Cognitive Computation

Resonance Elicits Diffusion: Modeling Subjectivity for Retweeting Behavior Analysis

--Manuscript Draft--

Manuscript Number:	COGN-D-14-00016R1
Full Title:	Resonance Elicits Diffusion: Modeling Subjectivity for Retweeting Behavior Analysis
Article Type:	S.I. : Sentic Computing
Keywords:	Twitter; subjectivity; retweeting behavior; LDA; sentiment analysis.
Corresponding Author:	Songxian Xie Changsha, CHINA
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	
Corresponding Author's Secondary Institution:	
First Author:	Songxian Xie
First Author Secondary Information:	
Order of Authors:	Songxian Xie
	Jingtao Tang
	Ting Wang
Order of Authors Secondary Information:	

Dear Amir Hussain,
Your comments and those of the reviewers were highly insightful and enabled us to greatly improve the quality of our manuscript. In the following pages are our point-by-point responses to each of the comments.
We hope that the revisions in the manuscript and our accompanying responses will be sufficient to make our manuscript suitable for publication in Cognitive Computation.
We shall look forward to hearing from you at your earliest convenience.
Yours sincerely,
Songxian Xie.

Responses to the comments of Reviewer #1:
1. The abstract seems to be not concise enough for summary. I suggest to keep the most important information about the proposed method and move some content to the section 1.
Response: This is an excellent point. We have rewrited the abstract and try to make it concise by only keeping such information as the problem we study, our proposed method and the effect of our method.
2. The section 1 and 2 overlap in some work. It should be better if combine two of them together.
Response: We agree that there are some works cited in both section 1 and 2. Section 1 is an overall introduction to the problem, and our mothed. The works we cite are used to support the claims we have made. Section 2 describe the works related to our method, so we give an detail introduction to every work we cite. The works overlapped take different effect in two section, and we have tried to make it more clear in the revision.
3. I am not clear about the definition 2, which appears to use the same model in definition 1, but different parameters.
Response: Definition 2 is the subjectivity model of a tweet, as we need to calculate the subjectivity similarity between the tweet and a user for retweeting analysis. It takes the same form of subjectivity model of a user, but different parameters. We have made it clear and explained the parameters of it in the revision.
4. The paper did not mention how to get eq. 3. Does this equation will adapt to other application of opinion analysis.
Response: Eq.3 is the transforming of the sentiment values of SentiStrength. SentiStrength output two sentiment values: one (within [1,5]) represents positive sentiment value, the other (within [-5,-1]) represents negative sentiment value. However, in order to use such output of SentiStrength in our subjectivity model, we need to transform it into one single value to represent sentiment, so we design eq.3. It is only used in our method, and need not to adapt to other application of opinion analysis.
5. The subjectivity model in algorithm 1 is based on LDA. Is there any original improvement for the algorithm? How does the paper validate the effectiveness of the proposed algorithm. More discuss should be concentrated on this algorithm.
Response: We agree that we should put more discuss on algorithm 1. In algorithm 1, LDA is trained on users' historical tweets and then used to find topics for each user and tweet. As we dicuss in section 3.3.1, LDA is more general in the concept of topic and more suitable for tweets to overcome the data sparsity. The effectiveness is validated qualitatively and quantitatively in the experiment by combining it with the retweeting analysis.
6. In the section 3.4, the author used the cosine similarity metric. More other similarity metrics are expected to compared in the performance evaluation.
Response: Great points. We have clarified why we use the cosine similarity metric in our mothod in section 3.4 in the revision.
7. What is the difference of star(*) and dagger in the table 2?
Response: Star(*) represents the model that outperforms the baseline significantly with a t-test ($p < 0.05$). Dagger represents the model that outperforms the LUO's method significantly with a t-test ($p < 0.05$).
8. For the fig.5, the author should give more explanation for the meaning of different colors.

Response: Thanks for the point. We have explained the meaning of the color in the caption of fig.5 in the revision.

Responses to the comments of Reviewer #2:
1. I would like to know more about the significance test method and quantitative significance analysis in more detail.
Response: In our experiment we adopt a t-test with significance level 5% to determine whether the performance improvement is significant.
2. I have to admit that I consider 5-fold cross-validation as unlucky in two respects: 1) reproducibility is limited, so the folds should be made available for download.2) there is usually no development part.
Response: We agree to the two points. As for reproducibility, as per request from Twitter, the folds can not be available for downloading publicly, but we will provide to those who require them for research.
3. Other than that, I am not sure if the DOI is really needed for each reference and some images are really (too?) small too read.
Response: Thanks. We have removed the DOI for each reference, and adjusted the images in the revision.

Responses to the comments of Reviewer #3:
1. The manuscript is well-written and is centered on an interesting topic. However, it is a bit lacking on clarity and presentation. Some bad English constructions, mix of British and American English, grammar mistakes, and misuse of articles: the manuscript needs to be proofread by a native English speaker.
Response: Thanks a lot. We have tried to correct the errors and asked a native English speaker to proofread the manuscript during the revision.
2. Authors are advised to view and integrate more recent approaches to opinion mining and natural language processing, e.g., see recent issues of Cognitive Computation, IEEE Intelligent Systems on concept-level sentiment analysis (vol. 28, no. 2 and 3) and IEEE Computational Intelligence Magazine on noetic NLP (vol. 9, no. 1 and 2).
Response: Thanks. We have gone over the recommended issues and integrate a few recent works of opinion mining and natural language processing into our work.

Cognitive Computation manuscript No.
(will be inserted by the editor)

Resonance Elicits Diffusion: Modeling Subjectivity for Retweeting Behavior Analysis

Songxian Xie · Jintao Tang · Ting Wang

Received: date / Accepted: date

Abstract Retweeting is the core mechanism of information diffusion on Twitter, few studies have investigated the subjective motivation of a user to retweet a message. In this paper, in light of psychological theory, we assume that a tweet is more likely to be retweeted by a user because of similar subjectivity, and propose a subjectivity model to combine both the topics and opinions to model subjectivity. With state-of-the-art topic model and sentiment analysis techniques, we establish subjectivity model by finding topics and determining opinions towards these topics from user-generated content simultaneously. We evaluate our model in the retweeting analysis problem to verify its impact on retweeting and effectiveness in the retweeting prediction performance.

Keywords Twitter · subjectivity · retweeting behavior · LDA · sentiment analysis

1 Introduction

Twitter is well-known for its freedom of publishing short messages (i.e. tweets), and viral spreading of information across complex social networks. In addition to large amounts of User-Generated Content (UGC), Twitter provides its social network functions for connection, communication and information diffusion by allowing users to message one another directly and follow one another publicly. The complex networks and large content volume of Twitter provide researchers with insights into people's social behaviors on a scale that has never been possible [36].

Songxian Xie
School of Computer, National University of Defense Technology
Tel.: +86-0731-84574627
E-mail: xsongx@nudt.edu.cn
Jintao Tang
Jttang@nudt.edu.cn
Ting Wang
tingwang@nudt.edu.cn

Information diffusion is a challenging problem which might be investigated on Twitter, because retweeting convention and complex networks of Twitter have provided an unprecedented mechanism for the spread of information despite the restricted length of tweets [19]. Actually almost 25% of the tweets published by users are retweeted from others [44]. Therefore, it is important to understand how retweeting behavior works so as to help study information diffusion on Twitter.

