Making Recommendations in a Microblog to Improve the Impact of a Focal User

Shanchan Wu
Department of Computer Science, University of Maryland
College Park, MD 20742
wsc@cs.umd.edu

Leanna Gong
University of Maryland
College Park, MD 20742
leannagong@yahoo.com

William Rand
Departments of Marketing and Computer Science, University of Maryland
College Park, MD 20742
wrand@umd.edu

Louiqa Raschid
University of Maryland
Institute for Advanced Computer Studies
College Park, MD 20742
louiqa@umiacs.umd.edu

ABSTRACT

We present a microblog recommendation system that can help monitor users, track conversations, and potentially improve diffusion impact. Given a Twitter network of active users and their followers, and historical activity of tweets, retweets and mentions, we build upon a prediction tool to predict the Top K users who will retweet or mention a focal user, in the future [10]. We develop personalized recommendations for each focal user. We identify characteristics of focal users such as the size of the follower network, or the level of sentiment averaged over all tweets; both have an impact on the quality of personalized recommendations. We use (high) betweenness centrality as a proxy of attractive users to target when making recommendations. Our recommendations successfully identify a greater fraction of users with higher betweenness centrality, in comparison to the overall distribution of betweenness centrality of the ground truth users for some focal user.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Retrieval models, Information filtering*

Keywords

Recommendation, social media, twitter, microblog

1. INTRODUCTION

The usage of social media has grown considerably in recent years, with microblogging sites being an important area of growth. On a site such as Twitter, one can follow a user and read their tweets. One can initiate a new conversation by tweeting or one can interact by mentioning a user. One can also participate in the diffusion of a topic by retweeting. Influence in a microblog can be captured in multiple ways. One can generate a lot of content or befriend a lot of users but this may not lead to a large follower network or increase diffusion. Someone who has high betweenness centrality, whose tweets diffuse rapidly or widely outside her immediate follower network, or someone who is mentioned frequently by other users may, is typically considered influential. Other factors such as the level of sentiment or persuasiveness may also play a role in diffusion.

We are interested in analyzing both diffusion and influence in microblogs such as Twitter, from the individual or personalized perspective. We want to understand who will be influenced by a particular focal user. Given a Twitter network of active users and their followers, and their historical activity of tweets, retweets and mentions, we build upon a prediction tool that uses history to predict the Top K users who will retweet or mention a focal user, in the future [10]. Our objective is to make high quality personalized recommendations for each focal user.

Social network and social influence analysis has drawn a lot of research interest. Previous research on social influence had a focus on the measurement of social influence [2] or attempted to maximize user influence [3, 5] at the aggregate level. Our objective is to track those users who will likely be influenced by an individual focal user and to improve the impact of the focal user.

We identify characteristics of focal users such as the size of the follower network and the level of sentiment averaged over all tweets. We demonstrate that these features have an impact on the quality of personalized recommendations, i.e., accuracy of predictions. As the focal user's follower network increases, prediction accuracy decreases. In contrast, we can improve prediction quality for focal users with higher levels of positive sentiment. We note that the focal users with higher levels of positive sentiment appear to have a larger following. Despite a larger following having been shown to decrease prediction accuracy, we are nevertheless able to successfully recommend users who will retweet the more positive focal user (in the future ground truth) with greater accuracy.

We use (high) betweenness centrality as a proxy of attractive and potentially influential users to target when making recommendations. Our recommendations successfully identify a greater fraction of users with higher betweenness centrality, in comparison to the overall distribution of centrality among the ground truth users.

RecSys'12, September 9–13, 2012, Dublin, Ireland

In summary, despite the difficulty of diffusion and influence prediction in evolving and noisy microblog networks, we have been successful in making personalized recommendations with improved accuracy for focal users with high(er) positive sentiment levels. We also are able to successfully recommend users with potentially greater influence (high betweenness centrality).

