measure social intimacy in the microblog network via their interrelationships and shared properties. Our primary idea of the recommendation framework is to combine tag correlations and social relations to predict users' preferences. Therefore we construct a user–user social relation similarity matrix by analyzing social relation and investigating the following/follower relationship between multi-users. And the final users' interests can be acquired via iteratively calculating with the updated user-tag matrix.

4.1. Social relation between multi-users

A follower follows another user, known as the followee, creating an explicit following/follower relationship. Generally, users would like to follow other users who post interesting posts. More specifically, in the social network of microblogging systems, users are connected via following/follower relationship. If a user u_i is interested in another user u_j*'s* microblogs, u_i will intentionally follow u_j, then u_i becomes u_j*'s* follower, and u_j is at the same time becomes u_i*'s* following. Furthermore, although users u_i and u_k are not connected through following/follower relationship, they simultaneously follow user u_j, which implicitly indicates users u_i and u_k share similar interests. We can therefore define the social relation between multi-users as follows.

The social relation of user u_i is defined as $SR(u_i)=(\boldsymbol{Fg}(u_i),\boldsymbol{Fr}(u_i))$, namely, *Social Relation*(u_i) =(***Following***(u_i), ***Follower***(u_i)). $SR(u_i)$ denotes the social relation of user u_i, which contains two attributes information, namely following and follower information. Theses two attributes are represented by two vectors, i.e. following vector $\boldsymbol{Fg}(u_i)$ and follower vector $\boldsymbol{Fr}(u_i)$, respectively. Specifically, let a binary 1-of-*C*-coding vector represent each vector, such that the corresponding bit of following/follower relationship is one and the rest bit is zero. For instance, if user u_i follows user u_j, then the value of the *j-th* component is 1 in the following vector $\boldsymbol{Fg}(u_i)$, otherwise the value is 0. Likewise, If the user u_i is followed by the user u_j, then the value of the *j-th* component is 1 in the follower vector $\boldsymbol{Fr}(u_i)$, otherwise the value is 0.

4.1.1. The similarity of social relation between users

For two microblog users u_i and u_j, the social relation is denoted as $SR(u_i)=(\boldsymbol{Fg}(u_i),\boldsymbol{Fr}(u_i))$ and $SR(u_j)=(\boldsymbol{Fg}(u_j),\boldsymbol{Fr}(u_j))$, respectively. The cosine similarity is adopted as the similarity calculation of social relationship.

The similarity of following information between u_i and u_j can be defined as:

$$sim\left(\boldsymbol{F}g(u_i),\boldsymbol{F}g\left(u_j\right)\right)=\frac{\boldsymbol{F}g(u_i)\cdot\boldsymbol{F}g\left(u_j\right)}{\|\boldsymbol{F}g(u_i)\|\times\left\|\boldsymbol{F}g\left(u_j\right)\right\|} \tag{12}$$

The similarity of follower information between u_i and u_j can be defined as:

$$sim\left(\boldsymbol{Fr}(u_i), \boldsymbol{Fr}\left(u_j\right)\right) = \frac{\boldsymbol{Fr}(u_i) \cdot \boldsymbol{Fr}\left(u_j\right)}{\|\boldsymbol{Fr}(u_i)\| \times \left\|\boldsymbol{Fr}\left(u_j\right)\right\|} \tag{13}$$

The similarity of social relation between u_i and u_j are defined as follows:

$$sim\left(SR(u_i), SR\left(u_j\right)\right) = sim\left(\boldsymbol{F}g(u_i), \boldsymbol{F}g\left(u_j\right)\right) + sim\left(\boldsymbol{Fr}(u_i), \boldsymbol{Fr}\left(u_j\right)\right) \tag{14}$$

The similarity of social relation is normalized to [0,1] as follows:

$$N - sim\left(SR(u_i), SR\left(u_j\right)\right) = \frac{sim\left(SR(u_i), SR\left(u_j\right)\right) - \min sim\left(SR(u_i), SR\left(u_j\right)\right)}{\max sim\left(SR(u_i), SR\left(u_j\right)\right) - \min sim\left(SR(u_i), SR\left(u_j\right)\right)} \tag{15}$$

4.1.2. The social relation matrix of multi-users

The interests of the microblog users can be revealed by the following/follower relationship to some extent. A $N \times N$ social relation similarity matrix $\boldsymbol{M}_{sr}$ of multi-users is defined, where N is the total number of users.

$$\boldsymbol{M}_{sr} = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1N} \\ m_{21} & m_{22} & \cdots & m_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ m_{N1} & m_{N2} & \cdots & m_{NN} \end{bmatrix} \tag{16}$$

The entry m_{ij} in $\boldsymbol{M}_{sr}$ denotes the similarity of social relation between users and it is defined as:

$$m_{ij} = \begin{cases} N - sim\left(SR(u_i), SR\left(u_j\right)\right), & \text{social relation exist} \\ 1, & i = j \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

4.2. Iterative recommendation algorithm

4.2.1. Matrix iteration

The result of $\mathbf{M}_{sr} \times \mathbf{M}_{re}$ is an $N \times n$ matrix, where the i-th row denotes the new weights of tags by considering the original tag weights of neighbors via the following /follower relationship. To obtain a converged tag weight, the process of matrix multiplication is repeated several times.

