A Multi-view Retweeting Behaviors Prediction in Social Networks

Bo Jiang, Ying Sha, and Lihong Wang

Institute of Information Engineering, Chinese Academy of Sciences,
Beijing, 10093, China
{jiangbo,shaying}@iie.ac.cn

Abstract. Retweeting is the most prominent feature in online social networks. It allows users to reshare another user's tweets for her followers and bring about second information diffusion. Predicting retweeting behaviors is an important and essential task for advertising product launch, hot event detection and analysis of human behavior. However, most of the methods and systems have been developed for modeling the retweeting behaviors, it has not been fully explored for this problem. In this paper, we first cast the problem of retweeting behaviors prediction as a classification task and propose a formally definition. We then systematically summarize and extract a lot of features, namely user status, content, temporal, and social tie information, for predicting users' retweeting behaviors. We incorporate these features into Support Vector Machine (SVM) model for our prediction problem. Finally, we conduct extensive experiments on a real world dataset collected from Twitter to validate our proposed approach. Our experimental results demonstrate that our proposed model can improve prediction effectiveness by combining the extracted features compared to the baselines that do not.

Keywords: retweeting behaviors, online social networks, SVM, extract feature, classification.

1 Introduction

Online social networks such as Sina Weibo, Twitter and Facebook have become an important information service platform for all walks of life. People not only share interesting information each other but also express their views on hot topic occurred in the real life. According to the study in [16], users post more than 500 million tweets every day in Twitter. These information is widely spread with thousands of millions of users participating over Twitter through retweeting mechanism. Retweeting is a social service function for users that one can reshare any users' tweets to his/her timeline. Through the feature, tweets can quickly reach all of their followers. This may cause information cascade where a tweet is reposted from one user to another or from one community to another. Moreover, retweeting has proven to be a significant factor for the form of large information cascades in social network [5]. Therefore, understanding the mechanisms of information diffusion and predicting users' retweeting behaviors are an

R. Cheng et al. (Eds.): APWeb 2015, LNCS 9313, pp. 756–767, 2015.
DOI: 10.1007/978-3-319-25255-1_62

important tasks for effective monitoring the trend of information diffusion and maximizing the popularity of new product. Specifically, in this paper, our goal is to choose a focal user and then try to predict who will retweet an incoming tweet published by the focal user in the near future. We propose a prediction model combining user status, content, temporal, and social tie information to model users' retweeting behaviors.

The main contributions of this work can be summarized as follows:

- We formulate the problem of retweeting behavior prediction as a classification task. Specifically, given an incoming tweet posted by a publisher, our goal is to predict who will retweet.
- We systematically summarize and extract a lot of features which are closely related with retweeting behavior prediction. We then propose a prediction model to incorporate user status, content, temporal, and social tie information for predicting users' retweeting behaviors. Meanwhile, we introduce semantic enrichment technologies to measure interests relevance between publishers, followers and transmissible tweets.
- We collect a large number of tweets from Twitter service. Experimental results on the constructed dataset demonstrate that our proposed method outperforms other baselines with a significant margin.

The rest of this paper is organized as follows. We review the related work in Section 2. In Section 3, we introduce a clear definition of retweeting prediction task. Section 4 discusses how to extract effective features for the problem. We present experiments and empirical analysis of our models in Section 5. Finally, conclusions are given in Section 6.

2 Related Work

The studies of retweeting behavior in social network have exist an explosion of research. We can roughly divide these works into two categories of models in this scope: (i) explanatory models and (ii) predictive models. In the following, we will summarize and review some representative efforts in both of them.

The goal of explanatory models is to understand why people retweet and analysis which factors impact retweet. These models are very useful to understand how human make decision and how information spread. For example, Boyd et al. [2] conduct a user survey to analysis the reasons on how people retweet, why people retweet, and what people retweet. Suh et al. [15] firstly collect a large number of Twitter data and extract a number of features to identify factors that might affect retweetability of tweets. Yang et al. [3] analyze how the retweeting behaviors is influenced based on user, message and time factors. They find that almost 25.5% of the tweets posted by users are actually retweeted from their friends. Macskassy et al. [11] make a better understanding of what makes people spread information in Twitter through the use of retweeting. Abdullah et al. [1] conduct a user survey to investigate what is the user's action towards spread message and why user decide to perform on the spread message. Their results reveal that users retweet a message due to the important and interesting of content and author's influence.

