

User Oriented Tweet Ranking: A Filtering Approach to Microblogs

Ibrahim Uysal
uysal@cs.umass.edu

W. Bruce Croft
croft@cs.umass.edu

Center for Intelligent Information Retrieval
Department of Computer Science
University of Massachusetts, Amherst, MA

ABSTRACT

The increasing volume of streaming data on microblogs has re-introduced the necessity of effective filtering mechanisms for such media. Microblog users are overwhelmed with mostly uninteresting pieces of text in order to access information of value. In this paper, we propose a personalized tweet ranking method, leveraging the use of retweet behavior, to bring more important tweets forward. In addition, we also investigate how to determine the audience of tweets more effectively, by ranking the users based on their likelihood of retweeting the tweets. Finally, conducting a pilot user study, we analyze how retweet likelihood correlates with the interestingness of the tweets.

Categories and Subject Descriptors

H.2.8 [Database Management]: Data Mining; J.4 [Computer Applications]: Social and Behavioral Sciences

General Terms

Algorithms, Experimentation, Human Factors

Keywords

Twitter, Retweet behavior, Filtering

1. INTRODUCTION

Although microblogs serve as a useful medium for various activities, the excessive number of posts users receive reduces their utility. In order to find interesting tweets, for example, users must scan potentially hundreds of less relevant posts every day from the people they subscribe to. This situation shows the potential value of an effective filtering system for microblogs.

Information filtering requires some criteria as the basis for the filtering process. In the case of Twitter, one of the potential criteria is whether a tweet is being disseminated by the users who read it. The practice of propagating a tweet

from another user is called *retweeting*. Since the practice of retweeting includes reading the tweet, deciding that it is worthwhile to share, and then acting on it; we can use retweet (RT) behavior as an explicit signal that the user considers the tweet to contain useful or interesting information.

In this paper, we approach the problem of reducing information overload in Twitter from two directions. First, we train a model using a coordinate ascent algorithm to rank the incoming tweets based on the likelihood that the user will RT them. Secondly, for a given tweet, we rank the users who receive that tweet, so that the user who is more likely to RT the given tweet will appear at the top of the ranking. Finally, we analyze how the tweet ranking approach based on RT likelihood correlates with the actual user preference by conducting a pilot user study.

2. RELATED WORK

One way to address the problem of improving the attractiveness of incoming tweets is choosing which user to follow in a more selective manner. The idea is that if a user receives the tweets of the right users, then she is less likely to get lost among uninteresting tweets. A few papers propose effective user recommendation approaches to overcome this issue. Hannon et al. utilize content and collaborative based approaches to evaluate a range of different user profiling and recommendation strategies [4]. In another paper, Golder et al. investigate a number of user similarity measures to be used in contact recommendation in Twitter [3]. Discovering both global and topic-sensitive influential users is also studied [12, 13].

A number of papers have studied how to classify tweets to improve information filtering. Sriram et al. trained a Naive-Bayes classifier to determine which predefined class a tweet falls into [10]. In their paper, they used categories such as news, events, opinions, deals, and private messages. Classifying the tweets solely on their news-relatedness was studied in [2, 9]. All these classifications are generic, not user oriented.

In order to shed light on what sort of information people propagate on Twitter, several studies investigated the RT behavior of the users. Boyd et al. focused on the qualitative aspect of the RT behavior and gave answers to the questions like “Why people retweet?” and “What people retweet?” [1]. Having a similar motivation to this paper, Suh et al. analyzed 9 features to see how they correlate with the retweetability of a tweet [11]. Their hypothesis is that recommen-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CIKM'11, October 24–28, 2011, Glasgow, Scotland, UK.
Copyright 2011 ACM 978-1-4503-0717-8/11/10 ...\$10.00.

ation and personalization could help to optimize the information diffusion, and help the users to catch interesting tweets. Apart from the global statistical analysis of these features, they did not build a personalized user model using the RT prediction.

To the best of our knowledge, there are four papers extending the analysis to building a retweet prediction model. Yang et al. used 22 features for observation, then built an RT prediction model using a factor graph model [14]. In another study, Zaman et al. used a probabilistic collaborative filtering model, called Matchbox, to make RT predictions [15]. Hong et al. used 5 features to train a logistic regression model and achieved a performance with relatively low recall score: 0.44 [5]. Lastly, Petrovic et al. introduced a few more features such as novelty of a tweet and the number of times the author is listed, to train a model using a passive aggressive algorithm [6]. All these papers approach the RT prediction problem from a global perspective, answering the question of “Whether a given tweet will be retweeted by anyone or not?”. Thus, they fail to capture which tweets are more retweetable from a user-centric perspective. In this paper, we build a model which is adaptable for individual users.