Although several works have concentrated on analyzing retweeting habits and influencing factors [4,21,37], most of them are generic, not user-oriented. From the point of a user, retweeting is a process that includes reading the tweet, estimating the content and deciding to share, and the crucial part of the process is to estimate whether a tweet contains information interesting to the user who might find it worth shareding. Therefore in this study we focus specifically on analyzing the retweeting behavior from the user modelling perspective.

Previous studies on retweeting analysis have shown that an enriched user model gives coherent and consistent explanation for retweeting motivation[1,23,12]. Specifically, users have been modeled from four types of information: profile features ("**Who you are**"), tweeting behavior ("**How you tweet**"), linguistic content ("**What you tweet**") and social network ("**Who you tweet**") [29]. Despite demographic profile, tweeting habits and network structure might determine the source and scope of information users could be exposed to, topics of interest encapsulated in rich linguistic content have been proved consistently dependable for retweeting behavior explanation. For example, whether a tweet will be propagated largely depends on its identification with the interests of users [30,15]. However, beyond merely publishing news and events, Twitter has become a platform where different opinions are presented and exchanged by allowing users to publish subjective messages on topics they are interested in. It is demonstrated that UGC with rich sentimental information can trigger more attention, feedback or participation [36], and tweets with high emotional diversity have a better chance of being retweeted [31]. Until recently, most efforts have been trying to find out whether and how sentiment of a tweet will influence its spreading, but none has realized that although users receive a great many of tweets on different topics every day, whether a tweet will be retweeted depends on the subjective choice of users.

Subjective initiative nature of human determines that his behavior pattern is subjectivity driven. Psychologist have identified subjectivity as the underlying factor that influences taking what behaviors to process incoming stimul [25]. According to the theory of Biased Assimilation, people are prone to choose and diffuse information according to their own biased subjectivity [18,38]. In this study we explore the UGC on Twitter to model the subjectivity of users, and investigate whether the subjectivity model could benefit the retweeting behavior analysis. Intuitively, subjectivity can be represented as topics and opinions articulated in the information generated by users on Twitter. We use the state-of-the-art topic model to find the topics users are talking about, and sentiment analysis techniques to determine user's opinions towards these topics from UGC simultaneously. We evaluate our model on the retweeting analysis problem to verify its impact on retweeting behavior.

Modelling subjectivity on Twitter is a challenging task because of the sparsity of textual information and the dynamic of topics and opinions. However, we are interested in understanding retweeting behavior at a local level rather than at a global level, since most of time retweeting pertains to a local network consisting of the tweet publisher and followers. The relatively tiny size and topic homophily of local network degrade the impact of sparsity. Given the biased nature of subjectivity, while new information may arise and old information may change their meaning, biased subjectivity is likely to be more consistent and less prone to external perturbations, therefore subjectivity model of a user is less likely to be influenced by changes of topics and opinions on Twitter.

Our work aims to define and establish the subjectivity model and identify the role of subjectivity in the processes of information diffusion on Twitter. Our contributions can be summarized as follows:

- In light of psychological theory, we firstly put forward a formal definition of subjectivity model which model both the topics and opinions simultaneously.
- Based on the state-of-the-art topic model and sentiment analysis techniques, we build subjectivity model from UGC on Twitter and apply it to the retweeting behavior analysis problem.
- We systematically evaluate the effectiveness of the subjectivity model in a series of experiments. It is demonstrated that our model outperforms other UGC-based models in rewteeting prediction and gives the most significant improvement over an off-the-shelf predicting model.

The rest of the paper is organized as follows: section 2 gives the related works to our research, the proposed subjectivity model is defined and specified in section 3, the qualitative and quantitative evaluation is described in section 4, and Section 5 summarizes the paper and points out future work.

2 Related Work

Retweeting Analysis. A lot of works have analyzed the characteristics of retweeting, examining factors that lead to increased retweetability and designing models to estimate the probability of being retweeted. As for factors influencing retweetability, Suh *et al.* [37] found that tweets with URLs and hashtags were more likely to be retweeted. Macskassy and Michelson [23] found that models derived from tweet content could explain most of retweeting behaviors. Comarela *et al.* [10] found previous response to the tweeter, the tweeters sending rate, the freshness of information, the length of tweet could affect followers response to retweet. Starbird and Palen [34] found that tweets with topical keywords were more likely to be retweeted. There are also many works extending the analysis to build retweeting prediction model. Osborne and Lavrenko [30] introduced features such as novelty of a tweet and the number of times the author is listed to train a model with a passive aggressive algorithm, and found that tweet features added a substantial boost to the performance. Jenders *et al.* [19] analyzed the "obvious" and "latent" features from structural, content-based, and sentimental aspects and found a combination of features covering all aspects was the key to high prediction quality. Naveed *et al.* [27,26] introduced interestingness based on such features as sentiments and topics to predict the probability of retweeting for an individual tweet. Feng and Wang [12] proposed a feature-aware factorization model to rerank the tweets according to their probability of being retweeted. Pfitzner *et al.* [31] proposed a new measure called emotional divergence and showed that high emotional diverse tweets have higher chances of being retweeted.

All works introduced above have tried to answer the question of "whether and why a tweet will be retweeted by anyone". But they are weak to capture "whether a tweet is retweetable from a user-centric perspective considering the interests and opinions ". In this paper, we will try to answer this question by building a subjectivity model which can capture both the interests and opinions of users.

User Modelling. With the popularity of social media, researchers have begun to pay close attention to model users on the massive amount of UGC. These studies provide researchers with insights into user online behaviors. Hannon *et al.* [14] proposed that Twitter users can be modeled by tweets content and the relation of Twitter social network. Macskassy and Michelson [23] discovered user's interests by leveraging Wikipedia as external knowledge to determine a common set of high-level categories that covers entities in UGC. Ramage *et al.* [32] made use of topic models to analyze tweets at the level of individual users with 4S dimensions, showing improved performance on tasks such as post filtering and user recommendation. Xu *et al.* [43] proposed a mixture model which incorporated three important factors, namely breaking news, friends' timeline and user interest, to explain user posting behavior. Pennacchiotti and Popescu [29] proposed a comprehensive method to model users for user classification, and confirmed the value of in-depth features by exploiting the UGC, which reflect a deeper understanding of the Twitter user and the user network structure.

Few works have identified the correlation between the opinions of users and their behaviors, motivated by the observation, we put forward subjectivity model to combine both interests and opinions to model a user.

