2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing

Opinion Mining and Sentiment Analysis in Social Networks: A Retweeting Structure-aware Approach

Lu Lin, Jianxin Li, Richong Zhang, Weiren Yu and Chenggen Sun
State Key Laboratory of Software Development Environment School of Computer Science Engineering
Beihang University, Beijing, China
{linlu, lijx, zhangrc, yuweiren, suncg}@act.buaa.edu.cn

***Abstract*—Microblogs have become quick and easy online information sharing platforms with the explosive growth of online social media. Weibo, a Twitter-like microblog service in China, is characterized by timeliness and interactivity. A Weibo message carries the user's views and sentiments, particularly forms a fission-like spreading structure while being retweeted. Such structure accelerates information diffusion, and reflects different topics and opinions as well. However, current researches mainly focus on sentiment classification, which neither efficiently combine tree-like retweeting structure nor analyze opinion evolutions with a holistic view. In light of this, we build an opinion descriptive model, and propose an opinion mining method based on this model. With a microblog-oriented sentiment lexicon being constructed, a lexicon-based sentiment orientation analysis algorithm is designed to classify sentiments. Finally, we design and implement a prototype which can mine opinions with respect to retweeting tree structures and retweeting comments.**

Keywords*—*opinion mining, sentiment analysis, opinion summarization, Weibo

I. Introduction

Recent years have witnessed the explosive development of online social media. Microblogs, which fuse information release and social networks, become one of the most popular information sharing platforms with high interactivity. Millions of messages are posted daily in microblogging web-site such as Twitter, Facebook and Tumblr. Weibo, a Twitter-like service launched in 2009, has attracted over 500 million users in less than five years, with over 100 million Chinese tweets being posted every day.

Information propagates through social networks, which is typically supported by the three features of Microblogs: short and simple contents, interactive comments and particularly an easy accessibility of retweets. More and more users of microblogs tend to share trifles and big deals, express opinions and sentiments on current issues and discuss various topics. Therefore, the huge amount of tweets provides us with rich informations about authors and real-world events, in which opinions and sentiments are essential. Such data can be efficiently used for marketing, public opinion analysis, social study and user behavior analysis.

However, researches mainly focus on sentiment classification of microblogs, which only demonstrates the sentiment orientation of one tweet or the sentiment fluctuation of tweets in a time interval. They do not reveal the sentiment propagation process and uncover the hot opinions behind it. Few works have been done to efficiently combine sentiment propagation and opinion evolution analysis with retweeting structures.

In light of this, we design and implement a real-time prototype to perform opinion mining for tweets, particularly concerning the retweeting structure. Given a hot tweet and an a set of retweeting comments, which are comments generated while the tweet is retweeted, the prototype can analyze sentiment propagation in retweeting structures and tell the hot opinions formed in retweeting procedure. Our task is performed in three main steps:

Firstly, building a descriptive model to depict opinions. A Weibo opinion mining method is proposed base on the model. This method mines opinion targets with the association rule mining algorithm, then identifies opinion words using syntactic relations, and classifies sentiment orientation of the sentence with lexicon-based method and summarizes all the opinion triples.

Secondly, constructing a microblog-oriented sentiment lexicon using the Semantic Orientation from Pointwise Mutual Information (or SO-PMI for short). Then a lexicon-based sentiment analysis algorithm is proposed to calculate the sentiment score of tweets effectively.

Finally, building an opinion mining prototype for tweets using the above processing modules. Functions of the prototype includes: (1) Analyzing sentiments of retweeting comments along a retweeting tree, and through a retweeting sentiment graph, showing the sentiment turning point while the tweet is retweeted. (2) Calculating numbers of positive and negative retweeting comments in a time internal and through a sentiment fluctuation line chart, presenting how sentiments aroused by a tweet change with time. (3) Mining opinions in retweeting comments, and through an opinion summarization chart, telling the hot opinions generated in retweeting procedure.