2. RECOMMENDATIONS IN A MICROBLOG

2.1 Problem Definition

DEFINITION 1. ***Future Retweet Prediction:*** *Given a focal microblog user u at a specific time point T and a time interval ΔT, identify K microblog users $S^u_{T,\Delta T}$ who will retweet one (or more) future tweet(s) of user u in the interval $(T, T + \Delta T]$.*

DEFINITION 2. ***Future Mention Prediction:*** *Given a microblog user u at a specific time point T and a time interval ΔT, identify K microblog users $S^u_{T,\Delta T}$ who will mention microblog user u one (or more) times in the interval $(T, T + \Delta T]$.*

2.2 Solution Approach

We formalize the retweet prediction and mention prediction problem to be a link prediction problem. Unlike traditional link prediction [7] over a homogeneous network, here we have an evolving hybrid network. In Twitter, there is an explicit Follower network, while Retweets and Mentions reflect a strong(er) relationship between users. These actions can be used to create communication-based networks independent of the follower network. In [10], we proposed and compared several methods to solve the previous two prediction problems. One of the best performing methods was based on a hybrid network and was labeled **WT-COM-BON**. The first step of **WT-COM-BON** is to create a composite weighted hybrid network with the best possible prediction potential as follows:

$$H = r \cdot R + m \cdot M + f \cdot F^*$$

R is the retweet network, M is the mention network, and F^* is a weighted follower network. In F^* a follower relationship is weighted inversely by the number of friends. If a user has a lot of friends, then her attention will be divided among those friends, and hence her weight for each friend should be lower.

The factors r, m, f are decided by a scale factor and a penalty factor. The scale factor scales the matrices so that no elements of any matrix can dominate the others. The penalty factor uses the ground truth from the training data to calibrate the influence of each network R, M, F^* with respect to retweet prediction and mention prediction. We note that the corresponding values of r, m, f, and hence H, are different for retweet prediction and mention prediction.

The next step is to apply the Bonacich centrality [1] metric to H. Bonacich centrality summarizes the total number of paths originating from a node to all other nodes; it uses an attenuation factor α to discount indirect links and β to discount direct links.

$$P = (\beta H + \beta\alpha H \cdot H + ... + \beta\alpha^n H^{(n+1)} ...)$$
$$= \beta H(1 - \alpha H)^{(-1)}$$

The ranking of the candidate users for future retweet and mention prediction is based on the rank of the values from the matrix P.

2.3 Example Recommendations

We illustrate using examples from a sample Twitter dataset; details of the dataset are provided later. We first consider users who are already widely retweeted. In our sample dataset, this includes `HumanCapLeague` (Human Capital League), a blog community of workforce management professionals, `Unibul`, a credit card merchant company, and `MomItForward`, a social media site for women. A personalized recommendation for such users, would identify specific users who (1) are more likely to retweet or mention the focal user, and (2) are influential. Targets could include users with high betweenness centrality, or users with a high level of positive sentiment. The recommended action is for the focal user to tailor tweet content, hashtags or links to external pages, to the target users, so that their message is effectively communicated.

Another type of recommendation is based on users who are active and engaged on Twitter but whose tweets have not been widely or effectively diffused as yet. `Zaibatsu` is a single father of two kids, and he tweets about social media usage, technology, photos, and humor. Users `Lizstrauss` and `TedRubin` are both social marketing strategists. Though the target recommendations might be similar in choosing users that have a large follower network, or a high betweenness centrality, the suggested actions would be different. The recommendations would encourage them to mention or retweet the target users, rather than tailor tweets to the target users.