Sentiment Analysis. Sentiment analysis is a popular research area and previous researches have mainly focused on reviews or news comments, and approaches can be grouped into three main categories: keyword spotting, lexical affinity, and statistical methods [6]. Recently, researchers began to pay more and more attention to social media such as Twitter. Hu *et al.* [16] interpreted emotional signals available in tweets for unsupervised sentiment analysis by providing a unified way to model two main categories of emotional signals: emotion indication and emotion correlation. Jiang *et al.* [20] focused on target-dependent Twitter sentiment classification, and proposed a method to improve performance by taking target-dependent features and related tweets into consideration. Asiaee T. *et al.* [2] presented a cascaded classifier framework for per-tweet sentiment analysis by extracting tweets about a desired target subject, separating tweets with sentiment, and setting apart positive from negative tweets. Hu emphet al. [17] extracted sentiment relations between tweets based on social theories, and proposed a novel sociological approach to utilize sentiment relations between messages to facilitate sentiment classification. Motivated by sociological theories that humans tend to have consistently biased opinions, Calais Guerra *et al.* [5] addressed challenges of topic-based real-time sentiment analysis by proposing a novel transfer learning approach with a suitable source task of opinion holder bias prediction. Thelwall *et al.* [40,39] designed SentiStrength, an algorithm for extracting sentiment strength from informal English text by exploiting the grammar and spelling styles in typical social media text. In this paper, we adopt SentiStrength for sentiment analysis to build our subjectivity model , because the fine-grained sentiment strength it outputs could give us more detailed opinions than binary polarity.

3 Subjectivity Model

In this section, we firstly give the definition of subjectivity model, then describe the method of building subjectivity model, and finally apply subjectivity model to the retweeting analysis problem.

3.1 Definition

Subjectivity has been extensively studied by psychologists to characterize the personality of a person based on his historic behaviors and remarks [11]. Linguists define the subjectivity of language as the speakers always show their perspectives, attitudes and sentiments in their discourses [35]. Social media provides users a platform to express their opinions towards topics of interest to show their personal subjectivity by publishing short messages. Therefore, for the term "subjectivity" , we refer to both topics and opinions articulated in the UGC. That is, we model subjectivity not only by interests of users, but also by "**what they think about the interests**". Here we firstly give our definition of subjectivity model on Twitter, while we emphasize that our model can be adapted to other social networks platforms as well.

For a set of users U on Twitter, we assume there is a topic space T containing all topics they talk about, and a sentiment valence space S evaluating their opinions towards these topics. As for S, it is often considered as a binary polarities consisting of positive and negative sentiment, however we argue that a more fine-grained sentiment space will indicate more detailed opinions of users.

Definition 1 (Subjectivity Model For User) The subjectivity model $P(u)$ of a user $u \in U$ is the combination of a set of topics $\{t_i\}$ the user talks about in a topic space T and the user's opinions $\{O_i\}$ towards the topics.

$$P(u) = \{(t_i, w_u(t_i), \{d_{u,t_i}(s_i)\)\}, |t_i \in T, s_i \in S\} \quad (1)$$

where:

- with respect to the given user u, for each topic $t_i \in T$, its weight $w_u(t_i)$ represents the distribution of the user's interests on it, subject to $\sum_{i=1}^{|T|} w_u(t_i) = 1$.
- opinion O_i of user towards topic t_i is a target-dependent sentiment distribution $d_{u,t_i}(s_i)$ over sentiment valence space S, subject to $\sum_{i=1}^{|S|} d_{u,t_i}(s_i) = 1$.

Users express themselves by posting tweets on Twitter, and each tweet generated by a user can be considered subjective in that it also contains topics and opinions. So we also give a subjectivity model definition for a tweet as follows:

Definition 2 (Subjectivity Model For Tweet) The subjectivity model $P(m)$ of a tweet m is the combination of a set of topics $\{t_i\}$ it talks about, and the opinions $\{O_i\}$ it expresses.

$$P(m) = \{(t_i, w_m(t_i), \{d_{m,t_i}(s_i)\}\}, |t_i \in T, , s_i \in S\} \quad (2)$$

where:

- with respect to the given tweet m, for each topic $t_i \in T$, its weight $w_m(t_i)$ represents the topic weight of the tweet, subject to $\sum_{i=1}^{|T|} w_m(t_i) = 1$.
- opinion O_i of tweet m towards topic t_i is a target-dependent sentiment distribution $d_{m,t_i}(s_i)$ over sentiment valence space S, subject to $\sum_{i=1}^{|S|} d_{m,t_i}(s_i) = 1$.

The definition of subjectivity model given above is in an abstract form by using latent concepts of topics and opinions which need to be derived from UGC. In this paper we combine subjectivity model with retweeting analysis problem and concrete the subjectivity model in such problem settings.

3.2 Retweeting Analysis Problem Statement

Retweeting is the core mechanism of information diffusion on Twitter. Many factors have been proved to influence retweeting behavior [37,23,10], however few researches have investigated the subjective motivation of a user to retweet a message. Therefore we will study whether subjectivity model can help understanding underlying reasons of a user's retweeting behavior.

In fact the likelihood of a tweet to be retweeted depends on both context constraints and its content. The context such as the network of the author and the time a tweet is published affects whether the tweet will be retweeted. A tweet with only few or passive followers is less likely to be retweeted, and a tweet published in the night have less chance to be retweeted than daytime. Apart from the context constraints, a tweet is more likely to be retweeted by subjective users who find its content worth to. Therefore, we are not interested in modelling the tweet by itself just as other researches [27,26,31], but understanding how the content resonate with the users who might want to retweet it. We put a much stronger emphasis on the content and try to model the user's subjective decision by deriving latent topics and opinions from UGC. Actually, none of contextual factors has any influence on the content of the tweet, therefore we deliberately ignore context constraints to avoid introducing contextual bias into our analysis by proposing Hypothesis 1.

Hypothesis 1 (H1) A tweet is evenly visible to the followers who subscribe to it by following its publisher.

The rationale behind this hypothesis is, the motivation of a user to retweet a message lies in that the user considers only the tweet content arousing his resonance without context perturbation.

On Twitter, the "following" relationship is a strong indicator of a phenomenon called "homophily", which has been observed in many social networks. Homophily is a phenomenon that people connected in a social network "are homogeneous with regard to many socio-demographic, behavioral, and intra-personal characteristics" [24]. In other words, homophily implies that a user follows another user because he finds they share similar interests. According to the principle of homophily, we put forwards the concept of **Local Topic Space**, which can be defined as:

Definition 3 (Local Topic Space) In a local network consisting of a user and his followers, all users concentrate on limited topics derived from their UGC, and these topics form a local topic space.

Since most of time retweeting pertains to a local network, we limit our research in understanding retweeting behavior at a local level rather than at a global level, and the relatively tiny size and topic homophily of local network degrade the impact of data sparsity.

According to our Hypothesis 1, if a tweet is published, all followers of its author will receive it in time, and followers are likely to retweet it if they find it worthwhile. Thus the retweeting analysis problem we study can be stated as follows:

Let F, A, M denote the follower set, author set and tweet set respectively. For each tweet m ($m \in M$) and its listener f ($f \in F$), we can define a quadruple $< f, a, m, r_{fam} >$ where:

- a ($a \in A$) is the author of the tweet m and f ($f \in F$) is a follower of author a.
- r_{fm} is a binary label indicating whether m is retweeted by f.
- Our work focuses on using subjectivity model to analyze the relation between the subjectivity of a follower f and his retweeting behavior. Hence we transform the quadruple into the Local Topic Space T formed by the author a and his followers $\{f\}$, and represent f, a, m with their subjectivity models to analyze their relations with the label r_{fm}.

