

UIR-PKU: Twitter-OpinMiner System for Sentiment Analysis in Twitter at SemEval 2015

Xu Han^{2,4}, Binyang Li^{1*}, Jing Ma², Yuxiao Zhang³, Gaoyan Ou³,
Tengjiao Wang³, Kam-fai Wong^{2,4,5}

¹School of Information and Technology, University of Information Relations, Beijing

²Dept. of Sys. Engineering&Engineering Management, The Chinese University of Hong Kong

³School of Information Science, Peking University, Beijing

⁴Shenzhen Research Institute, The Chinese University of Hong Kong

⁵MoE Key Laboratory of High Confidence Software Technologies, China

{xhan, jma, kfwong}@cuhk.edu.hk; byli@uir.cn; {yxzhang, gyoyou, tjwang}@pku.edu.cn

Abstract

Microblogs are considered as We-Media information with many real-time opinions. This paper presents a Twitter-OpinMiner system for Twitter sentiment analysis evaluation at SemEval 2015. Our approach stems from two different angles: topic detection for discovering the sentiment distribution on different topics and sentiment analysis based on a variety of features. Moreover, we also implemented intra-sentence discourse relations for polarity identification. We divided the discourse relations into 4 predefined categories, including *continuation*, *contrast*, *condition*, and *cause*. These relations could facilitate us to eliminate polarity ambiguities in compound sentences where both positive and negative sentiments are appearing. Based on the SemEval 2014 and SemEval 2015 Twitter sentiment analysis task datasets, the experimental results show that the performance of Twitter-OpinMiner could effectively recognize opinionated messages and identify the polarities.

1 Introduction

This year comes the third edition of SemEval Twitter sentiment analysis task consisting of new genres, including topic-based polarity classification, trends detection towards a topic, and the sentimental strength of association of terms (Nakov et al., 2013).

We only participated in the subtask of message sentiment analysis and built up a system, named Twitter-OpinMiner for the task. Twitter-OpinMiner stems from two different angles: LDA-based topic detection for discovering the opinionated features of trending tweets' topics and sentiment analysis based on a variety of features.

• Topic detection

Recent studies show that people often search Twitter to find temporally relevant information (Teevan et al., 2011), such as emergent events, trending topics. In fact, similar opinions were likely to express on the same topic/event in Twitter. For example, there are 20 tweets expressing similar opinions on "Blood moon" in SemEval 2015 dataset. Therefore, it can facilitate us to discover the sentiment distribution on different topics.

• Sentiment analysis

Unlike traditional news content, tweets are specialists in short texts with long compound sentences, and a number of irregular expressions, including emoticon, hashtag, and special punctuations. In order to better support tweets analysis, we extract features from following aspects: textual content, irregular expression, discourse relations, and word embedding. Then we introduce above features into a SVM classifier for sentiment analysis.

This paper is organized as follows. Section 2 describes the framework of our system. Section 3 introduces the details of our feature extraction. We

*Corresponding author

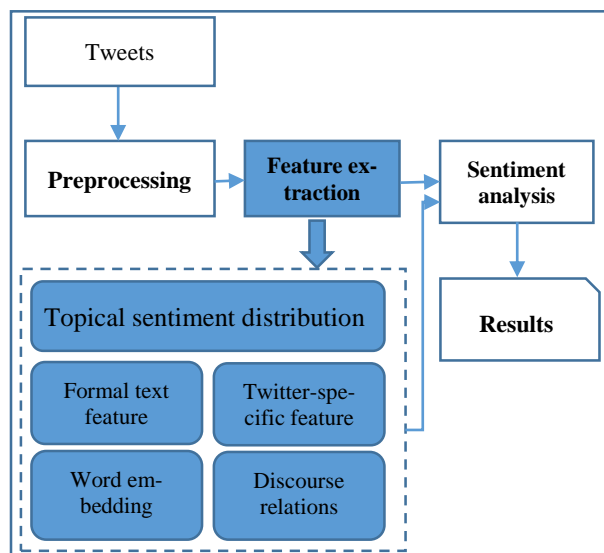


Figure 1. System architecture.

present the evaluation results in Section 4. Finally, Section 5 concludes the paper.