Finally, Ramage et al. investigated which topics users are interested in using a Labeled-LDA approach, to classify the tweets based on whether each one is worth to be read or not [8]. They also addressed the user recommendation task, however, what they do is basically deciding whether a user is in the followee list of a given user or not.

3. PROPOSED METHODS

The goal of ranking the Twitter feed of individual users is to display incoming tweets in a descending order based on their likelihood of being retweeted by the specific user. An effective ranking will help the user to find potentially more interesting tweets. Compared to a filtering approach where the incoming information either passes or fails the filter, the ranking method is more conservative in that all the relevant information is presented to the user in a more organized way.

Another approach for reducing the information overload is presenting the posted tweets to the potential audiences more selectively. Twitter already addresses this issue by hiding personal conversations from other users’ timelines. However, this is just a small step towards filtering out uninteresting tweets. Our approach for this problem is to rank the potential audience (users) of a tweet based on their likelihood of retweeting that tweet.

In the next two subsections, we explain our approach and experiments in detail.

3.1 Ranking Incoming Tweets

As a preliminary experiment, we trained a decision tree based classifier (J48), using the WEKA¹ toolkit, to see how accurately we can classify the tweets as retweetable or not for a specific user. To train the classifier, we used four groups of features explained below. Then, with the same set of features, we took a learning to rank approach using the Coordinate Ascent (CA) algorithm. The experimental setup and the results are explained below as well. In all our experiments, we used 5-fold cross-validation to split the training and test sets.

¹<http://www.cs.waikato.ac.nz/ml/weka/>

3.1.1 Features

Author-based: These are basically the features that appear on or can be inferred through the user profile: Is the author locally elite (i.e. the # of followers she has is in the range of (10K - 50K)), is she globally elite (i.e. she has more than 50K followers), followers count, friends count, age (i.e. # of days since the creation of her account), statuses count, favourites count, her tweet rate (i.e. avg. tweet count per week), # of times she is listed, is she a verified user, does she have a description, does she have a url, is her language English.

Author-related features mainly give clues about the authority of the author and how active user she is.

Tweet-based: These are the syntactic features of the tweet: Does it contain hashtags (#xyz), urls (<http://xyz..>), mentions (@user), question mark (?), exclamation mark (!), quotation mark (”), first person pronoun (I), same character consecutively three times (e.g. coooool), emoticons (e.g. :P), is the tweet retweeted by anyone, is the tweet an RT itself, the length of the tweet, and the tf-idf score of the tweet.

Some of these features give implications about how well-written the tweet is, the category, the audience, and the popularity of the tweet.

Content-based: Content based features are the ones related to the information contained by the tweet. These are the novelty of the tweet (i.e. minimum cosine distance to the other tweets that appear on the user’s timeline in the last week), the unexpectedness of the tweet from the author (i.e. minimum cosine distance to the other tweets of the author), and the Query Likelihood (QL) [7] score of the author based on the Language Model (LM) of her tweets.

User-based: These features are related to the user whose Twitter feed is being ranked. From the user’s point of view, the features are as follows: Is the author my friend (i.e. is she also following me), is the author my conversation friend (i.e. did we mention each other before), did I RT her before, did I list her, did I use the hashtag in the tweet (if any), did I use the url domain in the tweet (if any), is the tweet mentioning me, did I mention the any of the users mentioned in the tweet (if any).

The first four features show how closely the author and the user is interacting with each other. The remaining ones indicate how the tweet is related to the user and her interests.

3.1.2 Experimental Setup

Dataset: We used the Twitter REST API² to crawl information about the users and their tweets. To train the classifier, we need both positive and negative examples. The positive examples in our case are obvious: the user retweeting a given tweet. However, the negative examples are a little bit problematic since we do not know whether the user did not retweet a given tweet on purpose or she did not even see that tweet. In order to make sure that the latter is not the case, we randomly chose seed users who are considered as “ordinary”. We assumed that if a user has (10-1,000) friends/followers, tweets (1-200) times a week, and has tweeted more than 10 times so far, then she is considered to be an ordinary user. Most active Twitter users fall into this category. These users are periodically logging into

²<https://apiwiki.twitter.com/w/page/22554679/Twitter-API-Documentation>

Table 1: Classification results for predicting retweetability of the tweets. The metrics are based on the positive instances.

Features	Precision	Recall	F-Measure
All	0.771	0.683	0.724
Tweet-based	0.771	0.369	0.499
Author-based	0.636	0.201	0.305
User-based	0.728	0.191	0.302
Content-based	0.690	0.019	0.037

Table 2: Tweet ranking results based on the retweetability.

Features	NDCG@1	NDCG@5	NDCG@10
All	0.814	0.834	0.809
Tweet-based	0.797	0.722	0.718
User-based	0.306	0.381	0.392
Content-based	0.376	0.326	0.339
Author-based	0.239	0.263	0.277
Random_baseline	0.095	0.099	0.111

Twitter (i.e. tweeting 1-200 times a week), therefore, they are less likely to skip the tweets appearing on their timelines. Yet, we cannot guarantee that they read all their tweets.