3. EXPERIMENTAL EVALUATION

3.1 Dataset and Metrics

There have been several successful efforts to construct a proxy graph that characterizes the structure of a real network [4, 6]. For this experiment, our objective was different. It was to construct a dataset that reflected a comprehensive history of user interaction and tweet content, over an extended period, for a significant number of active users, given the strict limitations imposed by the Twitter API. We constructed a network of 15,000 users, as well as all their follower (friend) associations within this subnetwork. In choosing these 15,000 users, we focused on active users. Our premise is that the active users generate the most content and have the greatest influence. Thus, following the a large number (15,000) of active users provided us with a dataset that captured a majority of the activity that would have had an influence on these 15,000 users. We note that had we constructed a 15,000 user dataset to reflect the typical distribution of users in the network, we may have been severely limited in our ability to capture a majority of the relevant activity since the average user on Twitter is not very active.

We used the Twitter API to construct the network in the following way: Starting from a seed *active* user, we expanded her follower network and added further active users until we reached 15,000 active users. The test for an active user was as follows based on their most recent 100 tweets: (1) The user should have an average minimum tweet frequency of one tweet per day in this time period. (2)There was at least one retweet in the most recent 100 tweets. the twitter streaming API to collect all tweets published by the 15K active users between April 25, 2011 and June 25, 2011.

For each user, we collected the total number of followers and total number of friends in the entire twitter social network of that user. As we only collected a subset of the users, some of the users may only have a small fraction of her friends or followers in the subset of the users. We used a threshold $X\%$ to filter out those users in the subset by the following way:

- First get a set of the users who has at least $X\%$ of friends and also at least $X\%$ followers from the 15K users. Label this set of users as S.
- Repeat the following loop until the number of users in S is stable, i.e., $|S|$ does not change:
  For each user in S, if the number of her friends or the number

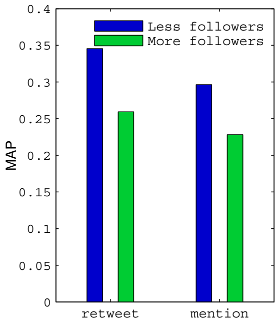


Figure 1: Prediction accuracy for the focal users with the number of followers less than and greater than the average number of followers of all focal users. The left part of the figure is for the focal users of retweet prediction; the right part is for the focal users of mention prediction.

of her followers from S is less than $X\%$ of the total number of his friends or the total number of his followers, remove this user from S.

- Return the set of users S.

We set the threshold $X\%$ = 2.4%. Our crawling statistics shows that 40% of Twitter users were "active users", and we only collected "active users". So with this threshold, we got a subset of users with at least around 6% of their active followers and 6% of their active friends in the subset. We used the first month of our data (from April 25th to May 25th) as a training dataset and we used the second month data (from May 26 to June 25) as a test dataset where we obtained the ground truth. We picked the sets of microblog users who had ground truth in the test dataset for evaluation. 2728 users had retweet ground truth and 4571 users had mention ground truth. The average number of ground truth (retwitterers) for the 2728 users is 4.23, and the average number of ground truth (mentioners) for the 4571 users is 8.64.

The metric that we used for evaluation is MAP (Mean Average Precision). MAP is widely used for evaluating for ranking methods. We set the K value to be 20.

3.2 Impact of User Network

Networking features such as the count of friends and followers, both from the global counts registered on the Twitter profile, as well as the local counts computed in our dataset, were found to be highly significant when creating a model to explain variants of user behavior and the impact of diffusion effectiveness, as reported in [9]. The same holds true for the accuracy of future retweet and mention prediction.

Figure 1 reports on the prediction accuracy for the focal users whose total number of followers is less than, or is greater than, the average number of followers of the focal users. The left part of the figure is for retweet prediction and the right is for mention prediction. The figure demonstrates that it is more difficult to predict for focal users with a larger following. When a user has more followers, more people will potentially read their tweets and retweet or mention her in the future. Some of the future users will be novel users who did not retweet her in the past. Both cases increase the difficulty of prediction.