3.3 Establishment of Subjectivity Model

According to the definition of subjectivity model , there are two distributions to model the subjectivity: the topic distribution and the opinion distribution for each topic. Both of them need to be inferred from historic content produced by users. However, content analysis on Twitter is challenging: the volume of tweets is so huge while a single tweet is very short with a limit of 140 characters, and informal languages are widely used, which make many supervised learning approaches and natural language processing techniques invalid [7]. Hence effectively modeling content on Twitter requires techniques that can readily adapt to these challenges and require little supervision. With state-of-the-art topic model and sentiment analysis techniques, we establish subjectivity model by identifying topics and opinions in an unsupervised way simultaneously.

3.3.1 Topic Analysis

The topics of a tweet are latent and have to be inferred from its content. Previous studies have tried to identify topics from tweets by finding key words [9], extracting entities [1], linking tweets to external knowledge categories [23], or using semantic framework [13]. However, the sparsity is a main problem for these methods because even if users have common local topics they still might refer to a topic with different vocabulary. Works show that topic models such as **Latent Dirichlet Allocation (LDA)** model and its extensions[3, 41] have been efficient ways to characterize latent topics of large volume corpus. Topics of LDA are broader in concept, since a single topic consists of the whole collection of related words. Therefore we adopt a user-level LDA model to find latent topics for users in their Local Topic Space, and the generative process can be graphically represented using plate notation in Figure 1. To distill the topics that users

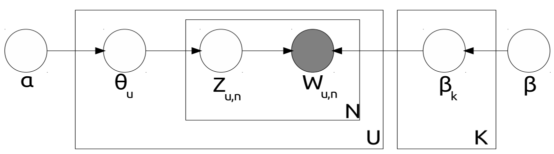

Fig. 1 Plate illustration of the user-level LDA model.

are interested in, documents of LDA should naturally correspond to tweets content. As our goal is to understand the topics that each user is interested in rather than the topics each single tweet talks about, we aggregate the tweets published by each user into a single document, and replace documents of LDA with the aggregated tweet documents. So a document stands for a user in our model, and a user can be represented as a multinomial distribution over topics, which corresponds to the topic weights of the user's subjectivity model.

Formally, given a set of users U and the number of topics K, a user u ($u \in U$) could be represented by a multinomial distribution θ_u over topics with a Dirichlet prior parameterized by α. A topic k ($k \in K$) is represented by a multinomial distribution β_k with another Dirichlet prior parameterized by η. The parameters θ_u and each β_k can be estimated by Gibbs sampling or variational inference. A variational inference-based topic model package Gensim [33] is used in our work.

3.3.2 Opinion Analysis

Users often express opinions towards their topics of interest by publishing topic-related tweets. In order to explore the opinions of users, we need to understand the sentiment embedded in each tweet. Machine learning and rule-based approaches are two main techniques of sentiment analysis. Machine learning approaches often need labelled data for the training process, which is often impossible for Twitter because of the large volume of tweets and its dynamic language characteristics. Therefore we adopt rule-based approaches, which could adapt to Twitter with good flexibility by changing its particular characteristics into rules [40, 16].

The SentiStrength package has been built especially to cope with sentiment analysis in short informal text of social media [40]. It combines lexicon-based approaches with sophisticated linguistic rules adapted to social media, which is suitable for analyzing sentiment of tweets in our research settings. SentiStrength assigns two values to each tweet standing for sentiment strengths: a positive (within $[1,5]$) and a negative (within $[-5,-1]$) sentiment value, both ranging from 1 to 5 on absolute integer scales, with 1 denoting neutral sentiment and 5 denoting highest sentiment strength. Sentiment assigned by SentiStrength is not a simple binary polarity but a fine-grained strength, which can catch fine opinions in a user's subjectivity model. For the convenience of calculation, we map the output of SentiStrength to a single-scaled sentiment valence space $[0,8]$ as follows:

$$o = \begin{cases} p+3 & if\ |p| > |n| \\ n+5 & if\ |n| > |p| \\ 4 & if\ |p| = |n| \end{cases} \tag{3}$$

Where p denotes the positive setiment value and n denotes negative sentment value. In the sentiment valence space $[0,8]$, value 4 and 5 indicates neutral sentiment, while values above 5 indicate positive sentiment and values below 4 indicate negative sentiment. With the sentiments of all tweets, we can aggregate opinion towards a topic as a sentiment distribution over sentiment valence space $[0,8]$.

3.3.3 Concreting Subjectivity Model

With statistical topic analysis and opinion analysis described above, we can concrete subjectivity model in a local network settings now. For user set U of a local network, we denote tweet set published by a user u as $M_u = \{m_i | i \in [1,\cdots,N]\}$.

Each M_u is concatenated to a single document d_u to construct Local Topic Space $T = \{t_i | i = 1, \cdots, K\}$. A topic model is built with parameter θ representing the distribution of users over topics in the Local Topic Space T, and parameter β represents the distribution of topics over the vocabulary of all tweets. SentiStrength is applied to each tweet m in collection M_u and outputs sentiment strength s_m for tweet m. We build the subjectivity model $P(u)$ for user u as Algorithm 1:

Algorithm 1 Establishment of subjectivity model .

Input:
The user set of a local network, U;
The tweet set published by each user u, M_u;
Output:
The subjectivity model for each user u, $P(u)$;

```
Topic analysis with a user-level LDA as Section 3.3.1, getting a topic model P(θ,β|M_u,U);
for all tweet m ∈ M_u do
    Sentiment analysis as Section 3.3.2, outputting sentiment of m, s_m;
end for
for user u ∈ U do
    the topic distribution is the corresponding component of parameter θ, θ_u;
    the topics he tweets about are Z_u = {t|p(t|θ_u) > 0, t ∈ T};
end for
for tweet m ∈ M_u do
    topics of m can be identified by the topic model, Z_m = {t|p(t|θ,β,Z_u) > 0, t ∈ T};
end for
for each topic t ∈ Z_u do
    for sentiment value s ∈ S do
        count the number of tweets which talk about topic t with sentiment value s, N_s = Σ_{m∈M_u} I(s_m), if s_m = s & t ∈ Z_m;
    end for
    calculating opinion towards topic t, O_t = { N_s / Σ_{s∈S} N_s };
end for
establishing subjectivity model of user u,
```

$$P(u) = \left\{ \left(t, p(t|\theta_u), \left\{ \frac{N_s}{\sum_{s \in S} N_s} \right\} \right) | t \in Z_u, s \in S \right\} \tag{4}$$

```
return P(u);
```

In the algorithm, we assume the sentiment of tweet m is related to every topic it talks about in Z_m for simplicity. Accordingly subjectivity model $P(m)$ for tweet m as:

$$P(m) = \{(t, p(t|\theta, \beta), d_{m,t}(s)) = 1.0 | t \in Z_m, s \in S\} \tag{5}$$

Noting that, the opinion towards each topic is a distribution of 1.0 on a single sentiment value s of tweet m.

