Journal of Computational Information Systems 9: 6 (2013) 9–17
Available at http://www.Jofcis.com

Mining Aspects and Opinions from Microblog Events

Xiao ZHOU*, Fang LI

Department of Computer Science and Engineering, Shanghai Jiaotong University, Shanghai 200240, China

Abstract

Microblogging platforms have experienced great growth during the last few years. Opinions are valuable for many applications. Most researches concentrate on classifying microblog posts into different polarities, which cannot provide concise opinion information. This paper presents an opinion mining system that extracts aspect related opinions for some specific event from Microblogs. Our approach first extracts aspect words, then cluster them as aspects of the opinions. Opinions are extracted by a CRF model that trained from automatically labeled data. Experimental results verify the effectiveness of the proposed system.

Keywords: Microblog; Opinion Mining; Information Extraction

1 Introduction

Microblogging is becoming a popular way of sharing information and personal feelings. Studies [1, 2] have shown that a substantial portion of microblog posts are about daily activities and events, which makes it a great resource for mining opinions about events. Most of current researches on opinion mining of microblogs [3, 4, 5] are about polarity classification, which doesn't always provide enough information. For example, it's easy to anticipate that majority of people will express negative feelings on a scandal event. In this paper, we provide an opinion mining system that will automatically extract popular aspects of the event, and their corresponding opinion expressions, which will be simply called opinions afterwards. The task is performed in three steps:

- **Mining aspects of events.** Nouns are extracted and ranked based on their frequencies in the event related microblogs and the inverted document frequencies in a background corpus. Top nouns are selected as aspect words, which will be clustered based on their relations. The result of clustering is the aspects of an event.
- **Identifying opinions in the microblogs.** A CRF model is trained to identify opinions in the event related microblogs. Emoticons and special features are also used.

*Corresponding author.
Email address: ucai@sjtu.edu.cn (Xiao ZHOU).

March 15 2013

- **Aspect-opinions generation.** Opinions and related aspects are connected based on their adjacency. For each aspect, opinions will be ranked based on their related degree.

The rest of this paper is organized as follows: next section gives an overview of related work in opinion analysis of microblogs and aspect-level opinion mining. Section 3 presents detailed information of the proposed system. In section 4, several experimental results are shown. Section 5 is the conclusion.

2 Related Work

Bo Pang and Lillian Lee [6] presented a comprehensive survey about opinion mining. This section focuses on two closely related fields: opinion mining of microblogs and aspect-level opinion mining.

2.1 Opinion mining of microblogs

Luciano Barbosa and Junlan Feng [3] leverage existing websites to get sentimental training data. They design a 2-step classification method which first labels Twitter posts as subjective or objective, and then classifies the subjective posts as positive or negative. Various Twitter syntax features are extracted to represent the Tweet post. Davidov et al. [4] use hashtags and smileys as a sentimental signal. Those microblogs containing hashtags and smileys are used as training data. A work similar to ours is presented in [5]. They address the sentimental classification problem in a target-dependent way. Syntactic information is used to generate a large set of features for opinion target identification. A graph-based sentiment optimization algorithm is utilized to readjust the classification result. Chen et al. [7] propose a method to find sentimental clues in microblogs. Slangs that contain opinions are selected to extend the sentiment dictionary. Phrases that contain at least one word in the sentiment dictionary are regarded as candidates. An optimization algorithm is used to generate the final result from those candidates.Our method differs from these works in that we focus on aspect-level opinion mining, and the result is popular opinions about specific aspects rather than just opinion polarities.

