Available online www.jocpr.com

Journal of Chemical and Pharmaceutical Research, 2014, 6(7):679-684

Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

Social group recommendation using topic models

[1]Jing Wang and [2]Hui Zhao

[1]Xidian University, Xi'an, China
[2]Xi'an Jiaotong University, Xi'an, China

ABSTRACT

In this paper, we take the user influence and their relationship into account to propose a new group recommendation method, which utilizes the topic oriented analysis to calculate the user influence on specific topic and to construct topical sub-groups. Based on these topical sub-groups, different user factors are used to depict the topical sub-group characteristics more properly and to calculate the user influence (including individual and social) on the topical sub-groups. Experimental results demonstrate the effectiveness and efficiency of the proposed method, and also show that the proposed method can improve the prediction accuracy of the group recommendation.

Key words: Group recommendation; topic oriented; user influence analysis; social networks

---

INTRODUCTION

Nowadays, social networks[1,2] are commonly used in our life and work, such as Facebook, Twitter, and MySpace. They provide invaluable contextual information concerning the preferences of individuals, e.g. their favorite books, music, movies, and restaurants, as well as their relationships with the others. In social networks, we know that the users have impact on each other. This is usually referred to emotional infection, which is proportional to the trust between the people, the more the people you trust, the greater influence he/she is on you.

With the development of social network services [2,3] like micro-blogging, many recommendation systems have recommend item to a group of users, called as group recommendation system. However, we know that the social network services always contain large amounts of information on different topics. For example, some users are interested in sports, so they will talk about or discuss NBA in social networks; while other people may be interested in movies, they will discuss more about directors or actors. Of course, user may have more than one interests and hobbies. We can know that a user may have different influences on different topics. There has been some researches show that the user influence in social may be different from different topics. For example, in social networks, A can have high influence to B about NBA, but B may have even higher influence to A on movies. So the recommendation system will not recommend appropriate items to user without topical user influence analysis in large-scale social networks.

In this paper, we take the user influence and their relationship into account to propose a new group recommendation method, which utilizes the topic oriented analysis to calculate the user influence on specific topic and to construct topical sub-groups. Based on these topical sub-groups, different user factors are used to depict the topical sub-group characteristics more properly and to calculate the user influence (including individual and social) on the topical sub-groups.

RELATE WORKS

SOCIAL RECOMMENDATION

Goyal et al. used the influence strength (or called probability) of user's historical log actions[4]. They proposed two concepts, user's impact probability and the behavioral probability. The purpose of research the influence probability is

to find a model (static or dynamic), which modeling for the user's influence as well as for user's behavioral influence. Goyal's method can handle large social network, but it ignores behavior correlation between users and the node attributes of user itself. To address this situation, Provost, etc. proposed social behavior tracking, mainly studying how to modeling with the social network structure, user attributes and user behavior [5]. Xiang etc. proposed a latent variable model, which is based on the similarity of user's profile and user interaction activity-based strength of the relationship. The purpose of the proposed model is to find a strong relationship among the user network [6].

GROUP RECOMMENDATION

With the emergence of group behavior, there are some studies for group recommendation. Reference [7] proposed an algorithm to recommend appropriate and novel content to groups of people. Based on the Power Balance Map and the Behavioral Tendency of each group, the algorithm recommends new content in or near high-density areas on the group's feature space. Reference [8] took the advantage of differences and the correlation of the dependencies between users in the group into consideration to formalize the notion of a consensus function, which achieves a balance between an item's aggregate relevance to the group and individual member's disagreements over the item. They designed and implemented efficient threshold algorithms to compute group recommendations.