The matrix $\boldsymbol{M}_x$ denotes the result of the xth iteration, and it can be defined as [57]:

$$\boldsymbol{M}_x = \beta \boldsymbol{M}_{re} + (1 - \beta)\boldsymbol{M}_{sr} \times \boldsymbol{M}_{x-1} \tag{18}$$

where $\boldsymbol{M}_0 = \boldsymbol{M}_{re}$, and $\beta(\beta \in (0,\ 1])$ is a damping factor to determine the importance of updated user-tag matrix and social relation similarity matrix between multi-users. If $\beta > 0.5$, the importance of updated user-tag matrix is more than social relation similarity matrix, and vice versa.

This iteration algorithm is similar to PageRank [3], however, there are still many differences. Firstly, PageRank is a link analysis algorithm and it assigns a numerical weighting to each element of a hyperlinked set of documents, with the purpose of measuring its relative importance within the set. The numerical weight that it assigns to any given element A is referred to as the PageRank of E and denoted by PR(A) as follows:

$$PR(A) = (1 - d) + d(PR(T_1)/C(T_1) + PR(T_2)/C(T_2) \ldots + PR(T_n)/C(T_n)) \tag{19}$$

where the parameter d is a damping factor which can be set between 0 and 1. The damping factor β in our iterative algorithm can be set similarly. Nevertheless, it is generally assumed that the parameter d will be set around 0.85 while β will be set around 0.6 through our experiments. Secondly, the PageRank of a page is obtained through iterative algorithm by calculating the importance of pages that are linked to it. Pages with more pages which point to it are of higher PageRank, and similarly, pages without any pages which point to it is of no PageRank. As Eq. (19) suggests that it assumes page A has pages T_1, $T_2 \ldots T_n$ which point to it (i.e., are citations). However, our algorithm investigates following/follower relationship to calculate the similarity of social relationship between multi-users, and constructs social relation similarity matrix. PageRank is unidirectional while our approach is bidirectional. Thirdly, the convergence of PageRank can be converted to the stationary distribution procedure of Markov chain Model. The most notable feature of Markov chain Model is that the transition probability distribution in each step is just related to the previous. The convergence of PageRank is assumed when the difference between two consecutive value PR_i and PR_{i+1} is less than a predefined threshold while the convergence of our algorithm is assumed when the difference between two successive matrix is less than a predefined threshold or after φ iterations. Then, the user u_i sets its the new tag weight vector $\boldsymbol{V}_i'$ to the i-th row of matrix $\boldsymbol{M}_\eta$.

In summary, we obtain the user-tag matrix via this iteration algorithm, which considers the multi-tag correlation and the social relation between multi-users in an iterative manner.

Algorithm 1 The microblog recommendation algorithm.

Input: *All microblogs* $D_i = \{d_{i1}, d_{i2}, \ldots, d_{iM_i}\}$*posted by each user i,* $i \in \{1, 2, ..., N\}$*, microblogs* $R=\{d_1, d_2, ..., d_M\}$ *to be recommended;*
Output: *The Recommended Microblog sequence for each user;*
1. For users with no tags, add tags based on tag retrieval procedure;
2. Construct the original user-tag matrix $\boldsymbol{M}_{ul}$, and calculate the original user tags weights;
3. Construct the multi-tags correlation matrix $\boldsymbol{M}_{lr}$;
4. Update the original matrix: $\boldsymbol{M}_{re} = \boldsymbol{M}_{ul} \times \boldsymbol{M}_{lr}$;
5. Calculate the similarity of social relation between multi-users;
6. Construct the $N \times N$ user-user social relation similarity matrix $\boldsymbol{M}_{sr}$;
7. Iteratively multiply the Matrix $\boldsymbol{M}_{re}$ with matrix $\boldsymbol{M}_{sr}$ until convergence: $\boldsymbol{M}_x = \beta \boldsymbol{M}_{re} + (1-\beta)\boldsymbol{M}_{sr} \times \boldsymbol{M}_{x-1}$;
8. Define the ranking function $f(u_i, d_p) = \underset{p}{\boldsymbol{E}} \cdot (\mathbf{V}_i')^{\mathrm{T}}$;
9. Predefine γ_i, if a microblog d_p exceeds threshold γ_i, then add it to its corresponding recommendation set.

Table 2
Number of training/test data in 20 categories.