2.2 Aspect-level Opinion Mining

Aspect-level opinion mining conveys more information, it summarizes opinions based on opinion targets. Minqing Hu and Bing Liu [8] propose a system to score features of products based on online reviews. They use a statistical approach to extract high frequency nouns and noun phrases as candidates of opinion targets. Adjectives are simply treated as opinions, which will be assigned to the nearest opinion target. The final score of a feature is calculated based on the sentiment orientations of the opinion words attached to the feature. Ana-Maria Popescu and Oren Etzioni [9] also use high frequency nouns as opinion targets, but a set of part discriminators is utilized to further filter the candidates. Their method trades a small percentage of recall for a large percentage of precision comparing with the method of [8]. The biggest difference between our system and previous aspect-level opinion mining systems is that we perform this task on chinese microblogs. Informal texts have strong impact on the mining effect [10]. This paper describes our method to solve this problem.

3 Proposed System

This section will present details of the proposed system. A corpus of microblogs about a specific event is given as input, then the following four steps will be conducted successively. The output are event aspects and their related popular opinions.

3.1 Preprocessing

Several strategies are used to do the filtering for an event-based microblog corpus:

- Regular expression patterns: a list of regular expression patterns are designed to identify spams and objective microblogs. For example, a pattern "^分享了一个投票.+" (share a vote) will identify those microblogs automatically generated when microbloggers participate in some votes.
- Microbloggers' credibility: Microbloggers with less than 5 posts or less than 3 followers are not considered. Their microblogs are filtered.
- Number of trending topics: If the number of trending topics (i.e. hashtags followed by successive words) in a microblog post exceeds 2, it is considered as a potential spam, hence is pruned.

After filtering, URLs and hashtags are removed. Consequent question marks, exclamatory marks, periods, and commas are replaced with placeholders, which will be utilized in the following steps. Finally, microblogs about an event are processed by a Chinese NLP toolkit[1].

3.2 Event aspect extraction

Two definitions are in the following:

Definition 1 *An* **aspect word** *is a target that people express their opinions on. It may be an object, a person, or an important concept.*

Definition 2 *An* **event aspect** *is a discussion theme of the event. It is represented as a cluster of aspect words.*

3.2.1 Aspect words extraction

According to the traditional TF-IDF, nouns (as aspect words) are extracted and ranked based on the following equation:

$$score(n) = freq(n) * \frac{1}{\log(freq_{bg}(n) + 1)} \tag{1}$$

Where n means a noun, $freq(n)$ is the number of times n occurred in event based microblogs, and $freq_{bg}(n)$ is the number of times n occurred in the background corpus. The background corpus is collected by monitoring the public timeline of Sina Weibo. It consists of general microblogs, not focusing on specific events. At last, the top 100 nouns are selected as aspect words.

[1]ICTCLAS: http://ictclas.org/

3.2.2 Aspect words clustering

A Chinese synonym dictionary[2] is leveraged to put synonymous aspect words into the same set at first. Then our system runs a Complete-Link Clustering algorithm on the synonym sets. Distance between two synonym sets is computed as follows:

$$dist(s_1, s_2) = 0.5 - \frac{NormalizedPMI(C_{s_1}, C_{s_2})}{2} \tag{2}$$

where s_1 and s_2 are two synonym sets, C_{s_1} and C_{s_2} are two centroids of the synonym sets. Centroid of a synonym set is the most frequent word in this set. $NormalizedPMI(C_{s_1}, C_{s_2})$ is the normalized Pointwise Mutual Information of C_{s_1} and C_{s_2}. Its value is between -1 and 1. After clustering, every aspect has a weight which equals to the sum of the occurrences of the aspect words contained in the aspect. The top 25 aspects are chosen as the hot aspects for each event.

3.3 Opinions identification

A straightforward way to identify opinions is using a large and precise sentiment dictionary. However, it will generally result in poor precision due to lack of consideration of context. To solve this problem, we use a CRF model to identify opinions. It consists of the following steps:

(1) Training data generation.Inspired by [4], our system selects those posts which contain at least one emoticon and one aspect word as training data. Word in the training data is automatically labeled as “Yes” if it is included in the sentiment dictionary; otherwise it is labeled as “No”. Two POS patterns proposed in [11] are used to extract two words opinion expressions, which means if an adjective or adverb is followed by an adjective, then these two words will be extracted as one opinion.