The main contributions of our paper are as follows:

- Proposing an efficient method to mine and summarize opinions using association rule mining algorithm.
- Building a microblog-oriented sentiment lexicon based on real tweets corpus to deal with the specificity of microblogs such as emoticons and online language. We then propose an effective algorithm to calculate sentiment orientations.
- Building a real-time system to perform opinion mining for tweets, which observes interesting or abnormal phenomena. Experiments on a set of real-world Weibo tweets show that our method is efficient and performs well.

978-1-4799-7881-6/14 $31.00 © 2014 IEEE
DOI

II. RELATED WORKS

Opinions and sentiments are the subjects of study of opinion mining and sentiment analysis, which first appear in [1]. While the exponential development of the research field coincide with that of the social media, such as reviews, blogs, microblogs and social networks, opinion mining and sentiment analysis have grown to be a field of interest for many researches.

A broad overview of existing works is presented in [2]. This survey covers existing techniques and approaches for an opinion-oriented information retrieval. An in-depth introduction to opinion mining is given in [3]. The author defines the problem and provides an abstraction to the problem. Moreover, the author decomposes the problem into key sub-problems and presents a comprehensive survey of techniques for solving these sub-problems. However, not many researches consider blogs and even much less take microblogs into account.

We mostly follow but simplify the abstraction of opinion mining in [3], because this abstraction gives us a common framework to unify research directions, and enables us to decompose the complex problem into inter-related sub-problems. Opinion mining is firstly studied in product reviews, and the opinion is defined as a quintuple (entity, feature, sentiment, opinion holder, opinion time).

Our work is related to but different from the work in [4] which studies the problem of generating feature-based summaries of customer reviews. The author defines a feature as an aspect of product attributes or functions. The task is performed in three sub-tasks: mining product features, classifying sentiment orientations and summarizing the results. However, their design is for product reviews and is not suitable for microblogs. In microblogs, we use ”opinion target” to represent entity and feature, because the differences between them are not obvious.

A. Opinion Target Extraction

An opinion always has a target which is the topic discussed in a sentence. We consider opinion targets to be nouns and noun phrases.

The first method is based on frequencies of nouns and noun phrases. A data mining algorithm is used in [4] to extract entities based on the observation that when people comment on different aspects of an entity, the vocabulary that they use usually converge. The authors count the occurrence frequencies of nouns and only keep the frequent ones. This method is simple yet effective. A frequent-based method is also used in [5]. The authors calculate TF-IDF scores and assign it as weight to each term both at the document level and at the paragraph level.

The second method is based on relations between sentiment words and opinion targets because sentiment words are often known. In [6], the authors use several syntactic relations that link opinion words to expand the initial opinion lexicon and to extract targets, use a dependency parser to identify these relations. This relation-based method is useful for extracting key targets because a target is unlikely important if no one express sentiments on it. Moreover, it can simultaneously extracting both sentiment words and opinion targets.

The third method is based on supervised learning because it is essentially an information extraction problem. Conditional Random Fields (CRF) is used in [7] to learn patterns and to extract targets. The authors use a set of domain independent features such as POS tags and syntactic dependency to train CRF on reviews from multiple domains. Other learning methods can also be used.

B. Sentiment Orientation Classification

At present, sentence-level sentiment classification work is mainly divided into the method based on sentiment lexicon and the approach based on machine learning.

For the machine learning methods, sentiment classification is essentially a text classification problem, therefore many learning methods can be applied, e.g., Naïve Bayes classification and support vector machines (SVM). The key is to engineer a set of effective features, such as terms and their frequency, part of speech, sentiment words, etc.

The features of bag-of-word and appraisal group such as ”very good” are used in [8]. The authors classify movie reviews using these features and report accuracy of 90.2%. In [9], the authors train SVM and conditional random field (CRF) classifiers at a sentence level. They find the importance of the last sentence in a document while applying these classifiers to the document level. In [10], the authors present a system for real-time twitter sentiment analysis of the 2012 U.S. presidential election. A system for sentiment analysis of Chinese tweets is built in [11], the authors train a fast naïve Bayes classifier to make the system capable of online real-time sentiment monitoring.