The aim of predictive models is to predict who will repost a tweet and the scale and depth of retweeting in a given network based on user, content, social and/or temporal features. For instance, Liu et al. [9] propose a probabilistic graph model to measure topic-level influence between users in order to predict user behaviors. Zhang et al. [18] propose the notion of social influence locality and construct a large ego network to study users' retweeting behaviors. Naveed et al. [12] only employ the content-based features of retweets to predict the probability of a tweet to be retweeted. Can et al. [4] exploit content- and structure-based features as well as image-based features to predict the retweet count of the tweets that contain links to images. Zaman et al. [17] present a collaborative filtering approach using user feature and tweet feature to predict individual retweets in Twitter. Luo et al. [10] use a wide range of followers' features, such as retweet history, social status and interests to predict retweet occurrences. Feng et al. [7] develop a feature-aware factorization model to recommend the tweets based on the their probability of being retweeted. Peng et al. [13] propose using conditional random fields (CRFs) to model and predict the retweet patterns with three types of user-tweet features. Petrovic et al. [14] propose a time-sensitive model combining social features and tweets features to predict retweets. Zhang et al. [19] develop using non-parametric Bayesian model adapted from the hierarchical Dirichlet process for predicting retweeting behaviors.

However, prior research work ignore some critical factors for users' retweeting behaviors. First, the more interactions between users show that they have a strong relationship, and retweeting behaviors are more likely to happen. Second, user's topics of interest play an important role when retweeting, previous work don't precisely identify the users interests due to only use Bag Of Words approach. We focus on how to tackle these problems in this paper.

3 Problem Statement

Retweeting is an important social function for information diffusion in social networks. It allows users to directly repost a tweet using the form of RT @username.

For convenience, we name RBP (**R**etweeting **B**ehavior **P**rediction) for our proposed model. Meanwhile, we formally define RT @username as a three tuple representation of retweeting as follows.

Definition 1. (RT @username) Suppose given a tweet t posted by a user u and a user v repost the tweet t to her timeline. An RT @username is a retweeting relation triple (u, t, v). Specifically, we denote u as retweetee, t as transmission tweet and v as retweeter, respectively.

We formally define the problem of user retweeting prediction as a classification problem. More specifically, given a tweet t published by a user u, and the candidate set C that is consist of followers and the interactive users, the goal of the work is to find who will retweet the tweet t in C where every candidate

$c_i (\in C)$ is tagged whether she retweeted tweet t in training data or not. Therefore, the problem can be solved by employing effective features in a supervised learning framework.

4 Features for Retweeting Prediction

In this section, we define a lot of features for modeling the retweeting behaviors. These features are roughly divided into four main categories, namely user status, content, temporal and social tie information. Next, we will introduce the extraction methods of these features.

4.1 User Status Features

User's personal attributes are an important feature for effecting one's retweeting or retweeted. Specifically, we also consider the personal profile of publisher and follower, including *the number of followers, followees, tweets, favorites and listed, verification status, the age of the account, location, whether the user profile has a self-description, and whether the user profile has a URL.* Intuitively, the richer profile user has, the stronger the social credibility. Hence, their tweets are more likely to be retweeted than those who have recently just create a new account or have a mall number of content in their profile.