2 System Overview

2.1 Architecture

The architecture of Twitter-OpinMiner is described in Figure 1. Twitter-OpinMiner system is comprised of three modules:

- (1) Pre-processing module: reads all data of training data and test data. It performs, POS tagging, named entity recognition, and semantic role labeling.
- (2) Feature extraction module: extracts the features including formal text features, tweet-specific features, discourse features, sentiment distribution among topics, and word embedding.
- (3) Sentiment analysis module: creates a SVM classifier that incorporates the above features classify the polarity of each tweet.

Finally, Twitter-OpinMiner outputs the polarity of each tweet.

2.2 Development Data and Lexicon

The development data are necessary in our system. We fully utilize the training tweets provided by SemEval 2013. The dataset consists of 9,912 annotated tweets.

Besides, for sentiment analysis, we also utilize several sentiment lexicons, including Liu’s sentiment lexicon (Liu, 2012), MPQA subjectivity lexicon (Wilson et al., 2005), and the sentiment lexicon generated from tweets (Mohammad et al., 2013).

Table 1. Features of text in our system.

Word-Level and entity-level features
The presence of sentiment word
The ratio of sentiment word in a sentence
The total number of positive words
The total number of negative words
The presence of negation words
The total number of the word in all-caps
Bi-gram features
Named entities + opinion operators
Pronouns + opinion operators
Nouns or named entities + opinion words
Pronouns + opinion words
Opinion words (adjective) + (noun)

3 Feature Extraction

The objective of this task is to determine whether a given message is positive, negative, or neutral. We train sentiment classifiers with LibLinear (Fan et al., 2008) on the training set and dev set, and tune parameter $-c$, $-w_i$ of SVM on the test set of SemEval 2013. SVM is a popular machine learning algorithm, the effectiveness of which has been proved in sentiment analysis on formal texts in related work (Pang and Lee, 2002; Liu, 2012). Since the performance of SVM classifier will be greatly influenced by the features selection, we explore a variety of features in the evaluation.

3.1 Features of topical sentiment distribution

The advancement of Twitter is fast response to the real world, so people often search Twitter to find temporally relevant information, such as emergent events, trending topics. In fact, tweets are likely to converge on some opinions for a specific topic, which will lead to different sentiment distributions among topics.

In our system, we adopt LDA-based approach for representing the typical sentiment distribution features. We use the Mallet toolkit, set the topic number as 50, and map each tweet into 50 dimensions to extract those features.

3.2 Features of formal text

Although the task is to analyze sentiment in Twitter, much research proved the effectiveness of the classic features of formal texts on tweets. The features we adopted in this task are partly the same with (Zhou et al., 2010) and listed in Table 1, and two types of features are incorporated in the classifier.

These features are also integrated into our SVM classifier for training and treated as the baseline in our experiment.

3.3 Twitter specific feature

Unlike formal texts, tweet has its own characteristics, including irregular expressions, emoticon, hashtag, ill format, and special punctuations. In our system, we combine the features proposed by Mohammad et al. (2013) with some new features as Twitter-specific features for supplementary to the formal text.

- Hashtags: the number of hashtags in one tweet;
- Ill format: the presence of ill format with some characters replacing by *, for example, f**k;
- Punctuation: the number of contiguous sequences of exclamation marks, question marks, and both exclamation and question marks; whether the last token contains an exclamation or question mark;
- Emoticons: the presence of positive and negative emoticons at any position in the tweet; whether the last token is an emoticon;
- OOV: the ratio of words out of vocabulary;
- Elongated words: the presence of sentiment words with one character repeated more than two times, for example, ‘coool’;
- URL: whether the tweet contains a URL.
- Reply or Retweet: Is the current tweet a reply/retweet tweet

3.4 Word embedding

We also utilize word embedding technique for feature extraction. We adopt sentiment-specific word embedding method (Tang et al., 2014) that could encode sentiment information in the continuous representation of words. In our approach, each term is extended into a 150 dimensional vector.