(2) CRF model training. Five sets of features are used by the CRF model:

- Token: the current word.
- POS tag.
- Successive punctuations: Some microblogs may contain successive punctuations such as ?? or !!!, which often indicate strong feelings.
- Microblog specific features: Appearances of hashtag and @ are used as binary features.
- Occurrence of aspect words: Number of times aspect words occur in this microblog.

(3) Identifying opinions. The trained classifier is used to identify the rest of the microblogs. The result of CRF is further checked by the sentiment dictionary. Only those opinions which are contained in the sentiment dictionary are the final result.

3.4 Aspect-opinions generation

Opinions are matched with the nearest aspect words in a window. Since aspect words are more likely to appear before the opinions, the system first searches words before the opinion, and if

[2]《同义词词林》扩展版:http://ir.hit.edu.cn/

no aspect word is detected, then it continues to search words after it. If a non-aspect noun is detected, or a punctuation that indicates the end of a clause (such as a period or a semicolon) is detected, the process will end immediately. Only if the aspect and opinion are both in the window, they are regarded as a related pair. For each aspect, the related opinions are ranked based on their relevance to the aspect. The relevance score is computed as follow:

$$relevance(e, a) = occur(e, a) * PMI(e, a) \tag{3}$$

where e is an opinion, a is an aspect. $occur(e, a)$ is the number of times they are within a window. $PMI(e, a)$ is the Pointwise Mutual Information of e and a which equals to:

$$PMI(e, a) = \log\left(\frac{occur(e, a)}{occur(e) * occur(a)}\right) \tag{4}$$

$occur(e)$ is the number of times e as an opinion in the event microblogs, $occur(a)$ is the number of times a as an aspect word. Finally, top 10 opinions are selected for each aspect of the event.

4 Experiments

4.1 Experiment setups

Event-based microblogs are collected by submitting key words to Sina Weibo API. Four events are chosen for experiments: "The Flowers of War", which is a movie showed in 2011; "Poison Milk of Mengniu" is a scandal of China's milk industry; "Event of Yue Yue" is a tragedy happened in Guangdong, China; "Death of Gaddafi" is an important political event happened in 2011. Over 1.3 million microblog posts are collected totally.

To build the background corpus for aspect extraction, over 5 million microblogs are collected from Sina Weibo in a week. The split threshold of clustering algorithm is set to 0.1, which means the distance of any two synonym sets in a cluster can't exceed 0.1. The window size for aspect-opinions generation is set to 5.

4.2 Result Empirical Analysis

Due to limitations of space, we only show part of the translated result of event "The Flowers of War" in table 1. It is a movie directed by a famous Chinese director Yimou Zhang. It is about 13 prostitutes sacrificed themselves to protect other citizens during the Nanjing Massacre. Table 1 lists top 10 aspects, and the top 5 opinions for each aspect.

The result has captured most of the interesting aspects and opinions of this event. The majority of the aspect words are correct, such as "Japanese, devil" and "woman, female, maiden", which are important elements of the movie. Some other aspects are also extracted such as director, movie theaters, and other movie that shows in the same time period. Opinions of each aspect also indicate informative information. On one hand, many audiences show great depression about this historic event, and strong condemnation of the invalders; On the other hand, the movie and its director are highly praised.

Table 1: Generated Summary of "The Flowers of War"

Aspect	Opinions
Japanese, devil	abhor, hate, hatred, inhuman, sick
movie	inexpensive, not bad, best, excellent, leave
Nanjing, Japanese army	judge, rape, disaster, sound, familiar
Yimou Zhang	best, new, war, grateful, deserve
movie theater, cinema	worthy, haven't been there, bright, elevation, support
enemy, NanKing	doom, interesting, just now, preparing, very good
film, picture	not bad, peak, heavy, best, hot
Long Men	interesting, not bad, like, close enemy, excellent
China	great, powerful, patriotic, participate, shame
woman, maiden, female	great, beautiful, escape, protect, beautiful