4.2 Content Features

The intuition behind is that whether a user retweet an incoming tweet or not depends on content's self-feature and user's interest to a certain extent. Hence, in the subsection, we extract three categories of features for retweeting behavior prediction, namely *self-feature, the interest similarity between publishers and followers, the topic similarity between an incoming tweet and followers.*

For tweet's self-feature, we extract a set of feature, such as the number of hashtags, URLs, media (containing images and videos) and mentions (referencing other users in tweet text). Previous study [15] has been found that these features have strong relationships with retweetability. Moreover, a lot of tweets are an express of personal sentiments in social network. [8] has shown that whether a tweet contains sentiment words or not may effect the retweetability of the tweet. To examine the influence of emotion in the retweeting, we use Stanford CoreNLP[1] that is a natural language analysis library including sentiment analysis tools to identify tweet emotional content. We classify sentiment into three classes, such as positive, neutral and negative. 1 represents the positive emotion of a tweet, 0 represents neutral, and -1 represents negative.

For the second feature, previous work [1,10] has been studied that the match of topics of interest between publishers and followers is an key important feature for retweeting behavior prediction. When facing a lot of incoming tweets, a follower is more likely to retweet these tweets that he/she is interested in.

[1] `http://nlp.stanford.edu/software/corenlp.shtml`

To calculate the match, an intuitive way is to leverage topic modeling methods like Latent Dirichlet Allocation (LDA) for extracting user topic distribution. However, these topic modeling methods may not fit for tweets due to their short text, noisy, ambiguous and dynamic nature features [6]. Consequently, we use semantic enrichment methods to generate user's topics of interest. The semantic enrichment methods explicitly attach entities in their tweets to semantic annotations by applying Linked Data, and therewith allows for making explicit topic tags. Since the focus of our work is on user's topics of interest identification and not semantic annotations technologies, we employee an existing solution.

A plethora of semantic annotation with linked data techniques and systems is available in general [6]. We opt to use OpenCalais[2] for our work because of the following reasons: (1) OpenCalais maps the entities identified in tweets to their corresponding topic label; (2) OpenCalais incorporates state-of-the-art semantic function with content; (3) OpenCalais provides a relatively high rate limit to other services, the API default usage quotas are 50,000 transactions per day, 4 transactions per second[3]. OpenCalais now offers 18 topics categorization, such as Sports, Education, Environment and Politics. Therefore, in this paper, we construct a 18 dimensions vector representation to profile user's topics of interest. We can then compute a cosine similarity of user's interest between publishers and followers. Specifically, give the publisher u and his/her follower v, u's topics vector is denoted as $P(u)$, and v's topics vector is denoted as $P(v)$. We denote *interest similarity* between u and v as follows:

$$InterSim(u, v) = \frac{P_u \cdot P_v}{\| P_u \| \| P_v \|} \tag{1}$$

Last but not least, the interesting tweet content also is a key factor in pushing the retweetablity of the users. Therefore, in this paper, we also propose a feature to measure the retweetable of the followers for an incoming tweet. Analogously, given a tweet t and the user v, we denote the topic vector of tweet t as $P(t)$, and then *topic similarity* between tweet t and user v can be estimated as below:

$$TopicSim(t, v) = \frac{P_t \cdot P_v}{\| P_t \| \| P_v \|} \tag{2}$$

4.3 Temporal Features

Intuitively, Twitter users have the same or similar activity time, the retweet action is more likely happen each other. This is because tweets published by one user in the morning are often overwhelmed by other users in the afternoon because of them being replaced by the more recent tweets. In this study, we consider four time based features as the recency of retweeting a tweet. The first feature can be extracted from user profile that whether two users are in the same

[2] `http://www.opencalais.com/`

[3] `http://www.opencalais.com/documentation/calais-web-service-api/usage-quotas`

timezone or not. We use *timezone*(1 indicates being lived in the same area and 0 indicates not being lived) to represent the time difference of user living city.

We also propose another feature *activity time overlap*, which is defined as the overlapping degree of posting time between publisher and follower. Specifically, we divide one day into 24 units per one hour and map user's activity time span to discrete time slice. Assume given publisher u and follower v, we denote u's activity time span as $t_a(u)$, and v's activity time span as $t_a(v)$. Similar to Jaccard Coefficient, we define *activity time overlap* between publisher u and follower v as follows:

$$ATO(u,v) = \frac{|t_a(u) \bigcap t_a(v)|}{|t_a(u) \bigcup t_a(v)|} \tag{3}$$

This feature could measure the consistency of publisher u's posting time habit and follower v's posting time.