3.3 Impact of Sentiment

Sentiment has also been widely identified as an important factor of influence and diffusion. We used a dataset and tool [8] trained for sentiment detection in tweets. In the training dataset, tweets containing positive emoticons like ":)" but not negative emoticons were labeled as positive, and tweets containing negative emoticons like ":(" but not positive emoticons were labeled as negative. A Na ive Bayes classifier (NBC) was constructed using the sentiment training dataset of 232K negative tweets and 232K positive tweets. We then used the NBC to classify our training dataset to assign a sentiment score to each tweet, in the range of [-1, +1]. Finally, we averaged the sentiment score over all the tweets of a user to determine a level of sentiment. Figure 2 reports on the distribution of the sentiment scores for each of the tweets of our dataset. Figure 3 reports on the distribution of the user sentiment level computed over all the tweets of each user.

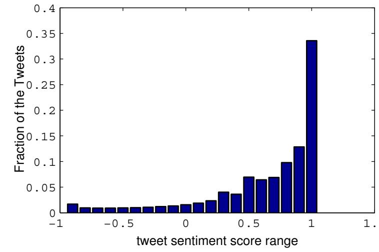


Figure 2: The distribution of the sentiment scores for all of the tweets.

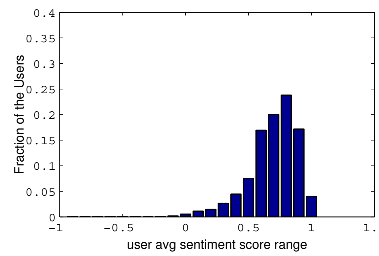


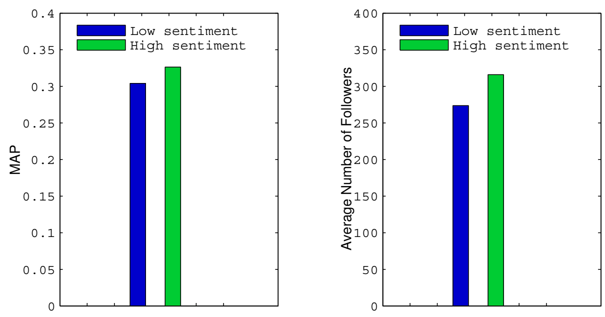
Figure 3: The distribution of the user sentiment scores in the training data for all of the users.

Figure 4 reports on the comparison of focal users with a sentiment level less than, and greater than, the average sentiment level of the focal users, for retweet prediction. Figure 4(a) compares the prediction accuracy while Figure 4(b) presents the number of followers. Figure 4(b) shows that users with a more positive sentiment level are more likely to attract a larger follower network. We have shown in a previous result, that it is more difficult to predict for focal users with more followers. However, for retweet prediction, Figure 4(a) shows that we can predict future users for focal users with more positive sentiment, with higher prediction accuracy. For example, for very positive focal users with user sentiment level > 0.9, the MAP value for retweet prediction is 0.395. In contrast, for very negative focal users with user sentiment score < 0.2, the MAP value for retweet prediction has reduced drastically is 0.253.

3.4 Impact of Centrality

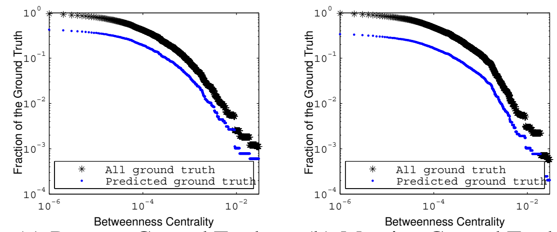
The betweenness centrality of a node v in a network is defined by the expression:

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$



(a) MAP (b) Average num of followers

Figure 4: Comparison of the focal users with the user sentiment scores less than and greater than the average user sentiment score of the focal users for retweet prediction.



(a) Retweet Ground Truth (b) Mention Ground Truth

Figure 5: Follower network betweenness centrality distribution of the ground truth users. X axis is the follower network betweenness centrality values; Y axis is the fraction of the users that have betweenness centrality values greater than or equal to the corresponding value on X axis. The upper darker distribution represents all ground truth; the lower lighter distribution is the predicted ground truth.

where σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(v)$ is the number of those paths that pass through node v.