3.4 Retweeting Analysis With Subjectivity Model

To understand the underlying reasons why a user retweet a message, we try to simulate the subjective decision-making procedure by investigating the relationship among the subjectivity models of a tweet, its author and followers. We assume that a user retweet a message because the user not only finds its topics interesting but also shares similar opinions towards these topics. In other words, if the subjectivity models of a tweet and a user are similar enough, the user will have a very high probability to retweet it. We call this phenomenon as "resonance", and assume that the resonance between a tweet and users will elicit retweeting behavior. With the subjectivity models built for users and tweets, we can define a similarity measurement to quantify the resonance among them.

Formally, for a tweet m, the corresponding author a, and a list of followers $F = \{f\}$, for each $f \in F$, we can define a quadruple $< f, a, m, r_{fm} >$ as Section 3.2. We firstly build subjectivity model $P(u)$ for each user $u \in F \bigcup a$ and $P(m)$ for tweet m, then define the similarity measurement as follows:

$$Sim(m, f) = similar(P(m), P(f)) \tag{6}$$

according to Equation 4,5:

$$\begin{aligned} Sim(m, f) = \lambda * Dist\left(p(t|\theta, \beta, Z_m), p\left(t|\theta_f, Z_f\right)\right) \\ + (1 - \lambda) * \left(\sum_{t \in T} Dist\left(O_{m,t}, O_{f,t}\right) \right) \end{aligned} \tag{7}$$

where

- λ is the coefficient used to control the proportions of topic similarity and opinion similarity in the holistic subjective similarity. We initiate it by setting $\lambda = 0.5$, and adjust its value in the range of $[0, 1]$ to optimize the best performance of our model.
- $Dist()$ is the similarity measurement between two distribution. The similarity between two distributions can be calculated with methods such as the cosine distance [8] or the Jensen-Shannon Divergence [42]. We adopt the cosine distance because it performs better than other measurements in our research settings. It is defined as:

$$Dist(\theta_m, \theta_u) = \frac{\theta_m \cdot \theta_u}{\parallel \theta_m \parallel \parallel \theta_u \parallel} \tag{8}$$

where θ_u denotes the distribution of user u and θ_m denotes the distribution of tweet m.

We also assume that a user might retweet another user because of their subjective resonance. Accordingly we define similarity between author a and follower f as:

$$\begin{aligned} Sim(a, f) = \lambda * Dist\left(p(t|\theta_a, Z_a), p\left(t|\theta_f, Z_f\right)\right) \\ + (1 - \lambda) * \left(\sum_{t \in T} Dist\left(O_{a,t}, O_{f,t}\right) \right) \end{aligned} \tag{9}$$

Table 1 Retweet Dataset Statistics

Tatget tweets	500
Average number of followers per target tweet	89
Total retweeters	5214
Total non-retweeters	40317

4 Experiment

In this section, we investigate whether subjectivity model can help retweeting analysis with an Twitter dataset.

4.1 Dataset

We adopt an off-the-shelf Twitter dataset of previous work [22], which was created with Twitter API[1]. To build the dataset, 500 randomly selected English target tweets were monitored in the next few days to find followers who would retweet them. Also each target tweet was chosen as starting point to collect historical data of its author and followers. Overall, there are 45,531 followers and 6,277,736 tweets, and 5214 followers who have retweeted at least one target tweet during the monitored period. Summary statistics of the dataset are listed in Table 1.

The relations among a target tweet, its author and followers are illustrated in Figure 2. There is a local network

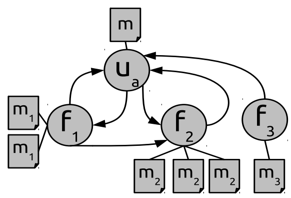

Fig. 2 Relations among a target tweet, its author and followers.

structure for each target tweet as figure shows, consisting of its author and followers.

4.2 Impact Evaluation of Different Factors

In Section 3.4, we model retweeting probability with subjectivity model in the form of similarity measurements 7,9. By setting different value to λ, the measurements can be transformed into different versions to model different factors that might influence user's retweeting behavior, which are:

- **TTF**: **T**opic similarity between **T**weet and **F**ollower ($\lambda = 1$ in measurement 7).
- **OTF**: **O**pinion similarity between **T**weet and **F**ollower ($\lambda = 0$ in measurement 7).
- **STF**: **S**ubjective similarity between **T**weet and **F**ollower ($\lambda \in (0, 1)$ in measurement 7).
- **TAF**: **T**opic similarity between **P**ublisher and **F**ollower ($\lambda = 1$ in measurement 9).
- **OAF**: **O**pinion similarity between **P**ublisher and **F**ollower ($\lambda = 0$ in measurement 9).
- **SAF**: **S**ubjective similarity between **P**ublisher and **F**ollower ($\lambda \in (0, 1)$ in measurement 9).

The six similarity measurements could be grouped into two aspects. One is consisted of TTF, OTF and STF, which is direct and explicit by modelling the tweet and its followers; the other is consisted of TAF, OAF and SAF, which is indirect and implicit by modelling the author and follower. The two aspects reflect properly the local information diffusion structure of Twitter at micro-level as illustrated in Figure 2.

To evaluate the impact of different factors on retweeting behavior, we compare six average similarity scores between 5214 retweeters and 5214 randomly selected non-retweeters. The values of λ for STF and SAF are tuned to produce the largest value difference between retweeters and non-retweeters, which are $\lambda = 0.5$ on our dataset. Figure 3 shows the result. As the figure illustrated, the simi-

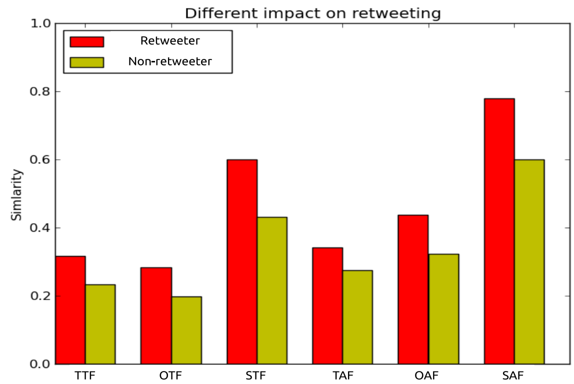

Fig. 3 Impact of different factors on retweeting behavior.

larities scores of retweeters are obviously higher than non-retweeters for all six factors. Specifically:

- TTF score shows that a tweet is more likely to be retweeted by followers who find topics it talks about interesting to them, which is consistent with other studies[23];
- OTF score shows that opinions in a tweet is an important indicator to be retweeted by by followers who hold similar opinions, although other studies[31,26] have shown that sentiment in tweet has impact on retweeting behavior, they haven't consider the opinions of followers and opinion similarity betweet tweet and followers;
- STF score shows the subjective similarity is the most distinguishable feature among the six factors with the largest value difference, which proves the importance of subjectivity model;

[1] https://dev.twitter.com/

– TAF score gives another perspective for retweeting analysis from the topic similarity between author and followers, indicating that followers are more likely to retweet author with similar interests , which verifies the homophily principle of following relation;
– OAF score indicates that similar opinions also influence followers' decision of retweeting another user, which proves opinion homophily of following relation.
– SAF score is interesting in that it implies that subjective similarity between author and followers might cause retweeting, and we call this phenomenon "tight homophily" of following relation because it requires both topic homophily and opinion homophily.