TOPIC LEVEL RECOMMENDATION

Reference [9] presented a novel approach iTop to discover topic-centric interaction based communities on Twitter. They evaluated the discovered communities along three dimensions: graph based (node-edge quality), empirical-based (Twitter lists) and semantic based (frequent n-grams in tweets), to discover topic-centric, interaction based communities on Twitter. Reference [10] discussed how they can extend probabilistic topic models to analyze the relationship graph of popular social-network data, so that they can "group" or "label" the edges and nodes in the graph based on their topic similarity. They first applied the well-known Latent Dirichlet Allocation (LDA) model and its existing variants to the graph-labeling task and argued that the existing models do not handle popular nodes (nodes with many incoming edges) in the graph very well. They then proposed possible extensions to the model to deal with popular nodes.

THE DEFINITION OF INFLUENCE TOPIC MODEL

THE MODEL DEFINITION

How to leverage both node-specific topic distribution and network structure to quantify social influence? In another word, a user's influence on others not only depends on their own topic distribution, but also relies on what kinds of social relationships they have with others. The goal is to design a unified approach to utilize both the local attributes (topic distribution) and the global structure (network information) for social influence analysis.

Definition (1): Given a network $G=(V,E)$, where *V* is the set of nodes (users, entities) and *E* is the set of directed edges, 2) *T*-dimensional topic distribution $\theta_v \in R^T$ for all node *v* in *V* , how to find the topic-level influence network $G_z=(V_z,E_z)$ for all topics $1 \le z \le T$? Here V_z is a subset of nodes that are related to topic *z* and E_z is the set of pair-wise weighted influence relations over V_z .

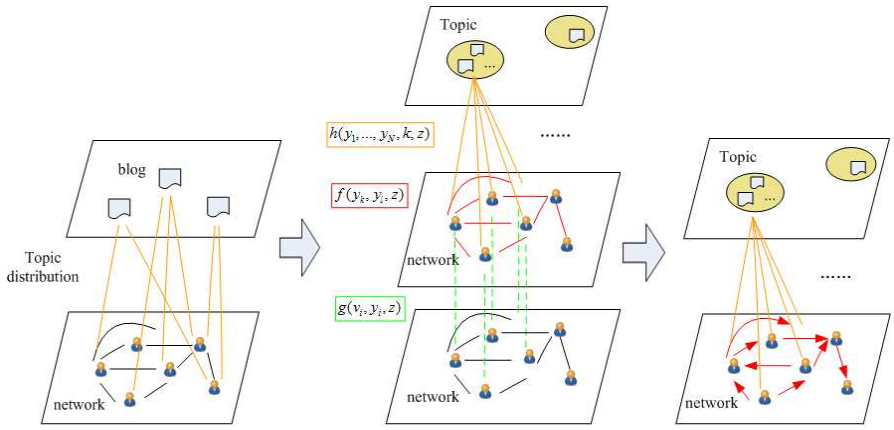

Fig. 1: The topic-level influence network

We formally define the improved TFG model tofind the topic-level influence network $G_z=(V_z,E_z)$ for all topics z .Based on this formulation, the task of social influence is to identify which node has the highest influence on the others on a specific topic along with the edge. That is, to maximize the likelihood function $P(V,Y)$, so we give the

definition of improved TFG model as following. Joint Distribution definition：

$$P(V,Y)=\frac{1}{Z}\prod_{k=1}^{N}\prod_{z=1}^{T}h(y_1,...,y_N,k,z)\cdot\prod_{i=1}^{N}\prod_{z=1}^{T}g(v_i,y_i,z)\cdot\prod_{e_{kl}\in E}\prod_{z=1}^{T}f(y_k,y_l,z) \quad (1)$$

where $V=[v_1,...,v_N]$ and $Y=[y_1,...,y_N]$ corresponds to all observed and hidden variables, respectively; $g(\cdot)$ and $f(\cdot)$ are the node and edge feature functions; $h(\cdot)$ is the global feature function; all these functions will be introduced detailedly in later. *Z* is a normalizing factor.

FEATURE COMPUTATION

Given a social network $G=(V,E)$, the goal is to find topic-level user influence. Based on the characteristics of the group, we can reflect different influence for the different types of groups on the theme of social.In this paper, we use four modules to represents the characteristics of the group, defined as: Leader, Expert, Social and Similarity. The four modules mainly are divided into two dimensions:

(a) Leader and Expert describe the single-user, and the Social and Similarity describe relationship between different users.