Moreover, the more recent the interaction happen, the more likely the behavior to be occurred in the near future. Specifically, we denote the timestamp of v first interact with u as $I_s(v,u)$, and the timestamp of their last interaction as $I_e(v,u)$. We define *interaction span* from v to u is defined as:

$$IS(v,u) = I_e(v,u) - I_s(v,u) \tag{4}$$

Correspondingly, *interaction frequency* is the average interacting interval from v to u:

$$IF(v,u) = \frac{freq(v,u)}{IS(v,u)} \tag{5}$$

where $freq(v,u)$ is the number of times v interact with u in the above given time interval.

Furthermore, the more recent the retweeting happen, the more likely the behavior to be occurred in the near future. Specifically, we denote the timestamp of follower v last retweet tweets posted by publisher u as $R_e(v,u)$. We then calculate *recent retweeting interval*, which is defined as the interval between $R_e(v,u)$ and the timestamp $T_t(u)$ of an incoming tweet t posted by user u:

$$RRI(u,t,v) = T_t(u) - R_e(v,u) \tag{6}$$

4.4 Social Tie Features

The intuition behind is, the more strong social tie between users has each other, the more likely the retweeting happen. We measure the strength of social tie combining structural, relationship, and interaction information together to predict users' retweeting behaviors.

To extract structural feature, we first construct an explicit network G_e by utilizing the followees and followers relationships in the data collection. In G_e, the nodes represent users and the directed edges represent following or followed relationship between user u and user v. We extract two features between two users by the number of *mutual followees* and *mutual followers* as prediction indicator.

In addition to the explicit network, we also construct an implicit network G_i that is consist of retweet network, reply network and mention network. We call G_i as interaction network. Therefore, we extract *interaction num* that sum of the number of retwteet, reply and mention from follower to publisher as one of the prediction features.

Moreover, we observe that users have four types of relationship each other in social network such as stranger, followee, follower, friend where stranger denotes no link between users, followee and follower denote a unidirectional follow relationship, friend denotes a bidirectional follow relationship. We define *relationship type* to measure the familiarity between users. Specifically, 3 represents friend relationship, 2 represents following relationship, 1 represents follower relationship and 0 represents stranger relationship.

To sum up, Table 1 gives a complete list of features described in this section for retweeting prediction task, where u denotes retweetee, t is transmission tweet and v denotes retweeter.

Table 1. The summary of features for retweet prediction model

Category	Feature	Description
User Status Feature	num_follower	number of who one is followed
	num_friend	number of whom one is following
	num_tweet	number of tweets that one post
	num_favourites	number of tweets that one like
	num_listed	number of group that one list
	is_verification	whether account is verified or not
	age_account	the account create time
	location	whether user's location enable or not
	self-description	whether profile has a self-description or not
	URL	whether profile has a URL or not.
Content Feature	num_hashtag	number of hashtag that a tweet contain
	num_URL	number of URL that a tweet contain
	num_media	number of media that a tweet contain
	num_mention	number of mention refer to another user
	sentiment	tweet emotional class
	u2u_interest	topic based user similarity between users
	t2u_interest	topic based preference of user v for t
Temporal Feature	timezone	whether u and v are lie the same area or not
	ATO	activity time overlap between u and v
	IF	interaction frequency between u and v
	RRI	recent retweeting interval between u and v
Social Tie Feature	mutual_followee	number of mutual followee between u and v
	mutual_follower	number of mutual follower between u and v
	num_interaction	interaction number between u and v
	relationship	the type of relationship between u and v

5 Experiments

In this section, we first describe the approach of data collection from Twitter service. Then the baseline methods and evaluation metrics are proposed. Finally, we compare the effectiveness of our approach with these baseline methods and analysis the results.