We crawled 242 ordinary seed users, all their followees and tweets. Then, for each seed user, we randomly selected 100 tweets that would appear on her Twitter feed. All together, we had 24,200 tweets, 2,547 of which were retweeted by the seed users. For each of the 24,200 tweets, we extracted the features using information from the author of the tweet, the tweet itself, and the user who is receiving that tweet. We followed the “RT @username” convention on deciding whether a tweet is RT or not. Twitter’s “retweet” button also adheres to this convention implicitly.

Results: The classification results using each group of features are shown in Table 1. *Precision* and *recall* metrics are used in the evaluation. The metrics are evaluated based on the positive instances. Since the RT ratio is almost 0.1 in our corpus, building an effective classifier on the positive instances is quite challenging. Nevertheless, the results obtained are reasonable.

Among the feature groups, the tweet-based ones have better performance. Since the tweet-based features are quite common in other microblogging services, this is a useful result. On the other hand, none of the feature groups gives satisfactory results alone.

We also applied the “leave-one-out” technique to see which features are more useful. From the tweet-based features, the presence of a URL, whether the tweet is retweeted, and the tf-idf score of the tweet were the most effective features. In addition, the “is she my conversation friend?” and “Did I use that URL domain” features from the user-based category are also useful features.

In the main part of the experiments, we ranked the incoming tweets for the 242 users to optimize the retweetability probability of the top ranked tweets. The results are shown in Table 2. For this part of the experiments, we used the *NDCG@k* metric. The relevance score for a tweet is 1 if the user retweets it later, 0 otherwise. Along with the results using each set of features, we also listed the performance of a random baseline, which is quite similar to how Twitter

Table 3: Classification results for predicting whether a user will retweet a given tweet.

Features	Precision	Recall	F-Measure
All	0.965	0.668	0.789
User-based	0.976	0.629	0.765
Author-based	0.679	0.095	0.166
Content-based	0.0	0.0	0.0

would rank the tweets. Again, tweet-based features outperform other groups of features. Using all the features, it is possible to achieve very effective ranking of the tweets for retweetability. In other words, the Coordinate Ascent algorithm can learn how to rank the tweets to place the the ones that will be retweeted on the top.

3.2 Ranking Targeted Users

Similar to the previous section, we attempted to answer the following two questions: “Can we classify the users based on whether they retweeted a given tweet?” and “Can we rank the users to potentially identify which users should receive a tweet?”. Again, we used the same classifier and learning to rank algorithm with a slightly different set of features. 5-fold cross-validation is also applied to split the dataset.

3.2.1 Features

Author-based: From the author’s point of view, the features are as follows: Is the user my friend (i.e. am I also following her), is the user my conversation friend, did I RT her before, did she RT me before, did she use the hashtags in the tweet (if any), am I mentioning her in the tweet.

Content-based: Novelty of the tweet (i.e. minimum cosine distance of this tweet to the user’s previous tweets) and the QL score of the user based on the LM of her tweets, where the tweet is the query.

User-based: We used the same set of author-based features with the tweet ranking task, which are applied to the users, not to the author of the given tweet.

We did not consider the tweet-based features here because our main focus is ranking the users and tweet-related features would be the same for all the users seeing this tweet.

3.2.2 Experimental Setup

Dataset: In this part of the experiments, we randomly selected 665 tweets which were retweeted at least by one user. For each tweet, we crawled all the users who follow the author of the tweet. The total number of users were 53,136. Due to the problem we stated about the negative instances of retweeting, we eliminated non-ordinary users from this pool. In addition to the described ordinary user criteria, we also enforced these users to tweet at least 10 times a week, to make sure that they are more likely to read all their incoming tweets. Finally we had 20,397 followers for the 655 tweets.

Results: Table 3 lists the classification results, where the the task was to decide whether a user will retweet a given tweet or not. This problem is more challenging than the previous classification task, because out of 20,397 people, just 780 retweeted something. In other words, the average number of RTs for the selected 665 tweets is 1.17. Even so, our classifier managed to achieve a relatively high F-measure. User-based features dominated other features, but still benefited from them in the overall model. The content-

Table 4: User ranking results based on the retweet likelihood.

Features	Best@1	Best@5	Best@10	RR@10
All	0.602	0.756	0.828	0.684
User-based	0.602	0.657	0.736	0.637
Author-based	0.528	0.675	0.782	0.592
Content-based	0.362	0.676	0.795	0.508

based features had unacceptably poor performance on the positive instances. Our feature analysis through “leave-one-out” technique revealed that the best feature was the tweet-rate of the users. From the author-based features, “did she use the hashtags in the tweet (if any)” and “am I mentioning her in the tweet” also proved to be effective.