(b) Leader and Similarity describe user influence depend on the content, while the Leader and Social are based on social structure.

According to the different of group's activities or group gathering, some modulesmay be more important than the others. For instance, in certain cases, for the characterization of the group, Expert is more important.

NODE FEATURE FUNCTION DEFINITION

Node feature function $g(v_i,y_i,z)$ is defined as distribution calculation for node v_i specific to topic *z*. Their quantitative calculations include: the user's behavior on the micro-blogs, $UM(v,z)$ and the user's influence on the group, $UG(y_i)$. Therefore, we can obtain the following equation.

$$g(v_i,y_i,z)=UG(y_i)\cdot UM(v_i,z) \quad (2)$$

In the micro-blog, a user can publish, brows, or follow micro-blogs, so we can get users' influence related with his published micro-blogs. The more micro-blogs the user publishes, the more active the user will be. We get the user activity $UM(v,z)$ by following formula:

$$UM(v_i,z)=d^{z}_{v_i,v_j}+(1-d^{z}_{v_i,v_j})\sum_{v\subseteq B(u)}A_{(v_i,v_j)}UM(v_j,z) \quad (3)$$

$$d^{z}_{v_i,v_j}=\frac{M_{v_i,v_j}}{\sum M_{v_i}}\cdot\frac{C^{AT}_{v_iz}+\alpha}{\sum_{z'}C^{AT}_{v_iz'}+T\alpha} \quad (4)$$

$$A_{(v_i,v_j)}=a_u\Big/\sum_{n=1}^{N}a_n,\ a_u=n_u/T \quad (5)$$

In Equation (4), $d^{z}_{v_i,v_j}$ is probability of user v_i forward the user v_j 's micro-blogs on topic *z*, where $M_{v_i,v_j}\Big/\sum M_{v_i}$ represents the probability of user v_i forward the user v_j 's micro-blogs; $(C^{AT}_{v_iz}+\alpha)\Big/(\sum_{z'}C^{AT}_{v_iz'}+T\alpha)$ represents the probability of user v_i on topic *z*; $C^{AT}_{v_iz}$ stands for total number of users assigned to topic *z*, α is the Dirichlet distribution over parameters. In Equation (5), $A_{(v_i,v_j)}$ is ratio of the value of UM, the user *v* assigned to the user *u*, which depend on the user *u*'s user activity and account of his all friends activity. a_u is the user's activity; n_u is the number of micro-blog user published; *T* is a uniform time scale, which can be able

more objective portrayal of the user's level of activity. $UG(y_i)$ stands for the user's influence on the group, which introduce the different group feature to users' behavior for the micro-blogs.

$$UG(y_i) = \sum_{u \in G} (desc_{leader}(u) + desc_{expert}(u)) \quad (6)$$

$$desc_{leader}(u) = \frac{\max(|N_u|)}{(\sum_{u \in G} |N_u|)/|G|} \quad (7)$$

$$desc_{expert}(u) = \sum_{u \in G} |M_u| / |G| \quad (8)$$

where $|N_u|$ refers to the number of friends of user *u* in the group, and $|M_u|$ is the number of the user's favorite micro-blogs on topic *z*; $|G|$ refers to the number of all members in the group.

EDGE FEATURE FUNCTION DEFINITION

Edge feature function $f(y_k, y_l, z)$ is defined as the edge of the input network specific to topic *z*:

$$f(y_k, y_l, z) = UM(v_k, z) \cdot UM(v_l, z) \cdot \{\sum_{u \in G} (desc_{similar}(y_{kl}, z) + desc_{social}(y_{kl}))\} \quad (9)$$

$$desc_{similar}(y_{kl}, z) = \sum_{y_{kl} \in G} Fr(y_{kl}, z) \Big/ |Pairs(G)| \quad (10)$$

$$desc_{social}(y_{kl}) = \sum_{y_{kl} \in G} Ps(y_{kl}) \Big/ |Pairs(G)| \quad (11)$$

where, $Fr(y_{kl})$ stands for the number of favorite micro-blogs of the user v_k and v_l on topic *z*; y_{kl} is the edge between v_k and v_l; $UM(y_{kl}, z)$ represents the probability of v_l's micro-blogs followed by user v_k; $Ps(y_{kl})$ is the number of same friends of v_k and v_l in the group; $|Pairs(G)|$ refers to the number of all friend members in the group.