4.3 Quantitative analysis of aspect extraction

4.3.1 Aspect words evaluation

Human judge evaluates the precision of extracted aspect words. The evaluators manually check the results: Qualified, if it is correct; Incomplete, if the result is part of the correct answer; Fail, if it is not a correct aspect. For example in table 1, the word "Long Men" makes no sense. It is a part of the correct answer "Long Men Fei Jia" which is a movie showing at the same time with the film "The Flowers of War". Therefore "Long Men" is judged as incomplete. Table 2 shows the result of four events. For each event, top 25 aspects are checked. The experimental result verifies that the frequency-based extraction method is reasonable.

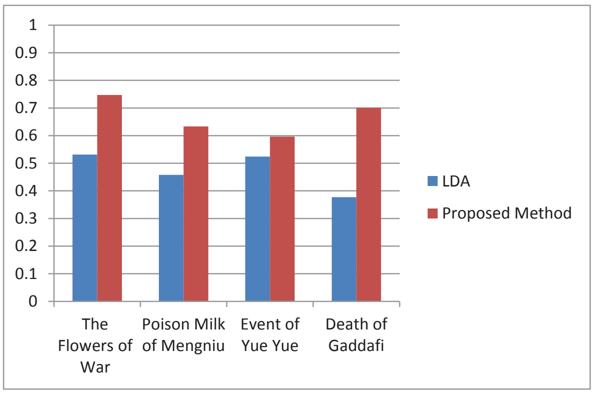

Fig. 1: Accuracy of aspect words clustering test

Table 2: Human judge result of aspect words extraction

Events name	Qualified	Incomplete	Fail
The Flowers of War	86.84%	7.89%	5.26%
Poison milk of Mengniu	80.39%	3.92%	15.69%
Event of Yue Yue	77.97%	6.78%	15.25%
Death of Gaddafi	81.67%	8.33%	10.00%
Total	81.25%	6.73%	12.02%

4.3.2 Aspect words clustering evaluation

In order to evaluate the clustering result, a method inspired by [12] is used. This method evaluates the consistency of the words in a cluster. For each evaluator, three aspect words are randomly selected for evaluation. Two of them come from the same aspect, while the third one comes from a different aspect. The evaluator is asked to pick out the third word from the different aspect. For example, there are three words "female", "maiden" and "movie", "movie" should be the right answer in this case. Methods based on topic model are used to find event aspects [13, 14]. A

LDA model using the Stanford Topic Model Toolbox[3] is used as the baseline method (similar to [14]). The topic number is set to 10 for each event and top 5 words of each topic are extracted as aspect words. Four evaluators joined the evaluation, and 825 test cases are collected. Figure 1 shows the result. As is shown in figure 1, our method overperforms LDA-based method in terms of consistency. This is because LDA-based method tends to cluster words that cooccur repeatedly without considering their consistence, for example the director "Yimou Zhang" and "Oscar" in the movie "The Flowers of War". The proposed system also considers cooccurrence by using PMI of apsect words, but we also use a synonym dictionary to leverage semantic information.

4.4 Quantitative analysis of opinions identification

Table 3: Results of opinions extraction

Events name	Method	Precision	Recall	F1
The Flowers of War	baseline method	0.4242	0.6788	0.5221
	Proposed method	0.6012	0.6121	0.6066
Poison Milk of Mengniu	baseline method	0.2982	0.5764	0.3931
	Proposed method	0.4082	0.5096	0.4533
Event of Yue Yue	Baseline method	0.4211	0.6178	0.5008
	Proposed method	0.4808	0.6126	0.5388
Death of Gaddifi	Baseline method	0.311	0.7956	0.4472
	Proposed method	0.5512	0.455	0.4985

For each event, two hundred microblogs are randomly selected. Opinions in these microblogs are manually labeled as standard answer.

Baseline method: dictionary-based method is used. Opinions are extracted based on the sentiment dictionary.