Category	#Train	#Test	Category	#Train	#Test
Sports	41,540	8000	Military	18,790	3000
Technology	29,700	6000	Parenting	25,100	6000
Estate	27,836	6000	Environmental protection	30,053	6000
Stock	23,000	4000	Health	27,820	6000
Emotion	46,703	8000	Travel	35,200	8000
Entertainment	55,430	8000	Medicine	24,991	5000
Political	18,905	3000	Commodity	39,860	8000
Religion	13,680	3000	Education	25,074	5000
Fitness	27,481	6000	Food	37,440	8000
Art	21,530	4000	Home improvement	15,740	3000

4.2.2. Recommendation algorithm

Given a microblog d_p, the ranking function $f(u_i, d_p)$ for a user u_i, is defined as:

$$f(u_i, d_p) = \underset{p}{\boldsymbol{E}} \cdot \left(\boldsymbol{V}_i'\right)^{\mathrm{T}} \tag{20}$$

which denotes the similarity between microblog d_p and user u_i. Each microblog d_p is represented as a vector $\underset{p}{\boldsymbol{E}} = (w_{p1}, w_{p2}, \ldots, w_{pn})$, if d_p contains tag l_j then $w_{pj} = 1$, $w_p = 0$ otherwise. The vector $\boldsymbol{V}_i' = (w_{i1}', w_{i2}', \ldots, w_{in}')$ is the updated tag weight vector for user u_i. The ranking function f is widely used to measure the relevance of microblogs and users. We predefine a threshold γ_i, if$f(u_i, d_p) > \gamma_i$, then the microblog d_p will be recommended to the user u_i.

The pseudo code of the propose algorithm is summarized in Algorithm 1.

5. Experimental results

In this section, we present the experimental results on four tasks, which are designed to explore the performance of our method. First, we test the proposed tag retrieval strategy and compare the performance with other classical methods. Second, we conduct the impact analysis on parameters α and β. Third, we verify the convergence of our iterative algorithm. At last, we compare the recommendation performance of our algorithm with other recommendation algorithms.

5.1. Experimental settings

We evaluate the performance of our algorithm on two real datasets, Sina Weibo and Twitter. The Sina Weibo dataset contains 8524 users with a large number of microblog from March 21th to April 27th, 2014. We preprocess these microblogs, and then the final experimental dataset is constructed. The number of microblogs varies across 20 categories, and there are at least 13,000 microblogs per category. These microblogs are divided into training set and the testing set, respectively, as shown in Table 2. The Twitter dataset is obtained from [57]. As Twitter does not provide the tagging service, tags are retrieved for each user as introduced previously in Section 3. The generally accepted evaluation metrics *Precision, Recall*, and *F-measure* are adopted to evaluate the performance of recommendation. Higher values indicate better recommendation performance.

5.2. Experimental results

The experiments include four parts: 1) Comparison with that of 4 other algorithm on tag retrieval; 2) The impact analysis of the parameter α and parameter β; 3) The Convergence of our iterative algorithm; 4) Comparison with that of other recommendation algorithms.

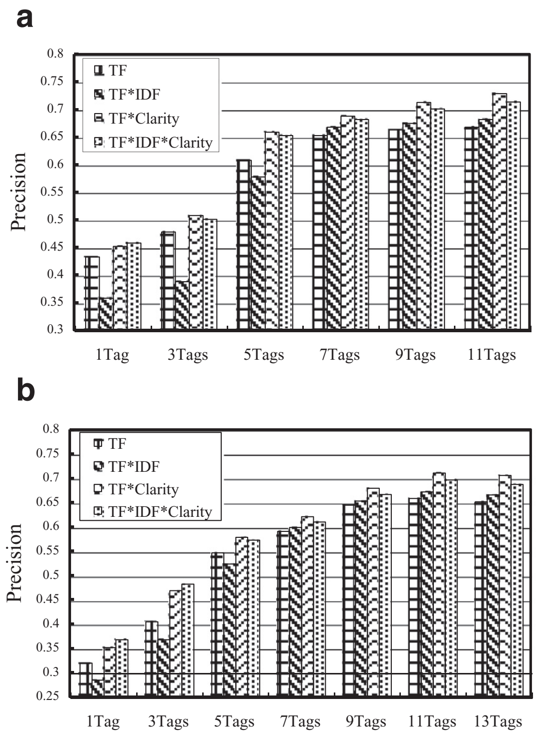

Fig. 3. The *Precision* scores of our recommendation algorithm on different datasets. (a) The Precision scores on Sina Weibo with different number of tags. (b) The Precision scores on Twitter with different number of tags.