For the lexicon-based method, the key is to compile a lexicon that has an obvious tendency and a broad coverage. The main approach is based on dictionary which use a few seed sentiment words to bootstrap based on the synonym and antonym structure of an online dictionary such as WordNet.

Pointwise mutual information (PMI) is used to compute the sentiment orientation of a given word in [12], the authors calculate the semantic orientation as the mutual information between the given phrase and the word ”excellent” minus the mutual information between the given phrase and the word ”poor”. Especially in [13], the association strength is measured using a set of positive words and a set of negative words other than one word such as ”excellent”. In [14], the authors use a WordNet distance based method to determine the sentiment orientation of a given adjective. The distance between two terms is the shortest path that connects the terms in WordNet. In [15], the authors use a positive seed set, a negative seed set and a neutral seed set to score each word based on a directed, weighted semantic graph in which neighboring nodes are synonyms or antonyms of words in WordNet but not part of the neutral deed set. Other methods based on word relatedness graphs includes those in [16] [17]. The dictionary-based approach can be easily performed but has the disadvantage that the generating sentiment words are domain independent, because a word can be positive in one context but negative in another.

III. Opinion Mining Methods

As mentioned in Section 2, the aim is to analyze all the comments generated while a hot tweet is retweeted. By mining and analyzing the opinions contained in those comments, we can discover some temporal or spatial patterns, e.g. sentiment propagation features, sentiment orientation turning points and sentiment fluctuation peaks. Problem definition is firstly needed.

Definition 1: An opinion in an opinion tweet document is a triple,

$$(g_i, w_{ij}, s_i),$$

where g_i is an opinion target, w_ij is an opinion word commenting on the opinion target g_i, and s_i is the sentiment orientation of the document about the opinion target g_i.

The definition is simple yet effective for reflecting the opinions generated in microblogs. With the definition, we now have the task of opinion mining.

Definition 2: Given an opinion document d, the task of opinion mining is to mine and summarize all the opinion triples (g_i, w_{ij}, s_i) in d.

It is quite obvious that the task of mining opinions in corpus D consists of four main sub-tasks: (1) opinion target extraction, which extracts entity nouns or noun phrases in D. (2) opinion words extraction, which extracts opinion words for every target. (3) sentiment orientation classification, which decides the sentiment tendency of every document on every target. (4) opinion summarization, which summarizes all the opinion triples.

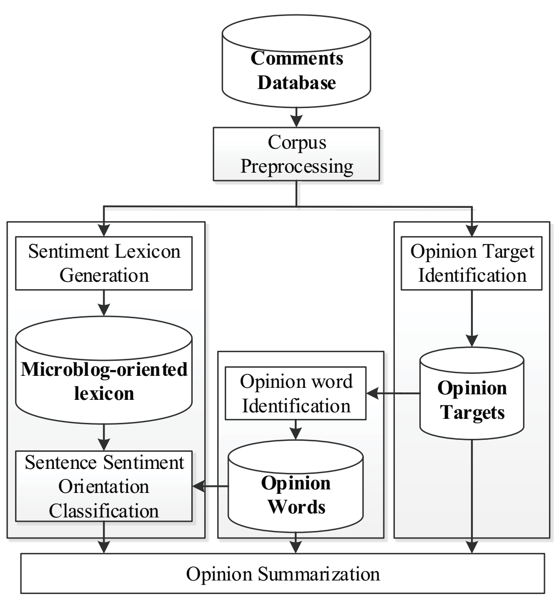


Fig. 1: Opinion mining framework

An architecture of our opinion mining framework is shown in Fig.1. Given a corpus of microblog comments, the framework will first do some preprocessing such as Chinese segmentation and POS tagging. It then generates a microblog-oriented sentiment lexicon using SO-PMI. At the same time, the system extracts the opinion targets that users have talked about, and sorts the result by the frequencies of targets appearing in corpus. After that, using those targets and semantic hints, the system extract the opinion words commented on every target. The sentiment orientation of each target in each document is decided based on the sentiment lexicon. In the last step, a summary of all the opinion triples is generated. The key subtasks is discussed below.