3.5 Discourse specific feature

Since tweets are usually expressed informally, there are many compound sentences in a tweet, which always contain positive sentiment and negative sentiment with ambiguity. For example,

*It may not be the biggest squad in the last 10yrs, **but** Ancelotti is working for quality over quantity. Everyone... <http://t.co/oCdPXOWggT>.*

Table 2. Examples of cue-phrases.

Relation	Cue Phrases
<i>Contrast</i>	<i>although, but, however, though</i>
<i>Condition</i>	<i>if, despite, in case of</i>
<i>Continuation</i>	<i>and, moreover, not only but also</i>
<i>Cause</i>	<i>because, so that, due to, in order that</i>

In this case, there are two segments in the tweet that holds a Contrast discourse relation, and the polarity is determined by “but” segment. In our system, we also take into consideration of intra-sentence discourse relation features for processing compound sentences.

Mann and Thompson (1988) defined a complete discourse scheme Rhetorical Structure Theory (RST). Since not all of the discourse relations in RST would help eliminate polarity ambiguities, the discourse relations were implemented in our system was on a subset (Zhou et al., 2011).

In our system, we use cue-phrase based method for discourse relation identification. We maintain a cue phrase lexicon and the examples of the cue phrases were shown in Table 2.

4 Experiment

We trained a SVM classifier on 9,912 annotated tweets (8,258 in the training set and 1,654 in the development set). We used the same evaluation metrics with SemEval 2013, including the macro-averaged F-score of the positive and negative classes. The experimental results obtained by our system on the training set (ten-fold cross validation), development set, and test sets on Twitter 2013 were shown in Table 3 where the baseline was achieved by using the formal text features as well as twitter-specific features. Since the effectiveness of these two types of features were analyzed in (Mohammad et al., 2013), we mainly evaluated the effectiveness of other features.

Table 3 showed that the most effective feature on Twitter 2013 dataset turned out to be the word embedding features: they provided gains of about 7%. For LDA, we set the numbers of topic from 10 to 100, and found it could achieve best performance when equaling 50. We then constructed the sentiment distribution among 50 topics for the further evaluation.

Besides, we also investigated the effectiveness of discourse features on compound sentences, and the statistics were shown in Table 6.

Table 3. Experimental results on Twitter 2013 dataset.

Approaches	Metrics						
	pos-P	pos-R	pos-F	neg-P	neg-R	neg-F	ave-F
Baseline (BL)	0.743	0.673	0.706	0.451	0.679	0.542	0.624
BL+LDA	0.752	0.679	0.714	0.465	0.707	0.561	0.634
BL+Word Embedding	0.772	0.685	0.724	0.561	0.798	0.659	0.692
BL+Discourse Relation	0.756	0.680	0.716	0.467	0.705	0.562	0.635
BL+All	0.791	0.704	0.745	0.563	0.809	0.664	0.704

Table 4. Experimental results of 2015 test.

Method	Metrics						
	pos-P	pos-R	pos-F	neg-P	neg-R	neg-F	ave-F
UIR-PKU	0.7518	0.6098	0.6734	0.4636	0.6110	0.5272	0.6003
Best run	0.7702	0.6975	0.7321	0.5171	0.6219	0.5647	0.6484

Table 5. Experimental results of progress test on Average F-value.

Approaches	Corpus			
	Live Journal 2014	SMS 2013	Twitter 2014	Twitter 2014 Sarcasm
UIR-PKU	0.7044	0.6741	0.6718	0.5258
Best run	0.7534	0.6716	0.7448	0.4286

Table 6. Distribution of discourse relations and the contribution in the evaluation.