5.2.1. Comparison with that of 4 other algorithm on tag retrieval

This experiment is designed to examine the tag retrieval performance of our model and to compare it with some baseline methods. Our method is denoted as *TF*Clarity*, we compare it against the following four methods: *TF, TF*IDF, TF*Clarity, TF*IDF*Clarity*. For each scheme, top {1, 3, 5, 7, 9, 11} words with the highest scores from Sina Weibo and top {1, 3, 5, 7, 9, 11, 13} words with the highest scores from Twitter are selected as users'tags, respectively. The precision scores are reported in Fig. 3.

From Fig. 3(a) and (b), we can make the following observations. First, more tags (from 1 tag to 7 tags in 3(a) and from 1 tag to 9 tags in 3(b)) lead to better recommendation precision. When more than 7 or 9 tags are selected the improvement becomes minor. Second, among all tag selection methods, *TF *Clarity* outperforms the others in most runs with three tags or more tags marked by user. The reason lies in that *TF *Clarity* offers a combination of both importance and topical indication to capture the representative of terms. Finally, when *IDF* is high, the number of keywords matching microblogs will decline, hence *TF*IDF*Clarity* is not an ideal weighting scheme. And we therefore decide to adopt *TF *Clarity* as our tag retrieval strategy in the rest of the experiments.

5.2.2. The impact analysis of the parameter α and β

Parameters α and β are the most important parameters in our method. The parameter α is used to balance the effects of inner and outer tag correlations, while the parameter β is used to determine the importance of updated user-tag matrix and social relation matrix between multi-users. We conduct experiments using different values of parameter α and β on different datasets. We set α as 0.5 when determining the value of parameter β and vice versa. In Figs. 4 and 5, *Precision, Recall* and *F*1 scores are reported to show the performance of our proposed approach, varying along with the value of parameter α and β (from 0 to 1 with a 0.1 step).

Fig. 4 reveals that the experimental results match favourably with our hypotheses and encourage us to further explore the reasons. Firstly, the performance of our algorithm is superior to either considering inner or outer correlation in isolation, indicating that an optimal performance comes from an appropriate combination of both the inner and outer correlation among tags. Secondly, the increase in recommendation is robust across a wide range of mixing proportions. When parameter α equals to 0.5, our recommendation algorithm shows its best performance, which means both inner correlation and outer correlation are equally important. Thirdly, when parameter α equals to 1, the performance of our algorithm is better than that of α equals to 0, suggesting that the explicit correlations between tags across entire users are important than of implicit correlations.

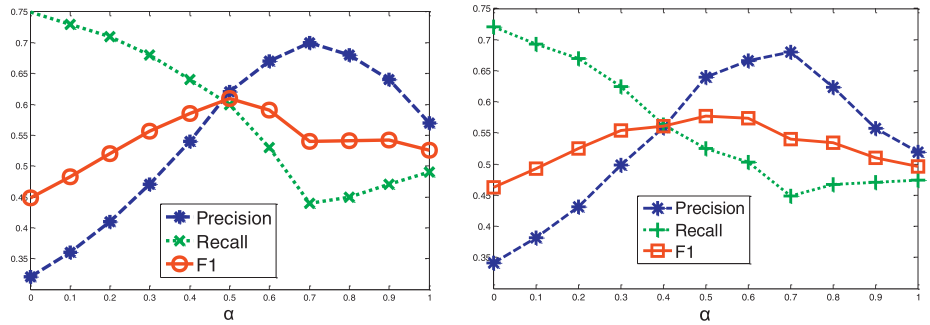

Fig. 4. Relationships between α and recommendation performance. (a) Sina Weibo. (b) Twitter.

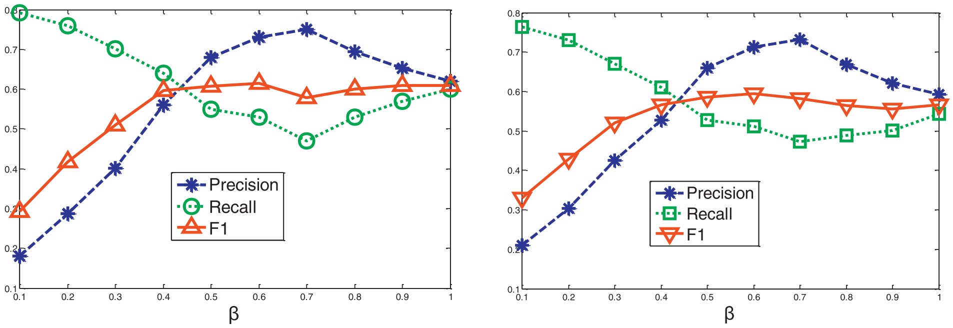

Fig. 5. Relationships between β and recommendation performance. (a) Sina Weibo. (b) Twitter.