A. Corpus Preprocessing

A tweet contains many redundant information, thus a corpus preprocessing is needed.

Useless stop words are removed first, e.g. "http://", "@", "#", etc. NLPIR [18] is used, which is a Chinese segmentation tool, to cut Chinese tweets into single words or phrases. The tool can also produce part-of-speech tag for each word or phrase, which decides whether the word or phrase is a noun, adjective, etc.

The preprocessed corpus will be used in two ways in the next steps. The first one is for opinion target identification which mainly uses nouns. We create a transaction file in which each record stands for one tweet and contains the nouns and noun phrases of this tweet. The second one is for sentiment lexicon generation and sentiment words extraction which mainly uses all the adjectives, adverbs and verbs in corpus.

B. Opinion Target Extraction

The system aims to discover what people are talking about, so find the opinion targets is a crucial task. At the same time, the targets are sorted because we are interesting about what targets are the discussion hotspot.

As discussed in Section 2, there are many methods to extract Opinion Targets. Base on the observation [4] that when people comment on opinion targets, the words they use converge, we use association mining to find the hot opinion targets. The transaction database is the transaction file generated in the corpus preprocessing step, and an item is a noun or a noun phrases.

Definition 3: Each record r is the nouns or noun phrases in a tweet document, all records consist of a tweet transaction database,

$$r = \{noun_1, noun_2, \ldots, noun_n\},$$

$$R = \{r_1, r_2, \ldots, r_m\}.$$

Apriori algorithm in [19] is used to generate 1 to k frequent itemsets which are candidate opinion targets. We define an itemset as frequent if its minimum support is 5%. Because microblogs posts are usually very short, the value of k is set as 5 to simplify the calculation.

We now generate and sort opinion targets from those frequent itemsets. Comments on a hot tweet usually contains things that appeared in the original tweet. Moreover, comments also talk about some other related things, which makes the tweet much more meaningful. Thus if a comment contains more nouns or noun phrases, the comment will have higher possibility to be a subjective opinion sentence and those nouns or noun phrases may be opinion targets.

Based on that observation, the nouns or noun phrases in k-frequent itemsets have higher possibility to be opinion targets compared with those in 1-frequent itemsets. By traversing the itemsets from k-frequent to 1-frequent, noun or noun phrase are put into opinion targets database one by one. If the noun or noun phrase exists in the database, then skip it. Then the opinion targets are extracted after finishing traversing 1-frequent itemsets. the value of k is adjustable according to the real situation.

Algorithm 1 Tweet sentiment orientation calculation

```
input: a tweet; a sentiment lexicon; a modifier lexicon
output: sentiment orientation of the input tweet
function TWEETSO(tweet)
    so ← 0
    weight ← 0
    orientation ← NEUTRAL
    for each sentence s_i ∈ tweet do
        if s_i ∈ [1, 9] then
            weight ← 2
        end if
        if s_i ∈ [10, 15] then
            weight ← −2
        end if
        so ← so + weight× SENTENCESO(s_i)
    end for
    if so > 0 then
        orientation ← POSITIVE
    end if
    if so < 0 then
        orientation ← NEGATIVE
    end if
    return orientation
end function
function SENTENCESO(sentence, Lexicon)
    so ← 0
    for each word w_i ∈ sentence do
        degree ← 1
        negation ← 1
        for each degree adverb d_j ahead of w_i do
            degree ← degree × Lexicon[d_j]
        end for
        for each negation word n_j ahead of w_i do
            negation ← negation × Lexicon[n_j]
        end for
        so ← so + negation × degree × Lexicon[w_i]
    end for
end function
```

C. Opinion Words Extraction

We are interested in how people think of a target, that is, what words they use to comment a target. Opinion words which express subjective opinions are crucial. Words showing opinions mainly adjectives, adverbs and verbs.

