Identifying Domain-Dependent Influential Microblog Users: A Post-Feature Based Approach

Nian Liu, Lin Li
School Of Computer Science & Technology
Wuhan University Of Technology
Wuhan 430070, China
{liunian, cathylilin}@whut.edu.cn

Guandong Xu
Advanced Analytics Institute
University of Technology, Sydney
NSW 2007, Australia
Guandong.Xu@uts.edu.au

Zhenglu Yang
College Of Computer & Control Engineering
Nankai University
Tianjin 300071, China
thxlifeyzl@gmail.com

Abstract

Users of a social network like to follow the posts published by influential users. Such posts usually are delivered quickly and thus will produce a strong influence on public opinions. In this paper, we focus on the problem of identifying domain-dependent influential users(or topic experts). Some of traditional approaches are based on the post contents of users users to identify influential users, which may be biased by spammers who try to make posts related to some topics through a simple copy and paste. Others make use of user authentication information given by a service platform or user self description (introduction or label) in finding influential users. However, what users have published is not necessarily related to what they have registed and described. In addition, if there is no comments from other users, its less objective to assess a users post quality. To improve effectiveness of recognizing influential users in a topic of microblogs, we propose a post-feature based approach which is supplementary to post-content based approaches. Our experimental results show that the post-feature based approach produces relatively higher precision than that of the content based approach.

Introduction

Recently, microblogs as a new social media have attracted researchers interests (Kwak et al. 2010). Compared with the traditional media, microblogs have many distinguished characteristics, such as rich information sources, quick transmission, large influence range, timeliness, and the strong interaction between users, and thus it is easy to form a hot topic. Like word-of-mouth diffusion, some users can provide valuable information as important messengers or often put forward original ideas for important events, which attracts more other users to participate in discussions, and even affects the publics viewponts. Such users are so called influential users. Therefore, identifying the influential users has important practical significance in microblogs. It is conducive to track the key characters in a hot topic, and to provide supports for microblog analysis (Ghosh et al. 2012).

A number of recent papers have addressed the matter of diffusion on networks in general, and the attributes and roles of influencers specifically (Bakshy et al. 2011). Identifying the influential users in a field is equivalent to judge topic experts. Usually, there are two main ideas. One is based on the information provided by a user herself (such as, authentication information or user self description) (Sullivan 2011). The other is analyzing microblog post contents of users. But the two ideas have limitations to judgeinfluential users. Firstly, though some users have provided authentication information or self description to show that they are experts in some topics, what they have published is not necessarily related to such topic fields. Secondly, if we rely on the post contents of a user, the results may be biased by spammers who try to relate posts and topics through a simple copy and paste. Also, text processing will greatly reduce efficiency of recognizing influential users. To overcome the above difficulties, in this paper we propose a post-feature based approach which mainly utilizes the nine kinds of post characteristics of users in microblogging.

A Post-feature Based Approach

Features of posts in this paper do not include user's authentication information, user's self description and user's post content. We summarize nine kinds of features as shown in Table . They are widely used in a social network and can reflect how users interact. For example, "Follwer number" means how many users like to read the posts of a user and follow her updates. In some extent, it shows her personal populairty. In this paper, we use three different ways to calculate user influence by aggregating the aforementioned nine features, i.e., score based aggregation, list based aggregation, SVM based aggregation.

Score Based Aggregation

Score based aggregation uses normalization formula to calculate microblog users scores of each kind of features as shown in Table . We sum nine scores together and get the total score. Finally, we order users by their total scores. The score is computed as

$$Score_{m,n} = \frac{p_{mn} - p_{min}}{p_{max} - p_{min}}, \quad (1)$$

where *score_m,n* represents the score of a feature *n* of a user *m*. *p_mn* represents the value of feature *n* of user *m*. *p_min* represents the minimum value of a feature *n* in all users. *p_max* represents the maximum value of a feature *n* in all users.

	Feature	Feature Description
1	Follower number	The follower number of a user
2	Attention number	The attention number of a user
3	Mutual follower number	The number of users to follow each other
4	Post number	The number of microblog posts that user have originally published over a period of time
5	Average agreement number	The average agreement number of user microblog posts over a period of time
6	Average retweet number	The average retweet number of user microblog posts over a period of time
7	Average comment number	The average comment number of user microblog posts over a period of time
8	Forwarded post number	The forwarded number of microblog posts that user have published over a period of time
9	Total microblog number	The total number of microblog posts that user have published over a period of time

Table 1: List of microblog user features

List Based Aggregation

According to the values of nine kinds of features, we can respectively get nine ranking lists of all users. Then we adopt sorting fusion method(e.g., Borda Count (Borda 1781)) to get total ranking list of all users. The borda count is a single-winner election method in which voters rank candidates in order of preference. It determines the winner of an election by giving each candidate a certain number of points corresponding to the position in which he or she is ranked by each voter. Once all votes have been counted the candidate with the most points is the winner. We sum nine ranking lists together and get the total ranking list of all users.

SVM Based Aggregation

By using SVM, we randomly devide our whole data set into training set (60%) and testing set(40%). Some students in our team manually give their answer in judgeing whether a user is influential or not in a topic. For cross validation, SVM is run three times.

Experiments

Dataset

We extract dataset from sina microblog (weibo.com). It contains 90 users in IT field from October 19, 2013 to November 18, 2013. The dataset is composed of three types: 30 users with service platform authentication information, 30 users with self description of IT related words, 30 users without any identifiers.

Experimental Results And Discussions

We compare our post-feature based approach with a baseline which identifies influential microblog users by service platform authentication information or user self description. The students in our team are asked to judge whether a microblog user is influential or not in IT field from their viewpoints. For evaluation, we define a measure called hit ratio which counts how many users are correctly identified. If our post-feature based apporach says YES to an influential user, one is added to hit ratio; if our approach says NO to a user which is not influential, one is also added to hit ration. By comparing three kinds of post-feature based approachs with the baseline, our results show that the post-feature based approach has higher precision of identifying influential users than our baseline, as shown in Table . The experimental results show that we can identify users independent of microblog content by using content irrelevant features.

Method	Score	List	SVM	Baseline
hit ration	35/90	43/90	46/90/	35/90

Table 2: Our Experimental Results

Conclusions

In this paper, we show that an effective solution to find influential users on microblog platform. We summarize the post features of user features and design three approach to judge domain-dependent influential users. The experimental results show that our proposed approach has higher precision than a post content based approach. In the future, we will consider to combine post-feature and post-content to identify influential users through feature selection. More on our research are on https://sites.google.com/site/gdxuau/.

Acknowledgments

This research was undertaken as part of Project 61003130 funded by National Natural Science Foundation of China.

References

Bakshy, E.; Hofman, J. M.; Mason, W. A.; and Watts, D. J. 2011. Everyone's an influencer: Quantifying influence on twitter. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, WSDM '11, 65–74. New York, NY, USA: ACM.

Borda, J. 1781. Mmoire sur les lections au scrutin. *Comptes rendus de lAcadmie des sciences* 44.

Ghosh, S.; Sharma, N.; Benevenuto, F.; Ganguly, N.; and Gummadi, K. 2012. Cognos: Crowdsourcing search for topic experts in microblogs. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, 575–590. New York, NY, USA: ACM.

Kwak, H.; Lee, C.; Park, H.; and Moon, S. 2010. What is twitter, a social network or a news media? In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, 591–600. New York, NY, USA: ACM.

Sullivan, D. 2011. Twitter improves who to follow results & gains advanced search page.