Considering that approximately one follower retweets each tweet, instead of $NDCG@k$, we used $Best@k$ and *Reciprocal Rank* metrics for evaluating how effectively we can rank the users. The results are demonstrated in Table 4. We can infer that, for instance, 75% of the times, the user who will RT the given tweet will appear on the top 5 rankings (e.g. Best@5: 0.756). It is also shown that the retweeter appears either on the top, or the 2nd position in the list on average (e.g. RR@10: 0.684). Although the feature groups do not have significant variances in performance, it is notable that the content-based features do work here even when used alone.

4. PILOT USER STUDY

In order to understand whether RT prediction can be an indicator of the interestingness of the tweets, we conducted a pilot user study on 10 volunteers. In the study, the participants rated a total of 700 tweets that they see on their Twitter feed, based on the following rating scheme:

- 2: I’m glad seeing this tweet.
- 1: I don’t mind seeing this tweet. I may enjoy reading it if I have time.
- 0: I prefer not to see this tweet at all. It’s a waste of my time.

The average rating was 0.95.

We then applied the tweet ranking model described in Section 3.1 to the tweets presented to the participants. As shown in Table 5, when the tweets are ranked temporally (as Twitter does), the NDCG values are lower than our method of ranking the tweets. The size of the pilot study is too small to make sound conclusions. Nevertheless, ranking the incoming tweets based on their likelihood of being retweeted by a specific user yields promising results in this user study, which supports our motivation.

5. CONCLUSIONS

In this paper, we proposed two methods for tackling the information overload problem in microblogs. Using four groups of features, we trained a Coordinate Ascent learning to rank algorithm to rank the incoming tweets to help the users interact with the tweets they are more likely to retweet. Additionally, we ranked users given a tweet, so that the users who are more likely to retweet have higher ranks. While our first approach effectively ranks the tweets, the second approach can help to decide the priority of a tweet for a specific user at an earlier stage. Finally, conducting a pilot user study, we observed that the tweets users want to read correlates with the ones they are likely to retweet.

Table 5: Performance of the RT likelihood based tweet ranking on interestingness.

Features	NDCG@1	NDCG@5	NDCG@10
Our model	0.4	0.543	0.541
Twitter	0.45	0.449	0.444

6. ACKNOWLEDGMENTS

This work was supported in part by the Center for Intelligent Information Retrieval. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the sponsor.

7. REFERENCES

- [1] D. Boyd, S. Golder, and G. Lotan. Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In *HICSS*, 2010.
- [2] C. Castillo, M. Mendoza, and B. Poblete. Information credibility on twitter. In *WWW*, 2011.
- [3] S. A. Golder, A. Marwick, and S. Yardi. A structural approach to contact recommendations in online social networks. In *Workshop on Search in Social Media at ACM SIGIR*, 2009.
- [4] J. Hannon, M. Bennett, and B. Smyth. Recommending twitter users to follow using content and collaborative filtering approaches. In *RecSys*, 2010.
- [5] L. Hong, O. Dan, and B. D. Davison. Predicting popular messages in twitter. In *WWW*, 2011.
- [6] S. Petrovic, M. Osborne, and V. Lavrenko. Rt to win! predicting message propagation in twitter. In *ICWSM*, 2011.
- [7] J. M. Ponte and W. B. Croft. A language modeling approach to information retrieval. In *SIGIR*, 1998.
- [8] D. Ramage, S. T. Dumais, and D. J. Liebling. Characterizing microblogs with topic models. In *ICWSM*, 2010.
- [9] J. Sankaranarayanan, H. Samet, B. E. Teitler, M. D. Lieberman, and J. Sperling. Twitterstand: news in tweets. In *GIS*, 2009.
- [10] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas. Short text classification in twitter to improve information filtering. In *SIGIR*, 2010.
- [11] B. Suh, L. Hong, P. Pirolli, and E. H. Chi. Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In *SOCIALCOM*, 2010.
- [12] J. Weng, E.-P. Lim, J. Jiang, and Q. He. Twitterrank: finding topic-sensitive influential twitterers. In *WSDM*, 2010.
- [13] Y. Yamaguchi, T. Takahashi, T. Amagasa, and H. Kitagawa. Turank: Twitter user ranking based on user-tweet graph analysis. In *WISE*. 2010.
- [14] Z. Yang, J. Guo, K. Cai, J. Tang, J. Li, L. Zhang, and Z. Su. Understanding retweeting behaviors in social networks. In *CIKM*, 2010.
- [15] T. R. Zaman, R. Herbrich, J. Van Gael, and D. Stern. Predicting information spreading in twitter. In *Workshop on Computational Social Science and the Wisdom of Crowds at NIPS*, 2010.