Based on opinion targets generated in the last step, opinion words are extracted, therefore only those tweets that contain those opinion targets are taken into account. For each tweets, if it contains an opinion target, we extract the adjectives/adverbs/verbs near the target. We say, if an adjective is adjacent to the target, that is, if the distance between them is no more than a threshold, the adjective modifies the targets. The distance is measure by the number of words between them in a sentence. Complex syntactic or semantic structures are not considered here for efficiency, but it is still an improvement work in our future work.

D. Sentiment Orientation Classification

After extracting opinion targets and opinion words, the sentiment orientation of a document toward every targets are decided. The sentiment orientation of a word/sentence/document indicates the direction that it deviates from the sentiment neutrality.

Sentiment words, which are the basic and concentrated reflections of sentiment features, are words that have obvious sentiment tendency, including adjectives, nouns, adverbs, etc. A sentiment lexicon is a set of sentiment words, which can be divided into positive sentiment lexicon and negative sentiment lexicon. There are some basic sentiment lexicon for Chinese, e.g. HowNet [20]. However, they are not sufficient for microblogs because expressions in microblogs are newfangled. That is why a microblog-oriented sentiment lexicon is constructed. The collection of sentiment words is an accumulated procedure, the methods of which are discussed in Section 2.

In this research, a microblog-oriented sentiment lexicon is built using SO-PMI based on real tweets corpus to deal with the specificity of microblogs such as emoticons and online language. With a set of seed adjectives, the orientations of which are manually labeled, SO-PMI between seed adjectives and candidate adjectives are computed to grow the set.

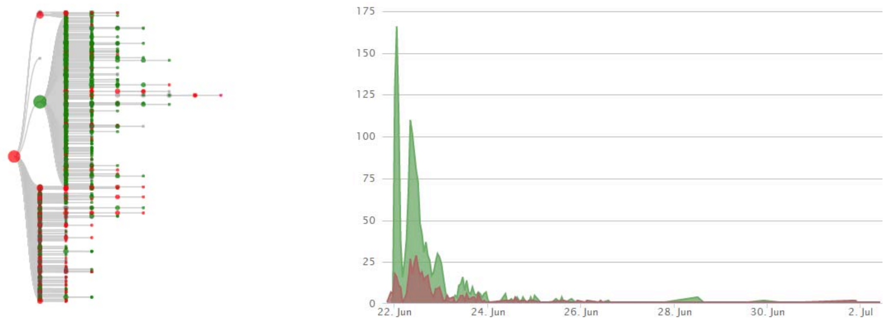
SO-PMI originates from PMI which is a measure of association in information theory. Suppose the probability of $word_1$ appearing in documents is $p(word_1)$, the probability of $word_2$ appearing in documents is $p(word_2)$, and the coincidence probability of $word_1$ and $word_2$ is $p(word_1, word_2)$, then we have:

$$PMI(word_1, word_2) = \log \frac{p(word_1, word_2)}{p(word_1)p(word_2)} \tag{1}$$

It can take positive or negative values, which indicates the two words related or exclusive, and is zero if they are independent. It can be used to measure the semantic similarity between words.

Suppose there are a set of positive words $PosWords$ and a set of negative words $NegWords$. For a candidate word w, we have:

$$\begin{aligned} SO-PMI(w) = &\sum_{posw \in PosWords} PMI(word, posw) \\ &- \sum_{negw \in NegWords} PMI(word, negw) \end{aligned} \tag{2}$$



(a) Sentiment propagation graph. The root node is the original tweet, while others are retweeting comments.

(b) Sentiment fluctuation chart. The horizontal axis shows the time line, and the vertical axis shows the number of retweeting comments with a certain sentiment orientation.

Fig. 2: A sample of real-world monitoring. The red color denotes negative sentiment, while green denotes positive and grey denotes neutral. Obvious sentiment propagation turning points and sentiment fluctuation peaks can be observed.