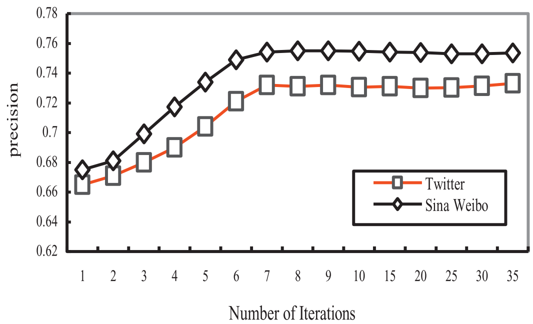

Fig. 6. Convergence of the iterative process on both datasets.

As for the parameter β, Fig. 5 reveals that our recommendation algorithm shows its best performance when $\beta = 0.6$, which means tag correlation and social relation are not equally important. Besides, β cannot be set as 0, that is because if $\beta = 0$, then $\boldsymbol{M}_0 = \boldsymbol{M}_{sr}$, on the one hand, this is not accordance with the initial setting for $\boldsymbol{M}_0 = \boldsymbol{M}_{re}$,; on the other hand, the algorithm only takes users' social relation into consideration which can not reveal users' interest. Besides, if $\beta = 1$, $\boldsymbol{M}_0 = \boldsymbol{M}_{re}$ and it corresponds to the initial value setting, even if users' social relation is not considered, the users' interests can still be investigated. Thus $\beta \in (0, 1]$. Similar results are demonstrated on both datasets.

5.2.3. The convergence of our iterative algorithm

To verify the convergence of the iterative algorithm, the maximum number of iterations φ is set as 35, and the precision of the algorithm with different numbers of iterations is calculated to determine the convergence condition of the algorithm.

As shown in Fig. 6, as the number of iterations increases, the precision of the algorithm sharply increases when $\varphi < 7$ and the precision of the algorithm gradually becomes stable when $\varphi > 7$, which suggests the iteration process of the algorithm tends to converge. Therefore, the convergence of the algorithm can be guaranteed through 7 iterations.

5.2.4. Comparison of our algorithm with other recommendation algorithms

To further demonstrate the effectiveness of our proposed approach, we compare our method *Iterative Tag Correlation And User Social Relation*, denoted as *ITCAUSR*, on different datasets against the following three algorithms: *Naive Approach* (*NA*) (57), *Tag Correlation* (*TC*) [35] and *Tag Correlation And User Social Relation* (*TCAUSR*). In this experiment, all the involved

Table 3
Performance of comparison with different algorithms on our datasets.

	AP	*Precision*	*Recall*	*F1*
NA	0.45	0.531	0.63	0.576
TC	0.49	0.62	0.6	0.609
TCAUSR	0.55	0.675	0.581	0.624
ITCAUSR(Sina Weibo)	0.58	0.754	0.546	0.633
ITCAUSR(Twitter)	0.57	0.732	0.559	0.629

parameters are carefully tuned and the parameters with best performance are used to report the final comparison results. It is observed from Table 3 that our approach achieves the highest performance on both datasets in terms of precision. This observation verifies the effectiveness of our method for microblog recommendation.

The superiority of *ITCAUSR* can be summarized as follows: the user tag retrieval strategy of tag assignment for users can select the most representative words as users' tags. These tags can reveal users' interests in the initial phase. Besides, the inner and outer correlations of these tags are investigated so that the sparsity of original user-tag matrix is reduced and semantic information are involved. What is more, the social relation multiplies with the updated user-tag matrix to accurately denote the user's interest. Last but not the least, an iterative mechanism is introduced to obtain the final user-tag weight. This can accurately reveal the user's interest and get better recommendation results.

6. Conclusions and future work

Microblog has recently become an influential Internet service for information publishing. In this paper, we present a tag correlation and user social relation based approach for recommending microblogs. We have designed the tag retrieval strategy to select proper tags from microblogs for users. Both inner and outer correlations between tags are investigated. The inner correlation captures correlation within tag pairs represented by patterns of explicit tag co-occurrences in the user set, while the outer correlation extracts correlation between tag pairs indicated by implicit couplings between tag pairs through indirectly linked tags. Thereafter the initial user-tag matrix can be updated. The relationship between the social relations of microblog users is also explored. The user–user social relation similarity matrix via investigating following/follower relationship is constructed, which iteratively multiplies with the updated user-tag matrix to obtain the final user tag weights. Future work will explore how to incorporate social influence of users to recommend microblogs.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (no. 61363058, 61163039,61165002), Youth Science and Technology support program of Gansu Province (145RJZA232, 145RJYA259) and the open fund of Key Laboratory of Intelligent Information Processing Institute of computing technology of Chinese Academy of Sciences (IIP2014-4).