5.1 Data Collection

For the work executed for this paper, we collect the microblog data from Twitter service. More specifically, we use Twitter API[4] to collect our experiment dataset from 15th September to 20th December 2014. The dataset was collected in the following ways. First, we randomly select 50 users as seed users. For each seed, we crawl four types of information including personal profile, follower and followee list, and all tweets. In the same way, we also collect all these information of their followers and followees. Finally, the dataset contains 14325 users, 63 millon edges relationships among them and 22 millon tweets.

We focus on retweeting behaviors in social networks. Thus, we preprocess the dataset by extracting popular tweets which are retweeted larger than 30 times from the data set. Each diffusion process contains the original retweetee and all its retweeters. After preparation, we have 2615 publishers, 31359 original tweets which give rise to 1,849,596 retweet instances. Table 2 lists statistics of the retweeted data.

Table 2. Retweeters data statistics

Dataset	#Users	#Relationships	#OriginalTweets	#Retweets
Twitter	2615	4,852,240	31,359	1,849,596

In order to limit problems like overfitting, we first utilize a ten-fold cross validation and split the data into training and testing data. We then report the average performance in ten rounds of tests.

5.2 Comparison Methods

To evaluate the performance of our prediction models, we compare our prediction results with four baseline prediction models as follows:

- **Random Guess(RG):** We randomly selects users and randomly assign the class label to each user with the prior probability.
- **Majority Vote(MV):** We observe that a lot of users ever retweet the same user's tweets many times. Therefore, we first rank candidates in our dataset by the number of times they ever retweeted the publisher's previous tweets before. Then, we choose top ranked users to retweet. This simple but powerful baseline has been used in most existing studies.

[4] `https://dev.twitter.com/docs/api/1.1`

- **Who Will Retweet Me(WRM):** Our work is similar to previous study [10] focusing on a wide range of features, such as retweet history, followers status, followers active time and followers interests to find retweeters in Twitter. We implement the method as the baseline.
- **LRC-BQ:** [18] formally define the feature of social influence locality, and also combine additional features(personal attributes, instantaneity and topic propensity). The authors use these features to train a logistic regression classifier for predicting users' retweeting behaviors. Moreover, the method and dataset mentioned in the paper has been released[5]. We employee the method as the baseline.

5.3 Evaluation Metrics

We use four common metrics to evaluate the performance of our prediction model, namely Precision, Recall, F1-measure, Accuracy.

Specifically, we assume Γ is the set of testing samples and N = $|\Gamma|$ is the size of testing samples. The ground truth of retweeted tweets is notated as Θ and M = $|\Theta|$ is the number of true retweeted tweet. We then denote an indicator function θ to indicate whether a tweet t is retweeted ($\theta = 1$) or not ($\theta = 0$). Let $\hat{y} = \{\hat{y_1}, \hat{y_2}, \cdots, \hat{y_n}\}$ be our prediction result vector and $y = \{y_1, y_2, \cdots, y_n\}$ be the ground truth vector. Therefore, the Precision, Recall, F1-measure, and Accuracy can be computed as follows:

$$Precision = \frac{\sum_{i=1}^{N} \theta_i}{N} \qquad Recall = \frac{\sum_{i=1}^{N} \theta_i}{M} \tag{7}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \qquad Accuracy = \frac{\sum_{i=1}^{N} \{\hat{y_i} = y_i\}}{N} \tag{8}$$

5.4 Results and Analysis

Overall Results and Analysis. In Table 3, we compare the baseline methods and our proposed approach (named as RBP) for different metrics. From the table, we can observe that our proposed prediction method significantly outperforms other baseline methods. Through comparing F1-score than other the methods, we can see that extracted these features which we proposed in the previous section is a better indicator for predicting users' behaviors. Meanwhile, we have found that the performance of WRM in the experiments dataset is lower than that of reported in [10]. The most likely cause of the low performance is the noisy and tweets' diverse of content. In addition, the performance of LRC-BQ in our experiment results is basically the same comparable with that of reported in [18]. This shows the robustness of influence locality method, and social network platforms between Twitter and Weibo also exist great similarities. In addition, we also compare other popular models: Naive Bayes, Logistic Regression, Random Forest. The experimental results show that SVM classifer outperforms other methods for the task. We omit the details due to space restrictions.