TABLE I: Parts of Opinion Summary Mined by Our System

Opinion targets (top 4)	Opinion words (top 4)	Sentiment contrast
能量(energy)	传播(propagate), 转发(retweet), 传递(pass on), 赌赢(win bets)	sum:1517
		positive:1400
		negative:117
赌球(soccer gambling)	小心(careful), 买(buy), 远离(keep away), 毁(ruin)	sum:119
		positive:75
		negative:44
人生(life)	押(bet), 赌(gamble), 精彩(splendid), 放心(relieved)	sum:43
		positive:12
		negative:31
世界杯(world cup)	悲惨(miserable), 赌(gamble), 悲伤(sad), 坑(deceptive)	sum:34
		positive:22
		negative:12

If SO-PMI takes a positive value, w is more related to positive words, so w has positive sentiment orientation.

The adjectives generated in corpus preprocessing step are firstly sorted by their frequencies, from which we manually choose a set of adjectives that have obvious orientations (we choose 40 adjectives for positivity and negativity respectively) as the seed set. The rest adjectives consist of the candidate set. For each candidate word, SO-PMI is computed to decide the word being classified to positive lexicon or negative lexicon. Moreover, microblogs provide various emoticons to users to express their mood, hence emoticons can be viewed as special sentiment words.

With the microblog-oriented sentiment lexicon, an effective algorithm is proposed to calculate sentiment orientations. The detail is showed in Algorithm 1.

Note that in the procedure TWEETSO($tweet$), we consider whether the sentence is exclamatory or rhetorical, because such feature will strengthen or invert the whole sentiment tendency. Likewise, modifiers ahead of opinion words such as privatives or degree adverbs are taken into account in the procedure SENTENCESO($sentence, Lexicon$). $Lexicon$ stores the sentiment lexicon and modifier lexicon in a array-like data structure.

IV. Implementation

Based on the opinion mining framework designed in Section 3, we build a real-time prototype. Through the visual analysis results, some interesting phenomena are observed, e.g. special sentiment turning points, sentiment groups and sentiment peaks. Opinion summary may tell the hot opinions. We now use a real-world sample and an experiment to prove the effectiveness of our system.

A. *Real-time Opinion Mining Sample*

Given an access to a hot tweet and related retweeting comments, the system will generate several intuitive charts to show how sentiment propagate and fluctuate, and what opinions are hot while the tweet is retweeted. The analysis results of a tweet about the 2014 FIFA World Cup is shown in Fig. 2 and TABLE I.

As can be seen, interesting results can be indeed mined. The original tweet talked about soccer gambling during the World Cup. The author was upset for losing bets, and made a self-mockery in this tweet, because as we know, there were many startling situations in this Wold Cup, e.g. many top teams being washed out at the early stage, with surprising scores. So the sentiment was negative before the tweet was retweeted. However, sentiments became positive, and formed an obvious group, as a turning point appeared, and positive sentiments peaked in a short time. Reasons behind the observation are found from the opinion summary. It turns out that one people said something about not gambling and stroke a chord. From the mining results, we can estimate some deep informations such as the proportion of soccer gambling users in Weibo.

B. Experiment Evaluation

An experiment is designed to evaluate our sentiment orientation classification method. We manually read and label the sentiments of two thousand tweets, and conduct a sentiment classification using two methods: lexicon-based method and SVM-based method. The results are shown in TABLE II. The results indicate that in the context of microblogs, lexicon-based method achieves much higher precision and recall, which shows the effectivity of our system.

TABLE II: Results of sentiment orientation classification

	Criterion	Positive sentiment	Negative sentiment	Whole
SVM-based method	p	70.26%	48.39%	63.52%
	r	75.31%	42.06%	–
	$F1-Measure$	72.20%	45.00%	–
Lexicon-based method	p	80.03%	61.61%	72.90%
	r	84.66%	77.82%	–
	$F1-Measure$	82.28%	68.77%	–

V. Conclusions And Future Works

In this research, we propose a framework to mine opinions in microblog domain and then build a real-time analysis system to monitor the sentiment propagation, sentiment fluctuation and hot opinions in Weibo. Opinions are abstracted as triples, based on which we then design an opinion mining method. We build a microblog-oriented sentiment lexicon and propose a lexicon-based sentiment analysis algorithm to classify sentiments. The experiment shows the effective of our system.