GLOBAL FEATURE FUNCTION DEFINITION

Global feature function $h(a = y_1, ..., y_N, k, z)$ is a feature function defined on all nodes of the input network to topic *z*:

$$P(z_i = j, x_i = y_n \mid w_i = k, \overline{z_i}, \overline{x_i}) = \frac{C_{kj}^{WT} + \beta}{\sum_{k'} C_{k'j}^{WT} + V\beta} \frac{C_{yj}^{AT} + \alpha}{\sum_{j'} C_{yj'}^{AT} + T\alpha} \quad (12)$$

where, $z_i = j, x_i = y_n$ represents the *i*-th word in article, assigned to the *j*-th topic and *y*-th author; $w_i = k$ means that *i*-th word is the *k*-th word in dictionary. In addition $\overline{z_i}, \overline{x_i}$ refer to the rest of the *i*-th for topic and author assignments. C_{kj}^{WT} is the total number assigned to topic *j* before assigning word *k*; C_{yj}^{AT} is the total number for author y_n assigned to topic *j*.After get the probability of a word belonging to a topic, we can get which topic the micro-blog is belonging to.

RESULTS

DATA SETS

In this paper, we get experimental data from Sina micro-blog. Using Sina micro-blog API, we choose a start seed and then crawl data, and finally get a total of 40,000 users, 2.5 million micro-blogs, and 120,000 users relationships.

Table 1: keywords of the topics

topic1: "Obama"	topic 2: "Family"	topic 3: "Military Operations"	topic 4: "Imprisonment"
presid (0.0146) obama (0.0143) govern (0.0113) countri (0.0103) offici (0.0094) secur (0.0092) militari (0.0082) year (0.0080) minist (0.0079) state (0.0077)	women (0.0272) school (0.0160) children (0.0157) year (0.0149) old (0.0149) woman (0.0126) men (0.0124) man (0.0121) girl (0.0109) student (0.0103)	cia (0.0130) report (0.0112) million (0.0107) work (0.0103) compani (0.0100) money (0.0095) oper (0.0093) blackwat (0.0082) use (0.0077) state (0.0071)	prison (0.0194) court (0.0137) case (0.0104) charg (0.0099) releas (0.0092) detaine (0.0083) tortur (0.0080) alleg (0.0072) investig (0.0069) guantanamo (0.0061)

After data preprocessing, we get a small social group (as **whole group**) from this data set composed of 21 users, 172 relationships and more than 500 blogs with 4 topics. On average, each user has 2 interested topics, 10 friends and 25 blogs. Then we classify it into four sub-groups by topic. Table 1 shows the name of the topics and keywords.

EVALUATION MEASURES

In group recommendation system, recommendation quality metrics commonly use group mean absolute error(GMAE). The smaller the GMAE is, the better recommendation is.

$$GMAE = \sum_{m \in M} | pred(G,m) - r(G,m) | / | M | \quad (13)$$

where, $| M |$ is the size of the test set. In our experiments, to evaluate our proposed methods for group recommendation, we predict an item to a topical sub-group compared with the whole group. In order to verify the effectiveness of the improved model TFG, we use the user preferences found in testing for the accuracy by the follow four algorithms, and thenwe compare the actual user rate with the results of the test. Table 2 shows the results comparisons.

①Base Line method(TFG): The algorithm ignores the influence among users.

②Social Model method: it only considers the user influence based on node feature function.