[5] `http://arnetminer.org/billboard/Influencelocality`

Table 3. Performance of retweeting behaviors prediction

Model	Precision	Recall	F1	Accuracy
RG	0.342	0.338	0.340	0.339
MV	0.475	0.478	0.477	0.476
WRM	0.514	0.505	0.509	0.508
LRC-BQ	0.708	0.643	0.674	0.673
RBP	**0.876**	**0.868**	**0.872**	**0.871**

Feature Evaluation. As discussed in Section 4, we extract a lot of features and then roughly divide them into four categories: user status, content, temporal and social tie features. In order to explore the importance of the four features to the prediction performance, here we eliminate one feature at a time from our proposed model. Specifically, we denote each comparison method and a simple explanation as follows: NU-RBP represents the model without user based features into consideration, NC-RBP represents the model without content based features into consideration, NT-RBP represents the model without temporal based features into consideration, NS-RBP represents the model without social based features into consideration.

Table 4. The performance of RBP with deleting the kth feature.

Model	Precision	Recall	F1	Accuracy
NU-RBP	0.770	0.603	0.676	0.683
NC-RBP	0.763	0.604	0.674	0.674
NT-RBP	0.767	0.588	0.666	0.688
NS-RBP	0.759	0.623	0.684	0.723
RBP	0.876	0.868	0.872	0.871

The measurement results of above mentioned features are shown in Table 4, where the larger the value is, the less important the feature is. From the table, we clearly observe that basically the descending order of importance for all features is S > C > T > U. More specifically, we first can see that social based features are the most important factors, which means that the stronger social ties are individually more influential, thus retweeting behaviors are more likely to happen each other. This conclusion agrees with the views reported in [18]. Second, an interesting content is the significant indictor to trigger more retweeting. Third, time based feature is relatively important. This is because a large number of tweets are generated in all the time on social network platforms, if a follower has no the same habit of posting time with her followee, she is more likely to not see the tweets posted by the followee. Thus, time period especially most recent interaction affects future retweeting behaviors. Lastly, the performance for user based feature works the worst than other. On the one hand, one possible reason is that the type of selected users is lack in our constructed dataset, this causes less obvious differentiate with different users. On the other hand, it also indicates that social authority is also a significant factor for predicting users' retweeting behaviors.

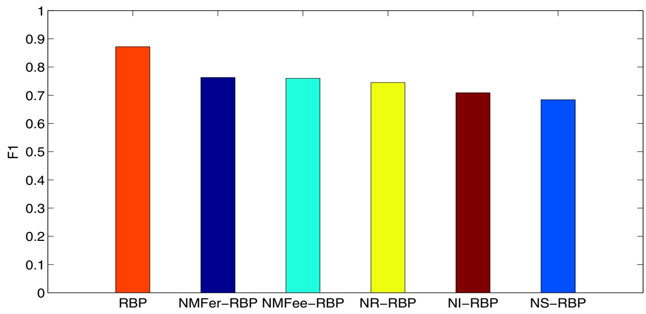

Fig. 1. Comparison of importance of each feature in the social tie feature

As we see that the best features are from the group of social tie feature, we further explore the effectiveness of each feature in the group by removing each feature and examining how the prediction performance is affected in terms of F1. Similarly, we denote RBP without mutual_follower based feature as NMFer-RBP, RBP without mutual_followee based feature as NMFee-RBP, RBP without num_interaction based feature as NI-RBP, and RBP without relationship based feature as NR-RBP. The performance is shown in Figure 1. From the figure, we can clearly conclude that past frequent interactions are more likely to retweet in the future.