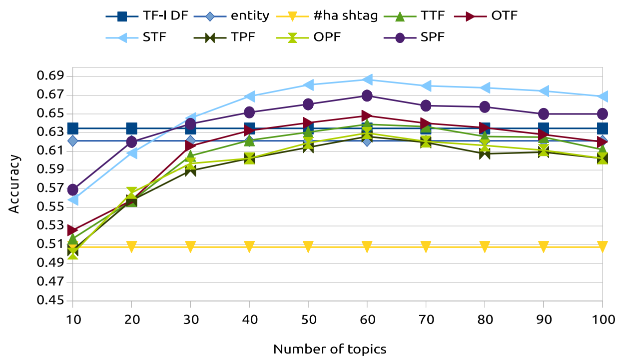

Fig. 4 Comparison of different models.

4.3 Performance of Retweeting Prediction

The main purpose of retweeting analysis is to help users find interesting information from the overwhelming information streams. Retweeting is an important signal of interestingness because users are prone to broadcast their favorite messages to their followers. Thus, the performance of retweeting prediction is a suitable evaluation for the utility of subjectivity model in retweeting analysis problem. In our experiment, we evaluate the subjectivity model in supervised machine learning framework.

As Section 3.2 introduced, the retweeting analysis problem could be formulated as a quadruple $< f, a, c, r_{fm} >$. For retweeting prediction, we need to estimate the label r_{fm} when $m, a, and f$ are known. There are 5,214 retweeters in our dataset who retweet at least one target tweet, so we extract 5214 quadruples as positive instances with their label $r_{fm} = 1$. For the other 40,317 non-retweeters, we also extract quadruples as negative instances with label $r_{fm} = 0$. To avoid unbalance bias of training data, we randomly sample 5,214 negative instances into the test dataset.

4.3.1 Comparison With Other User Models

Firstly the comparison between our model with other UGC-based user models (TF-IDF model [22], entity-based model and hashtag-based model [1]) in retweeting prediction is investigated. As defined in Section 4.2, the six similarities derived from our model are used for comparison, because they model different factors that influence retweeting behavior. For the comparing models, cosine similarities are calculated between tweets and followers. We use the logistic regression classifier of Scikit-learn machine learning package [28], with 5-fold cross-validation on our balance test dataset. Accuracy is our evaluation metric. Performances of our model and all other models are shown in Figure 4. Figure 4 also shows that the impact of topic number of LDA on the predicting accuracy, our model arrives its peak when the number is set to 60, so we fix the topic number as 60 in all our experiment.

As Figure 4 illustrates, the best accuracy of 68.67% is achieved by the STF (Subjective similarity between tweet and followers). The accuracies of TF-IDF model and entity-based model are 63.45% and 62.12%, which are very close to TTF (Topic similarity between Tweet and Followers, 63.88%) and OAF (Opinion similarity between Publisher and Followers, 62.96%). While for hashtag-based model, its accuracy is 50.76% , which is only a little better than random selection (50%) but not significant. The reason might lie in a very low usage of hashtag in our data. The accuracies of the other three model are OTF (Opinion similarity between Tweet and Followers, 64.80%), TAF (Topic similarity between Publisher and Followers, 62.58%) and SAF (Subjective similarity between Publisher and Followers, 66.95%) model. The results show that subjectivity model can better help understanding retweeting behavior than the other user models.

4.3.2 Comparison with Other Factors

In this section, we feed the six similarities of our model as features into a retweeting classification framework to verify the effectiveness of subjectivity model. We compare the performance of our model with method of Luo *et al.* [22] which uses four feature families: Retweet History (follower who retweeted a user before is likely to retweet the user again), Follower Status (for a follower, the number of tweets, followers, friends, being listed and whether he is verified), Follower Active Time (the time users interact with others) and Follower Interests (common interests between tweet and followers, TF-IDF model).

We use LinearSVM of Scikit-learn package to build a retweeting prediction framework, leveraging two different features sets. One includes the six features derived from subjectivity model (marked as "SM6"). The other is the feature

Table 2 Prediction Accuracy of Different Models. Significant improvement over baseline with star(∗) and LUO' model with dagger(‡) ($p<0.05$).

Feature Set	Accuracy(%)
RB	60.85
LUO	68.76 ∗
SM6	69.12 ∗
LUO(⊖)+TTF	69.20 ∗
LUO(⊖)+TAF	71.04 ∗ ‡
LUO(⊖)+OTF	71.88 ∗ ‡
LUO(⊖)+OAF	70.27 ∗
LUO(⊖)+STF	72.86 ∗ ‡
LUO(⊖)+SAF	72.05 ∗ ‡
LUO(⊖)+All	72.93 ∗ ‡

set from Luo *et al.* [22] (marked as "LUO"). We use the same dataset as Section 4.3.1 with 5-fold cross-validation, and accuracy as evaluation metric. In addition, we set a baseline (marked as "RB"), for which followers who have retweeted the author's previous tweets are predicted as retweeters of current tweet. The result is listed in Table 2. The accuracy of baseline is 60.85%, and two prediction models (LUO and our SM6) both outperform the baseline significantly. But the prediction model based on our feature set shows no significant improvement over LUO feature set. The reason might be that our model only tries to reflect the retweeting motivation of users based on content, whereas other important factors associated with retweeting behavior are not considered, such as network topology and tweeting habit of the user, etc.

Since it is proved that subjectivity model outperforms TF-IDF model in Section 4.3.1, which is used in LUO feature set, we propose that retweeting prediction performance could be improved by using features derived from subjectivity model. As denoted by "LUO(⊖)+" in the table, the Follower Interests features of LUO are replaced with our six features one by one. The accuracies are all improved. It shows that our model is of great importance for retweeting prediction. Noticing that, the most significant improvement (LUO(⊖)+STF, 72.86% versus 68.76%) is the subjective similarity feature between tweet and followers, which verifies our assumption that subjective resonance between tweet and followers can be considered as the underlying reason that elicits retweeting behavior. Besides, the improvement by adding subjective similarity features between author and followers (LUO(⊖)+SAF, 72.05% versus 68.76%) is also obvious in that the resonance between author and follower indicates the tight homophily between them. Finally, the last row of table is the complete combination of two sets of features (LUO(⊖)+All) by adding all six features into LUO feature set. The performance shows no significant improvement over adding STF feature only, in that subjectivity model combines both topic and opinion information, and STF is an integral feature to model both topic similarity and opinion similarity between tweet and followers, so it is redundant to add other separate parts.