③Link Model method: it only considers the user influence based on edge feature function.

④Full Model method: it considers various topics based on global feature function.

⑤Proposed model: combining the methods of □□□.

Table 2: Comparison of GMAE results

Methods	GMAE				
	Topic1	Topic2	Topic3	Topic4	whole group
①	1.1354	1.1157	1.1403	1.1890	1.5369
②	1.0166	0.9531	0.8487	0.9899	1.4832
③	1.0348	0.9742	0.9562	0.8795	1.4634
④	0.9241	*0.8557*	0.9797	1.1407	1.3945
⑤	**0.8231**	**0.8145**	**0.7846**	**0.8056**	**1.2715**

From Table 2, we can see that the results of all topical sub-groups are better than that of the whole group. This is because each sub-group focuses on a special topic, it will be more better when recommending a topical blog to the sub-group, which shows that leveraging the topic oriented user influence analysis is a more efficient method. Besides,

we also can see that all our methods (②③④) are much better than TFG on all topics. This confirms that four user modules can calculate the user influence on the topical sub-groups more accurately. The all-factors achieve even better result compared to others, which prove that it can depict the topical sub-group characteristics more properly.

CONCLUSIONS

In this paper, we proposed a new group recommendation method based on the topic oriented user analysis to obtain the user influence on specific topic. Furthermore, four user modules are used to calculate the user influence (including individual and social) on the topical sub-groups, which can depict the topical sub-group characteristics more properly. At last, we give our group recommendation algorithm utilizing user influence on specific topic and user modules, and the experimental results show that our method outperforms than the others.

As for future work, we intend to improve our method by using more sensible functions to formalize the user influence on specific topic and topical sub-group. Moreover, we will evaluate our method with real groups and data.

Acknowledgments
This research project was promoted by The National Natural Science Foundation of China (No. 61202177); The Fundamental Research Funds for the Central Universities.

REFERENCES

[1] S Aral; D Walker, *Science*, vol.337, no.6092, pp.337-341, July **2012**.
[2] Q Gao; F Abel; GJ Houben; Y Yu, A comparative study of users' microblogging behavior on Sina Weibo and Twitter. User *Modeling, Adaptation, and Personalization*, pp.88-101, **2012**.
[3] H Kwak;C Lee;H Park; S Moon, What is Twitter, a social network or a news media?. Proceedings of the 19th International Conference on World Wide Web, pp.591-600, **2010**.
[4] A Goyal;F Bonchi; LV Lakshmanan, Learning influence probabilities in social networks. Proceedings of the third ACM International Conference on *Web Search and Data Mining*, pp.241-250, **2010**.
[5] F Provost;B Dalessandro;R Hook;X Zhang;A Murray,Audience selection for on-line brand advertising: privacy-friendly social network targeting. Proceedings of the 15th ACM SIGKDD International Conference on *Knowledge Discovery and Data Mining*, pp.707-716, **2009**.
[6] R Xiang;J Neville; M Rogati, Modeling relationship strength in online social networks. Proceedingsof the 19th International Conference on World Wide Web, pp.981-990, **2010**.
[7] S Seko;T Yagi;M Motegi;S Muto, Group recommendation using feature space representing behavioral tendency and power balance among members. Proceedings of *the fifth ACM conference on Recommender Systems*, pp.101-108, **2011.**
[8] AY Sihem; R Senjuti Basu; C Ashish; D Gautam; Y Cong,Group recommendation: Semantics and efficiency. *Proceedings of the VLDB Endowment*,vol.2, no.1, pp.754-765, **2009**.
[9] D Correa;A Sureka;M Pundir,iTop: interaction based topic centric community discovery on twitter. Proceedings of the 5th Ph. D. workshop on Information and knowledge, pp.51-58, **2012**.
[10] YC Cha; J Cho, Social-network analysis using topic models. Proceedings of the 35th international ACM SIGIR conference on research and development in information retrieval, pp.565-574, **2012.**