We calculated the betweenness centrality of each user using the follower network in our dataset. We want to evaluate how well our recommended target users are also users with a high betweenness centrality, so that our recommendations are more valuable.

Figure 5 reports on the follower network betweenness centrality distribution of all the ground truth users and the subset that can be predicted by our system. The figures were drawn on the log-log scale and the distribution is somewhat close to a power law distribution. The range of the betweenness centrality values for all users in the network is [0,0.06553]. First, we consider users with low betweenness centrality in the range of [0,0.00002). While 28.9 percent of the retweet ground truth and 24.3 percent of the mention ground truth is in that range, the corresponding values for our target recommendations are 27.7% and 21.7% respectively, i.e., we make a lower fraction of our recommendations in this range.

When the betweenness centrality value increases, the curve of the predicted ground truth is closer to the curve of all ground truth for both retweet prediction and mention prediction. Thus, when we consider users with high betweenness centrality, in the range (0.001,0.06553], we see the opposite effect. A higher fraction of target recommendations is in that range. While 8.2 percent of the retweet ground truth and 12.0 percent of the mention ground truth is in that range, we recommend 8.6% and 12.4% respectively. To summarize, we are successful in recommending users in the retweet and mention ground truth that have a higher betweenness centrality.

4. CONCLUSIONS

We present results on a recommendation system for microblogs. We make recommendations of future retweet and future mention users; we also recommended suggested actions including tailored messages or targeted interactions with users. We show that focal users with higher levels of positive sentiment are associated with a larger follower network and a larger follower network typically reduces the accuracy of our predictions. Nevertheless, we are successful and can make recommendations for focal users with higher levels of positive sentiment with greater prediction accuracy. Our recommendations target future ground truth users with high betweenness centrality values. Those users are potentially more influential. The reason that we are able to identify users with high betweenness centrality values is because our solution is based on a composite network. The users with high centrality values are more likely to receive all tweets in the system; this increases their likelihood of appearing in both the mention and retweet network. Thus, our prediction method that exploits the composite hybrid network is more likely to identify these more influential users.

Acknowledgements: This research was partially supported by NSF awards CMMI 0753124, IIS 0960963, and IIS 1018361.

5. REFERENCES

[1] P. Bonacich. Power and centrality: A family of measures. *The American Journal of Sociology*, 92(5):1170–1182, 1987.

[2] M. Cha, H. Haddadi, F. Benevenuto, and K. Gummadi. Measuring user influence in twitter: The million follower fallacy. In *ICWSM '10*.

[3] W. Chen, C. Wang, and Y. Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *Proceedings of the International Conference on Knowledge Discovery and Data Mining (SigKDD)*, 2010.

[4] M. Gjoka, M. Kurant, C. T. Butts, and A. Markopoulou. Walking in facebook: a case study of unbiased sampling of osns. In *INFOCOM'10*, pages 2498–2506, 2010.

[5] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *KDD '03*, pages 137–146, 2003.

[6] J. Leskovec and C. Faloutsos. Sampling from large graphs. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '06, pages 631–636, 2006.

[7] D. Liben-Nowell and J. Kleinberg. The link prediction problem for social networks. In *CIKM '03: Proceedings of the twelfth international conference on Information and knowledge management*, pages 556–559. ACM, 2003.

[8] D. Monner. Tweet sentiment computation. *National Science Foundation SM3 Project Wiki, Smith School of Business*, 2011.

[9] P. Swaroop, Y. Joshi, W. Rand, and L. Raschid. Modeling behavior and its effect on diffusion effectiveness in microblogging. *Technical Report, Smith School of Business*, 2012.

[10] S. Wu and L. Raschid. Prediction with microblogs: Learning in a hybrid network. In *Technical Report, University of Maryland*, 2012.