Discourse Relation	Occurrence	Contribution
Cause	26.9%	33.9%
Condition	12.6%	22.1%
Contrast	18.2%	10.1%
Continuation	42.3%	33.9%

By adopting discourse features, around 59% sentences with discourse relations were identified. Among these four types of relations, better performance were achieved on *cause* and *condition* relations. Especially for the sentences with *condition* relation, they were all classified correctly. It is because that more cue-phrase of *cause* and *condition* relations were used to explicitly denote the discourse relations in tweets, but more likely use context to imply *contrast* and *continuation* relations.

Table 4 and Table 5 showed the evaluation results in SemEval 2015 Task 10. Compared with the best run in Table 5, our system achieved comparable results on Twitter sentiment analysis and better performance on the evaluation of *sarcasm*. In fact, many sarcasm are likely expressed in ironic, hence most feature types are ineffective for this case. In our system, we also used the features of topical sentiment distribution, which assumed the polarity of *sarcasm* tweet the same with *non-sarcasm* tweets.

5 Conclusion

We describe our Twitter-OpinMiner systems for participating in SemEval 2015 sentiment analysis in Twitter. Our approach stems the features from two different aspects: topical sentiment distribution and a variety of short text based features. In our paper, we also implemented intra-sentence discourse relations for polarity identification in compound sentences where both positive and negative sentiments are appearing. In this way, the polarity ambiguities will be eliminated. Based on SemEval 2015 and SemEval 2014 datasets for Twitter sentiment analysis task, we examined the performance of Twitter-OpinMiner, which could achieved comparable results on recognizing opinionated messages and identifying the polarities.

Acknowledgments

This research is partially supported by Fundamental Research Funds for the Central Universities (3262014T75, 3262015T20), Shenzhen Fundamental Research Program (JCYJ20130401172046450), General Research Fund of Hong Kong (417112). We also thank Liyu Chen, Jianxiong Wu, and anonymous reviewers for their helpful comments.

References

- Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. 2008. LIBLINEAR: A Library for Large Linear Classification. *Journal of Machine Learning Research*, 9: 1871-1874.
- Bing Liu, Mingqing Hu, and Junsheng Cheng. 2005. Opinion Observer: Analyzing and Comparing Opinions on the Web. In *Proceedings of the 14th international conference on World Wide Web*, pages 342-351, Chiba, Japan, ACM.
- Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1): 1-167.
- William C. Mann and Sandra A. Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text-Interdisciplinary Journal for the Study of Discourse*, 8(3): 243-281.
- Saif M. Mohammad, Svetlana Kiritchenko, and Xiaodan Zhu. 2013. NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets. In *Proceedings of the 2nd Joint Conference on Lexical and Computational Semantics*.
- Preslav Nakov, Sara Rosenthal, Zornitsa Kozareva, Veselin Stoyanov, Alan Ritter, and Theresa Wilson. 2013. SemEval-2013 task 2: Sentiment analysis in twitter. In *Proceedings of the International Workshop on Semantic Evaluation*, volume 13.
- Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment Classification Using Machine Learning Techniques. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 79-86.
- Duyu Tang, Furu Wei, Nan Yang, Ming Zhou, Ting Liu, and Bing Qin. 2014. Learning sentiment-specific word embedding for twitter sentiment classification. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pages 1555-1565.
- Jaime Teevan, Daniel Ramage, and Meredith R. Morris. 2011. #TwitterSearch: A Comparison of Microblog Search and Web Search. In *Proceedings of the 4th ACM International Conference on Web Search and Data Mining*, pages 35-44.
- Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the International Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 347-354.
- Lanjuan Zhou, Yunqing Xia, Binyang Li, and Kam-fai Wong. 2010. WIA-Opinmine System in NTCIR-8 MOAT Evaluation. In the 8th NTCIR Workshop Meeting, pages 286-292.
- Lanjuan Zhou, Binyang Li, Wei Gao, Zhongyu Wei, and Kam-fai Wong. 2011. Unsupervised discovery of discourse relations for eliminating intra-sentence polarity ambiguities. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 162-171.