References

[1] F. Abel, Q. Gao, G.J. Houben, K. Tao, Analyzing user modeling on Twitter for personalized news recommendations, in: Proc. 19th Int. Conf. User Modeling Adaption and Personalization, UMAP 2011, Girona, Spain, Springer International Publishing, 2011, pp. 1–12.
[2] F. Benevenuto, T. Rodrigues, M. Cha, V. Almeida, Characterizing user navigation and interactions in online social networks, Inf. Sci. 195 (2012) 1–24.
[3] S. Brin, L. Page, The anatomy of a large-scale hypertextual web search engine, Compt. Networks & ISDN Syst. 30 (1-7) (1998) 107–117.
[4] F. Buccafurri, G. Lax, A. Nocera, D. Ursino, Discovering missing Me Edges across social networks, Inf. Sci. 319 (2015) 18–37.
[5] F. Buccafurri, G. Lax, S. Nicolazzo, A. Nocera, A model to support multi-social-network applications, in: Proc.Int. Conf. on the Move to Meaningful Internet Systems, OTM 2014, Amantea, Italy, Springer International Publishing, 2014, pp. 639–656.
[6] W. Bi, J. Kwok, Efficient multi-label classification with many labels, in: Proc. 30th Int. Conf. on Machine Learning, ICML 2013, Atlanta, USA, 2013, pp. 405–413.
[7] A. Cano, J.M. Luna, E.L. Gibaja, S. Ventura, LAIM discretization for multi-label data, Inf. Sci. 330 (C) (2016) 370–384.
[8] X. Cai, M. Bain, A. Krzywicki, W. Wobcke, Y. Kim, P. Compton, A. Mahidadia, Reciprocal and heterogeneous link prediction in social networks, in: Proc. 16th Int. Conf. Advances in Knowledge Discovery and Data Mining, PAKDD 2012, Springer International Publishing, Kuala Lumpur, Malaysia, 2012, pp. 193–204.
[9] Y.N Chen, H.T Lin, Feature-aware label space dimension reduction for multi-label classification, in: Proc. 26th Int. Conf. on Neural Information Processing Systems, NIPS 2012, Lake Tahoe, Nevada, United States, 2012, pp. 1529–1537.
[10] Y.L. Chen, S.Y. Wu, Y.C. Wang, Discovering multi-label temp oral patterns in sequence databases, Inf. Sci. 181 (3) (2011) 398–418.
[11] J. Chen, R. Nairn, L. Nelson, E. Chi, Short and tweet: experiments on recommending content from information streams, in: Proc. 28th Int. Conf. on Human Factors in Computing Systems, CHI 2010, Atlanta, Georgia, 2010, pp. 812–815.
[12] J. Chen, W. Geyer, C. Dugan, M. Muller, I. Guy, Make new friends, but keep the old: recommending people on social networking sites, in: Proc. 27th Int. Conf. on Human Factors in Computing Systems, CHI 2009, Boston, MA, USA, 2009, pp. 201–210.
[13] T. Chen, Combining factorization model and additive forest for collaborative followee Recommendation, KDD-Cup Workshop Publishing, 2012.
[14] W. Cheng, E. Hullermeier, Combining instance-based learning and logistic regression for multilabel classification, Mach. Learn. 76 (2-3) (2009) 211–225.
[15] B. Chikhaoui, S. Wang, T. Xiong, H. Pigot, Pattern-based causal relationships discovery from event sequences for modeling behavioral user profile in ubiquitous environments, Inf. Sci. 285 (2014) 204–222.
[16] K. Dembczynski, W.W. Cheng, E. Hullermeier, Bayes optimal multilabel classification via probabilistic classifier chains, in: Proc. 27th Int. Conf. on Machine Learning (ICML-10), ICML 2010, Haifa, Israel, 2010, pp. 279–286.
[17] M.P. De, Ferrara E., D. Rosaci, G.M. Sarne, Trust and compactness in social network groups, IEEE Trans. Cybern. 45 (2) (2015) 205–216.