6 Conclusion

In this paper, we focus on retweeting behaviors prediction in social network. Specifically, we cast the retweet prediction problem as a classification task. We extract four categories of features including user status, content, temporal, and social tie features from the observed retweets for the prediction task. Furthermore, these features are incorporated into Support Vector Model to predict the class label of candidate. Finally, we collect a large number of data from Twitter, and validate the effectiveness and efficiency of our approach on the constructed dataset. The experimental results clearly show that our approach outperforms other baselines with a significant margin.

Acknowledgments. This work was supported by National Key Technology R&D Program(No.2012BAH46B03), and the Strategic Leading Science and Technology Projects of Chinese Academy of Sciences(No.XDA06030200).

References

1. Abdullah, N.A., Nishioka, D., Tanaka, Y., Murayama, Y.: User's Action and Decision Making of Retweet Messages towards Reducing Misinformation Spread during Disaster. Journal of Information Processing 23(1), 31–40 (2015)
2. Boyd, D., Golder, S., Lotan, G.: Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In: HICSS, pp. 1–10 (2010)
3. Yang, Z., Guo, J., Cai, K., Tang, J., Li, J., Zhang, L., Su, Z.: Understanding retweeting behaviors in social networks. In: CIKM, pp. 1633–1636 (2010)
4. Can, E.F., Oktay, H., Manmatha, R.: Predicting retweet count using visual cues. In: CIKM, pp. 1481–1484 (2013)
5. Cheng, J., Adamic, L., Dow, P.A., Kleinberg, J.M., Leskovec, J.: Can cascades be predicted? In: WWW, pp. 925–936 (2014)
6. Derczynski, L., Maynard, D., Aswani, N., Bontcheva, K.: Microblog-genre noise and impact on semantic annotation accuracy. In: HT, pp. 21–30 (2013)
7. Feng, W., Wang, J.: Retweet or not?: personalized tweet re-ranking. In: WSDM, pp. 577–586 (2013)
8. Kanavos, A., Perikos, I., Vikatos, P., Hatzilygeroudis, I., Makris, C., Tsakalidis, A.: Modeling ReTweet diffusion using emotional content. In: Iliadis, L. (ed.) AIAI 2014. IFIP AICT, vol. 436, pp. 101–110. Springer, Heidelberg (2014)
9. Liu, L., Tang, J., Han, J., Jiang, M., Yang, S.: Mining topic-level influence in heterogeneous networks. In: CIKM, pp. 199–208 (2010)
10. Luo, Z., Osborne, M., Tang, J., Wang, T.: Who will retweet me?: finding retweeters in Twitter. In: SIGIR, pp. 869–872 (2013)
11. Macskassy, S.A., Michelson, M.: Why do people retweet? anti-homophily wins the day! In: ICWSM (2011)
12. Naveed, N., Gottron, T., Kunegis, J., Alhadi, A.C.: Bad news travel fast: A content-based analysis of interestingness on twitter. In: WebSci, p. 8 (2011)
13. Peng, H.K., Zhu, J., Piao, D., Yan, R., Zhang, Y.: Retweet modeling using conditional random fields. In: ICDMW, pp. 336–343 (2011)
14. Petrovic, S., Osborne, M., Lavrenko, V.: RT to win! predicting message propagation in twitter. In: ICWSM (2011)
15. Suh, B., Hong, L., Pirolli, P., Chi, E.H.: Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In: SOCIALCOM, pp. 177–184 (2010)
16. Yang, S.H., Kolcz, A., Schlaikjer, A., Gupta, P.: Large-scale high-precision topic modeling on twitter. In: SIGKDD, pp. 1907–1916 (2014)
17. Zaman, T.R., Herbrich, R., Van Gael, J., Stern, D.: Predicting information spreading in twitter. In: NIPS, pp. 17599–17601 (2010)
18. Zhang, J., Liu, B., Tang, J., Chen, T., Li, J.: Social influence locality for modeling retweeting behaviors. In: IJCAI, pp. 2761–2767 (2013)
19. Zhang, Q., Gong, Y., Guo, Y., Huang, X.: Retweet behavior prediction using hierarchical dirichlet process. In: AAAI (2015)