4.4 Case Study

In this section we give a vivid example to illustrate the subjectivity model and its ability in explaining the retweet behavior. The subjectivity models for one of the 500 target tweets, its author, and two followers (one retweeter, the other non-retweeter) are shown as Figure 5. The right part of each sub-figure illustrates topic distribution and the left part illustrates opinions towards each topic. It is the 14^{th} topic that the tweet talks about in the local topic space. Figure 6 shows top words of the 14^{th} topic, the tweets of author and two followers in a word cloud[2]. Content of the tweet is:

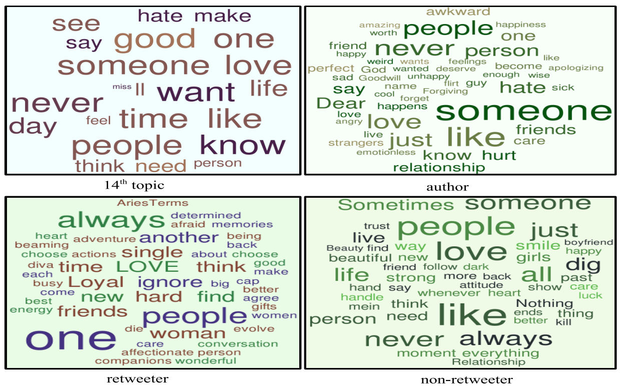

Fig. 6 Word cloud of 14^{th} topic, author and followers.

Tweet: "Sometimes the right person for you was there all along. You just didnt see it because the wrong one was blocking the sight"

The topic of this tweet is about "love between people" and the opinion is neutral, which is in accordance with the 14^{th} topic word cloud in Figure 6 and subjectivity model of tweet in Firgure 5. The author concentrates on the 14^{th} topic with 208 tweets, and his opinions are mainly neutral (as Figure 5, 6 demonstrate). As for two followers, the retweeter has published 250 tweets about two topics (the 14^{th} and 52^{nd} topic) uniformly and his opinions towards the two topics are mainly neutral. While the other one, the non-retweeter has also talked about two topics (14^{th} and 56^{th} topic) with 188 tweets, but he is mainly interested in the 14^{th} topic and his opinion is positive. Although two followers have the same interest (the 14^{th} topic), their different opinions elicit their different decision, which verifies subjectivity model can help better understanding the retweeting behavior not only from topics but also opinions.

[2] We use TagCrowd (http://tagcrowd.com/) to produce word cloud.

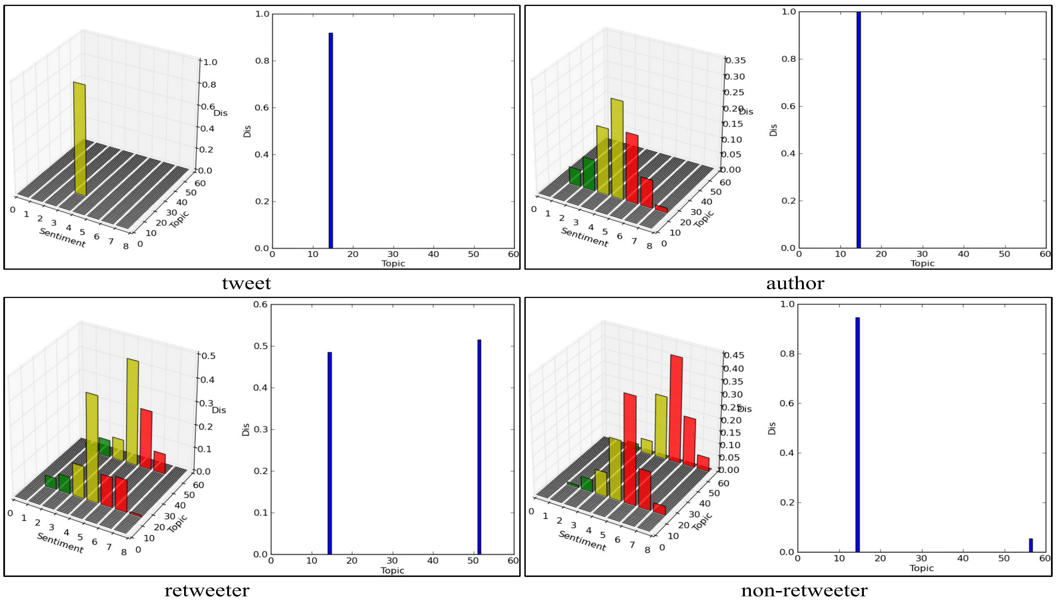

Fig. 5 Subjectivity models of tweet, author and followers. In each opinion distribution sub-graph, color red stands for positive setiment (value>5), green for negative sentiment (value<4) and yellow for neutral sentiment (value 4,5).

5 Conclusion

In this paper, we propose a subjectivity model to analyze user retweeting behavior on Twitter. We assume that retweeting should be elicited by the subjective resonance between the tweet and its followers. We define the subjectivity model formally as the combination of topics and opinions, and we put forward an algorithm to establish the subjectivity model leveraging statistical topic model and sentiment analysis techniques. We demonstrate the effectiveness of our model for retweeting analysis problem and show that subjectivity model is able to reach better understanding of retweeting behavior.

Our future work mainly lie in two directions. Firstly, the subjectivity model is established through simple combination of topics and opinions. It is an interesting direction to establish it under the framework of generative topic-sentiment model, which has been applied in reviews and citation network. Secondly, we will apply subjectivity model to other social networks analysis tasks such as connection prediction and friend recommendation.

References

1. Abel, F., Gao, Q., Houben, G.J., Tao, K.: Analyzing user modeling on twitter for personalized news recommendations. In: Proceedings of the 19th international conference on User modeling, adaption, and personalization, UMAP'11, pp. 1–12. Springer-Verlag, Berlin, Heidelberg (2011)
2. Asiaee T., A., Tepper, M., Banerjee, A., Sapiro, G.: If you are happy and you know it... tweet. In: Proceedings of the 21st ACM international conference on Information and knowledge management, CIKM '12, pp. 1602–1606. ACM, New York, NY, USA (2012)
3. Blei, D., Ng, A., Jordan, M.: Latent dirichlet allocation. the Journal of machine Learning research **3**, 993–1022 (2003)
4. Boyd, D., Golder, S., Lotan, G.: Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. In: 2010 43rd Hawaii International Conference on System Sciences, vol. 0, pp. 1–10. Kauai, HI (2010)
5. Calais Guerra, P.H., Veloso, A., Meira Jr., W., Almeida, V.: From bias to opinion: a transfer-learning approach to real-time sentiment analysis. In: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '11, pp. 150–158. ACM, New York, NY, USA (2011)
6. Cambria, E., Schuller, B., Xia, Y., Havasi, C.: New avenues in opinion mining and sentiment analysis. IEEE Intelligent Systems p. 1 (2013)
7. Cambria, E., White, B.: Jumping nlp curves: A review of natural language processing research. IEEE Computational Intelligence Magazine **9**(2), 48–57 (2014)
8. Cha, S.H.: Comprehensive survey on distance/similarity measures between probability density functions. City **1**(2), 1 (2007)
9. Chen, J., Nairn, R., Nelson, L., Bernstein, M., Chi, E.: Short and tweet: experiments on recommending content from information streams. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10, pp. 1185–1194. ACM, New York, NY, USA (2010)
10. Comarela, G., Crovella, M., Almeida, V., Benevenuto, F.: Understanding factors that affect response rates in twitter. In: Proceedings of the 23rd ACM conference on Hypertext and social media, HT '12, pp. 123–132. ACM, New York, NY, USA (2012)
11. Engbert, K., Wohlschläger, A., Thomas, R., Haggard, P.: Agency, subjective time, and other minds. Journal of Experimental Psychology: Human Perception and Performance **33**(6), 1261–1268 (2007)
12. Feng, W., Wang, J.: Retweet or not?: personalized tweet reranking. In: S. Leonardi, A. Panconesi, P. Ferragina, A. Gionis (eds.) WSDM, pp. 577–586. ACM (2013)
13. Gangemi, A., Presutti, V., Reforgiato Recupero, D.: Frame-based detection of opinion holders and topics: A model and a tool. Computational Intelligence Magazine, IEEE **9**(1), 20–30 (2014)