In future, we plan to refine our techniques by combining the semantic analysis. Moreover, the efficiency of our system can be improved when analyzing large-scale data, by distributed data processing. For further research, we will consider the problem of sentiment propagation on retweeeting tree as an influence propagation problem. By analyzing plenty of retweeting data, we believe some special patterns in retweeting graph will be discovered, such as adversarial followers which means that someone keeps attack another in retweeting comments.

VI. Acknowledgment

This work is supported by China MOST project (No. 2012BAH46B04), NSFC Program (No. 61472022), and SKLSDE-2014ZX-04.

References

[1] K. Dave, S. Lawrence, and D. M. Pennock, "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews," in *Proceedings of the 12th International Conference on World Wide Web*. New York, NY, USA: ACM, 2003, pp. 519–528.

[2] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Retr.*, vol. 2, no. 1-2, pp. 1–135, Jan. 2008.

[3] B. Liu, "Sentiment analysis and opinion mining," *Synthesis Lectures on Human Language Technologies*, pp. 1–167, 2012.

[4] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: ACM, 2004, pp. 168–177.

[5] L. Ku, Y. Liang, and H. Chen, "Opinion extraction, summarization and tracking in news and blog corpora," in *Proceedings ofAAAI-CAAW-06, the Spring Symposia on Computational Approaches to Analyzing Weblogs*, 2006.

[6] G. Qiu, B. Liu, J. Bu, and C. Chen, "Opinion word expansion and target extraction through double propagation," *Comput. Linguist.*, vol. 37, no. 1, pp. 9–27, Mar. 2011.

[7] N. Jakob and I. Gurevych, "Extracting opinion targets in a single- and cross-domain setting with conditional random fields," in *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2010, pp. 1035–1045.

[8] C. Whitelaw, N. Garg, and S. Argamon, "Using appraisal groups for sentiment analysis," in *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*. New York, NY, USA: ACM, 2005, pp. 625–631.

[9] C. Yang, K. H.-Y. Lin, and H.-H. Chen, "Emotion classification using web blog corpora," in *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*. Washington, DC, USA: IEEE Computer Society, 2007, pp. 275–278.

[10] H. Wang, D. Can, A. Kazemzadeh, F. Bar, and S. Narayanan, "A system for real-time twitter sentiment analysis of 2012 u.s. presidential election cycle," in *Proceedings of the ACL 2012 System Demonstrations*. Association for Computational Linguistics, 2012, pp. 115–120.

[11] J. Zhao, L. Dong, J. Wu, and K. Xu, "Moodlens: An emoticon-based sentiment analysis system for chinese tweets," in *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2012, pp. 1528–1531.

[12] P. D. Turney, "Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews," in *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2002, pp. 417–424.

[13] P. D. Turney and M. L. Littman, "Measuring praise and criticism: Inference of semantic orientation from association," *ACM Trans. Inf. Syst.*, vol. 21, no. 4, pp. 315–346, Oct. 2003.

[14] J. Kamps, M. Marx, R. J. Mokken, and M. D. Rijke, "Using wordnet to measure semantic orientation of adjectives," in *National Institute for*, 2004, pp. 1115–1118.

[15] S. Blair-goldensohn, T. Neylon, K. Hannan, G. A. Reis, R. Mcdonald, and J. Reynar, "Building a sentiment summarizer for local service reviews," in *In NLP in the Information Explosion Era*, 2008.

[16] A. Hassan and D. Radev, "Identifying text polarity using random walks," in *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2010, pp. 395–403.

[17] L. Velikovich, S. Blair-Goldensohn, K. Hannan, and R. McDonald, "The viability of web-derived polarity lexicons," in *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2010, pp. 777–785.

[18] K. Zhang, "NLPIR." [Online]. Available: http://ictclas.nlpir.org/

[19] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules in large databases," in *Proceedings of the 20th International Conference on Very Large Data Bases*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1994, pp. 487–499.

[20] Z. Zhang, "HowNet." [Online]. Available: http://www.keenage.com/