Evaluation metrics: Precision, recall and F1 score are used. Table 3 shows the result. Our method gets a better F1 score. Comparing to the baseline method, the proposed method generally trades a little drop of recall for a big improvement of precision. Precision is more important than recall in the information explosion era.

5 Conclusions and Future Work

In this paper, a system that performs opinion mining of event-related microblogs is presented. Unlike most of the existing approaches, our system goes beyond polarity classification. Specifically, it extracts the hottest aspects talked on microblog, and presents relevant opinion expressions expressed on each event aspect. To provide more concise result, PMI and synonymous information are leveraged to further cluster the extracted aspect words. Moreover, a method that combines dictionary-based method and CRF model is proposed to identify opinion expressions. Experiment results show our methods outperform the baseline methods dramatically.

However, there is still a lot work should be done to improve the result. As mentioned before, the performance of clustering process and opinion expressions identification are the bottlenecks

[3]Stanford Topic Model Toolbox: `http://nlp.stanford.edu/software/tmt/tmt-0.4/`

of our system. We are looking forward to find better method to avoid these disadvantages in the future.

References

[1] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. What is Twitter, a social network or a news media?. In: Proceedings of the 19th international conference on World Wide Web, WWW '10, 2010, pp. 591 – 600.

[2] Jaime Teevan, Daniel Ramage, and Merredith Ringel Morris. #TwitterSearch: a comparison of microblog search and web search. In: Proceedings of the fourth ACM international conference on Web search and data mining, WSDM '11, 2011, pp. 35 – 44.

[3] Luciano Barbosa and Junlan Feng. Robust sentiment detection on Twitter from biased and noisy data. In: Proceedings of the 23rd International Conference on Computational Linguistics: Posters, COLING '10, 2010, pp. 36 – 44.

[4] Dmitry Davidov, Oren Tsur, and Ari Rappoport. Enhanced sentiment learning using Twitter hashtags and smileys. In: Proceedings of the 23rd International Conference on Computational Linguistics: Posters, COLING '10, 2010, pp. 241 – 249.

[5] Long Jiang, Mo Yu, Ming Zhou, Xiaohua Liu, and Tiejun Zhao. 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, 2011, Vol. 1, pp. 151 – 160.

[6] Bo Pang and Lillian Lee. Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval Volume 2, Issue 1-2 (January 2008), 2008, pp. 1 – 135.

[7] Lu Chen, Wenbo Wang, Meenakshi Nagarajan, Shaojun Wang and Amit P. Sheth. Beyond Positive/Negative Classification: Automatic Extraction of Sentiment Clues from Mi-croblogs. Kno.e.sis Center Technical Report, 2011.

[8] Minqing Hu and Bing Liu. Mining and summarizing customer reviews. In: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '04, 2004, pp. 168 – 177.

[9] Ana-Maria Popescu and Oren Etzioni. Extracting product features and opinions from reviews. In: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, HLT '05, 2005, pp. 339 – 346.

[10] Tengfei Ma and Xiaojun Wan. Opinion target extraction in Chinese news comments. In: Proceedings of the 23rd International Conference on Computational Linguistics: Posters, COLING '10, 2010, pp. 782 – 790.

[11] Peter 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, ACL '02, 2002, pp. 417 – 424.

[12] Jonathan Chang, Jordan Boyd-Graber, Sean Gerrish, Chong Wang, and David M. Blei. Reading tea leaves: How humans interpret topic models. In: 23rd Annual Conference on Neural Information Processing Systems, NIPS’ 09 , 2009, pp. 288 – 296.

[13] Ivan Titov and Ryan McDonald. Modeling online reviews with multi-grain topic models. In: Proceedings of the 17th international conference on World Wide Web, WWW '08, 2008, pp. 111 – 120.

[14] Samuel Brody and Noemie Elhadad. An unsupervised aspect-sentiment model for online reviews. In: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, HLT '10, 2010, pp. 804 – 812.