14. Hannon, J., Bennett, M., Smyth, B.: Recommending twitter users to follow using content and collaborative filtering approaches. In: Proceedings of the fourth ACM conference on Recommender systems, RecSys '10, pp. 199–206. ACM, New York, NY, USA (2010)
15. Hong, L., Dan, O., Davison, B.D.: Predicting popular messages in Twitter. In: Proceedings of the 20th international conference companion on World wide web, WWW '11, pp. 57–58. ACM, New York, NY, USA (2011)
16. Hu, X., Tang, J., Gao, H., Liu, H.: Unsupervised sentiment analysis with emotional signals. In: Proceedings of the 22nd international conference on World Wide Web, WWW '13, pp. 607–618. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland (2013)
17. Hu, X., Tang, L., Tang, J., Liu, H.: Exploiting social relations for sentiment analysis in microblogging. In: Proceedings of the sixth ACM international conference on Web search and data mining, WSDM '13, pp. 537–546. ACM, New York, NY, USA (2013)
18. Hyman, J.: Three Fallacies about Action. Behavioral and Brain Sciences **23**, 665–666 (2000)
19. Jenders, M., Kasneci, G., Naumann, F.: Analyzing and predicting viral tweets. In: Proceedings of the 22nd international conference on World Wide Web companion, WWW '13 Companion, pp. 657–664. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland (2013)
20. Jiang, L., Yu, M., Zhou, M., Liu, X., Zhao, T.: Target-dependent twitter sentiment classification. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1, HLT '11, pp. 151–160. Association for Computational Linguistics, Stroudsburg, PA, USA (2011)
21. Kwak, H., Lee, C., Park, H., Moon, S.: What is twitter, a social network or a news media? In: Proceedings of the 19th international conference on World wide web, WWW '10, pp. 591–600. ACM, New York, NY, USA (2010)
22. Luo, Z., Osborne, M., Tang, J., Wang, T.: Who will retweet me?: finding retweeters in twitter. In: Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval, SIGIR '13, pp. 869–872. ACM, New York, NY, USA (2013)
23. Macskassy, S.A., Michelson, M.: Why do people retweet? anti-homophily wins the day! In: L.A. Adamic, R.A. Baeza-Yates, S. Counts (eds.) ICWSM. The AAAI Press (2011)
24. McPherson, M., Smith-Lovin, L., Cook, J.M.: Birds of a feather: Homophily in social networks. Annual Review of Sociology **27**(1), 415–444 (2001)
25. Moore, J., Haggard, P.: Awareness of action: Inference and prediction. Consciousness and Cognition **17**(1), 136–144 (2008)
26. Naveed, N., Gottron, T., Kunegis, J., Alhadi, A.C.: Bad news travel fast: A content-based analysis of interestingness on twitter. In: WebSci '11: Proceedings of the 3rd International Conference on Web Science (2011)
27. Naveed, N., Gottron, T., Kunegis, J., Alhadi, A.C.: Searching microblogs: coping with sparsity and document quality. In: Proceedings of the 20th ACM international conference on Information and knowledge management, CIKM '11, pp. 183–188. ACM, New York, NY, USA (2011)
28. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
29. Pennacchiotti, M., Popescu, A.M.: A Machine Learning Approach to Twitter User Classification. In: International AAAI Conference on Weblogs and Social Media (2011)
30. Petrovic, S., Osborne, M., Lavrenko, V.: Rt to win! predicting message propagation in twitter. In: ICWSM (2011)
31. Pfitzner, R., Garas, A., Schweitzer, F.: Emotional divergence influences information spreading in twitter. In: J.G. Breslin, N.B. Ellison, J.G. Shanahan, Z. Tufekci (eds.) ICWSM. The AAAI Press (2012)
32. Ramage, D., Dumais, S., Liebling, D.: Characterizing microblogs with topic models. In: Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media. AAAI (2010)
33. Řehůřek, R., Sojka, P.: Software Framework for Topic Modelling with Large Corpora. In: Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks, pp. 45–50. ELRA, Valletta, Malta (2010)
34. Starbird, K., Palen, L.: (how) will the revolution be retweeted?: information diffusion and the 2011 egyptian uprising. In: Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work, CSCW '12, pp. 7–16. ACM, New York, NY, USA (2012)
35. Stein, D., Wright, S.: Subjectivity and Subjectivisation: Linguistic Perspectives. Cambridge University Press (2005)
36. Stieglitz, S., Dang-Xuan, L.: Political communication and influence through microblogging-an empirical analysis of sentiment in twitter messages and retweet behavior. In: HICSS, pp. 3500–3509. IEEE Computer Society (2012)
37. Suh, B., Hong, L., Pirolli, P., Chi, E.H.: Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network. In: Proceedings of the IEEE Second International Conference on Social Computing (SocialCom), pp. 177–184. IEEE, Minneapolis (2010)
38. Sunstein, C.: On Rumors: How Falsehoods Spread, Why We Believe Them, What Can Be Done. Farrar, Straus and Giroux (2009)
39. Thelwall, M., Buckley, K., Paltoglou, G.: Sentiment strength detection for the social web. J. Am. Soc. Inf. Sci. Technol. **63**(1), 163–173 (2012)
40. Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., Kappas, A.: Sentiment in short strength detection informal text. J. Am. Soc. Inf. Sci. Technol. **61**(12), 2544–2558 (2010)
41. Weng, J., Lim, E.P., Jiang, J., He, Q.: Twitterrank: finding topic-sensitive influential twitterers. In: B.D.D. 0001, T. Suel, N. Craswell, B.L. 0001 (eds.) WSDM, pp. 261–270. ACM (2010)
42. Weng, J., Lim, E.P., Jiang, J., He, Q.: Twitterrank: finding topic-sensitive influential twitterers. In: Proc. of the third ACM WSDM, pp. 261–270. ACM (2010)
43. Xu, Z., Zhang, Y., Wu, Y., Yang, Q.: Modeling user posting behavior on social media. In: Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval, SIGIR '12, pp. 545–554. ACM, New York, NY, USA (2012)
44. Yang, Z., Guo, J., Cai, K., Tang, J., Li, J., 0007, L.Z., Su, Z.: Understanding retweeting behaviors in social networks. In: J. Huang, N. Koudas, G.J.F. Jones, X. Wu, K. Collins-Thompson, A. An (eds.) CIKM, pp. 1633–1636. ACM (2010)