[18] V.D. Graaff, A.V. Venis, M.V. Keulen, R.A. By, Generic knowledge-based analysis of social media for recommendations, in: Proc. 2nd Conf. on New Trends on Content-Based Recommender Systems co-located with 9th ACM Conference on Recommender Systems (RecSys 2015), 2015, Vienna, Austria, 2015, pp. 1712–1723.
[19] Y. Guo, S. Gu, Multi-label classi-fication using conditional dependency networks, in: Proc. Int. Conf. on Artificial Intelligence, IJCAI 2011, Catalonia, Spain, 2011, pp. 1300–1305.
[20] I. Guy, I. Ronen, E. Wilcox, Do you know? recommending people to invite into your social network, in: Proc. Int. Conf. Intelligent User Interfaces, ACM 2009, Sanibel Island, Florida, 2009, pp. 77–86.
[21] S. Guimarães, M.T. Ribeiro, R. Assunção, W. Meira, A holistic hybrid algorithm for user recommendation on Twitter, J Inf. Data Manage. 4 (3) (2013) 341–356.
[22] J. Hannon, K. McCarthy, B. Smyth, Finding useful users on Twitter: Twittomender the followee recommender, in: Proc. Int. Conf. Advances in Information Retrieval 33rd European Conference on IR Research, ECIR 2011, Dublin, Ireland, Springer International Publishing, 2011, pp. 784–787.
[23] Y. Jiang, Z. Deng, K.S. Choi, F.L. Chung, S. Wang, A novel multi-task TSK fuzzy classifier and its enhanced version for labeling-risk-aware multi-task classification, Inf. Sci. 357 (2016) 39–60.
[24] I. Kahanda, J. Neville, Using transactional information to predict link strength in online social networks, in: Proc. 3rd Int. Conf. on Weblogs and Social Media, ICWSM 2009, San Jose, California, USA, 2009, pp. 74–81.
[25] D.P. Karidi, From user graph to Topics Graph: Towards Twitter followee recommendation based on knowledge graphs, in: Proc. IEEE Int. Conf. on Data Engineering Workshops, ICDE Workshops 2016, IEEE Computer Society, Helsinki, Finland, 2016, pp. 121–123.
[26] J. Lee, D.W. Kim, Memetic feature selection algorithm for multi-label classification, Inf. Sci. 293 (2015) 80–96.
[27] J. Lee, H. Kim, N.R. Kim, J.H. Lee, An approach for multi-label classification by directed acyclic graph with label correlation maximization, Inf. Sci. 351 (C) (2016) 101–114.
[28] J. Leskovec, D. Huttenlocher, J. Kleinberg, Predicting positive and negative links in online social networks, In: Proc. 19th Int. Conf. on World Wide Web, ACM 2010, Raleigh, USA, 2010, pp. 641–650.
[29] H.Y. Lo, S.D. Lin, H.M. Wang, Generalized k-Labelsets ensemble for multi-label and cost-sensitive classification, IEEE Trans. Knowl. Data Eng 26 (7) (2014) 1679–1691.
[30] Y. Lin, Q. Hu, J. Zhang, X. Wu, Multi-Label feature selection with streaming labels, Inf. Sci. 372 (2016) 256–275.
[31] Z.Y. Liu, X.X. Chen, M. Sun, Mining the interests of chinese microbloggers via keyword extraction, Front. Comput. Sci. 6 (1) (2012) 76–87.
[32] D.R. Liu, P.Y. Tsai, P.H. Chiu, Personalized recommendation of popular blog articles for mobile applications, Inf. Sci. 181 (9) (2011) 1552–1572.
[33] W.Y. Liu, X.J. Quan, M. Feng, B. Qiu, A short text modeling method combining semantic and statistical information, Inf. Sci. 180 (20) (2010) 4031–4041.
[34] H. Liang, Y. Xu, Y. Li, R. Nayak, X. Tao, Connecting users and items with weighted tags for personalized item recommendations, in: Proc. 21st ACM Conf. on Hypertext and Hypermedia, ACM 2010, Toronto, Ontario, Canada, 2010, pp. 51–60.
[35] H.F. Ma, M.H.Z. Jia, M. Xie, X.H. Lin, A microblog recommendation algorithm based on multi-tag correlation, in: Proc. 8th Int. Conf. Knowledge Science, Engineering and Management, KSEM2015, Chongqing, China, Springer International Publishing, 2015, pp. 483–488.
[36] P.D. Meo, G. Quattrone, D. Ursino, A query expansion and user profile enrichment approach to improve the performance of recommender systems operating on a folksonomy, User Model. User-Adapt. Interact. 20 (1) (2010) 41–86.
[37] P.D. Meo, G. Quattrone, D. Ursino, Exploitation of semantic relationships and hierarchical data structures to support a user in his annotation and browsing activities in folksonomies, Inf. Sys. 34 (6) (2009) 511–535.
[38] J. Read, A. Puurula, A. Bifet, Multi-label classification with meta-labels, in: Proc. IEEE Int. Conf. on Data Mining, ICDM 2014, Shenzhen, China, 2014, pp. 941–946.
[39] T. Ma, Y. Yang, L. Wang, B. Yuan, Recommending people to follow using asymmetric factor models with social graphs, KDD-Cup Workship, 2012.
[40] J. Read, F. Perezcruz, Deep Learning for Multi-label Classification, Mach. Learn. 85 (3) (2014) 333–359.
[41] S. Rendle, L. Schmidt-Thieme, Pairwise interaction tensor factorization for personalized tag recommendation, in: Proc. 3rd ACM Int. Conf. on Web Search and Data Mining, WSDM 2010, New York, USA, 2010, pp. 81–90.
[42] X.D. Song, B.L. Tseng, C.Y. Lin, M.T. Sun, Personalized recommendation driven by information flow, in: Proc. 29th annual ACM SIGIR Int. Conf. on Research and development in information retrieva, ACM 2006, Washington, USA, 2006, pp. 509–516.
[43] P.N.D. Silva, E.C. Gonçalves, A. Plastino, A.A. Freitas, Distinct chains for different instances: an effective strategy for multi-label classifier chains, in: Proc. European Conf. Machine Learning and Knowledge Discovery in Databases, ECML PKDD 2014, Nancy, France, 2014, pp. 453–468.
[44] A. Sun, Short text classification using very few words, in: Proc. 35th Int. ACM SIGIR Conf. on Research and development in information retrieval, SIGIR 12, Portland, USA, 2012, pp. 1145–1146.
[45] C.C. Tu, Z.Y. Liu, M.Y. Sun, Tag Correspondence model for user tag suggestion, J. Comput. Sci. Technol. 30 (5) (2015) 1063–1072.
[46] J.L. Tang, X.F. Wang, H.J. Gao, X. Hu, H. Liu, Enriching short text representation in microblog for clustering, Front. Comput. Sci. 6 (1) (2012) 88–101.
[47] T. Takata, S. Ozawa, A neural network model for learning data stream with multiple class labels, in: Proc. 10th Int. Conf. on Machine Learning and Applications and Workshops, ICMLA 2011, HONOLULU, Hawaii, USA, 2011, pp. 35–40.
[48] R. Wetzker, C. Zimmermann, C. Bauckhage, S. Albayrak, I tag, you tag: translating tags for advanced user models, in: Proc. 3rd ACM Int. Conf. on Web search and data mining, WSDM 2010, New York, 2010, pp. 71–80.
[49] W. Wu, B. Zhang, M. Ostendorf, Automatic generation of personalized annotation tags for Twitter users,, in: Proc. Conf. Human Language Technologies, the North American Chapter of the Association of Computational Linguistics, HLT-NAACL 2010, California, USA, 2010, pp. 689–692.
[50] F. Wu, Y Song, Y. Huang, Microblog sentiment classification with heterogeneous sentiment knowledge, Inf. Sci. 373 (2016) 149–164.
[51] R. Xia, C. Zong, S. Li, Ensemble of feature sets and classification algorithms for sentiment classification, Inf. Sci. 181 (6) (2011) 1138–1152.
[52] R. Xiang, J. Neville, M. Rogati, Modeling relationship strength in online social networks, in: Proc. 19th Int. Conf. on World wide web, WWW 2010, Raleigh, North Carolina, USA, 2010, pp. 981–990.
[53] Z.M Xu, D. Li, T. Liu, S. Li, G. Wang, S.L. Yuan, Measuring similarity between microblog users and its application, Chin. J Comput. 37 (1) (2014) 207–218.
[54] Q. Yan, L. Wu, L. Zheng, Social network based microblog user behavior analysis, Physica A 392 (7) (2013) 1712–1723.
[55] R. Yan, M. Lapata, X. Li, Tweet recommendation with graph co-ranking,, in: Proc. 50th Annual Meeting of the Association for Computational Linguistics, ACL 2012, Jeju Island, 2012, pp. 516–525.
[56] Z. Yu, C. Wang, J. Bu, X. Wang, Y. Wu, Friend recommendation with content spread enhancement in social networks, Inf. Sci. 309 (2015) 102–118.
[57] X.K. Zhou, S. Wu, C. Chen, G. Chen, S.S. Ying, Real-time recommendation for microblogs, Inf. Sci. 279 (2014) 301–325.

Huifang Ma is currently an associate Professor in the College of Computer Science and Engineering at Northwest Normal University, P. R. China. She received the B.E. degree from Northwest Normal University, P. R. China., in 2003, and the M.S. degree from Beijing Normal University, P. R. China., in 2006. She received the Ph.D. degree from Institute of Computing Technology, Chinese Academy of Sciences, in 2010. Her general area of research is data mining and machine learning.

Meihuizi Jia is currently a postgraduate student in the College of Computer Science and Engineering at Northwest Normal University, P. R. China. She received the B.E. degree from Northwestern Polytechnical University Ming De College, P. R. China., in 2013. Her general area of research is text mining and machine learning.

Di Zhang is currently a postgraduate student in the College of Computer Science and Engineering at Northwest Normal University, P. R. China. He received the B.E. degree from Northwest Normal University, P. R. China in 2015. His general area of research is text modeling and graph modeling.

Xianghong Lin is currently a Professor in the College of Computer Science and Engineering at Northwest Normal University, P. R. China. He received the B.E. degree from Northwest Normal University, P. R. China., in 1998. And he received the M.S. degree and the Ph.D. degree from Harbin Institute of Technology, P. R. China., in 2004 and 2009 respectively. His general area of research is simulation and evolution of large-scale spiking neural networks.