

Sentiment Classification for Movie Reviews in Chinese by Improved Semantic Oriented Approach

Qiang Ye

Wen Shi

Yijun Li

*School of Management,
Harbin Institute of
Technology
Harbin, 150001, China
yeqiang@hit.edu.cn*

*Department of Computer
Science, Northeast
Agricultural University
Harbin, 150030, China
shiw@neau.edu.cn*

*School of Management,
Harbin Institute of
Technology
Harbin, 150001, China
liyijun@hit.edu.cn*

Abstract

Sentiment classification aims at mining reviews of customers for a certain product by automatic classifying the reviews into positive or negative opinions. With the fast developing of World Wide Web applications, sentiment classification would have huge opportunity to help people automatic analysis of customers' opinions from the web information. Automatic opinion mining will benefit to both consumers and sellers. Up to now, it is still a complicated task with great challenge. There are mainly two types of approaches for sentiment classification, machine learning methods and semantic orientation methods. Though some pioneer researches explored the approaches for English movie review classification, few jobs have been done on sentiment classification for Chinese reviews. The improved semantic approach for sentiment classification on movie reviews written in Chinese was proposed in this paper. Data experiment shows the capability of this approach.

Key words

Sentiment classification, movie reviews, Chinese, Semantic Oriented

1 Introduction

If you are considering a vacation travel to Kauai Island, Hawaii, you might go to a search engine, say Google, and search reviews of other visitors for it. However, in this case, Google would report about many matched pages for you. It would be useful to know what fraction of these matches recommend Kauai Island as a good travel destination. With an algorithm for automatically classifying a review as positive or negative, it would be possible for a search engine to report such summary statistics. It is called sentiment classification.

Sentiment classification aims to mining the text of written reviews of customers for a certain product by classifying the reviews into positive or negative opinions^[1, 3, 5]. According to the result of sentiment

classification, consumers would know which products to buy or not to buy and sellers would know the response of their costumers and the performances of their competitors'. Given the advance in machine learning and computing resources, opinions and reviews on several genres of products and services can be semi-automatically classified into recommended or not recommended. Examples of past work include mining reviews of automobiles, banks, movies, travel destinations^[3], electronics^[1, 3] and mobile devices^[3]. Potential applications include extracting opinions or reviews from discussion forums efficiently, and integrating automatic review mining with search engines to provide quick statistics of search results.

Special challenges are associated with movie review mining. As it has been pointed out elsewhere^[5], movie review mining is very domain specific and word semantics in a particular review could contradict with overall semantic direction (good or bad) of that review. For example, an "unpredictable" camera gives negative meaning to that camera model, whereas a movie with "unpredictable" plot sounds positive to moviegoers.

There are mainly two types of approaches for sentiment classification. One is machine learning methods, the other are semantic orientation methods.

Turney (2002)^[3] presented a Semantic Orientation (SO) mining method based on PMI-IR algorithm for sentiment classification by combining the Point Mutual Information (PMI) and the statistic data collected by Information Retrieval (IR).

The first step of the algorithm is to use a part-of-speech tagger to identify phrases in the input text that contain adjectives or adverbs^[10]. The second step is to estimate the *semantic orientation* of each extracted phrase^[6]. A phrase has a positive semantic orientation when it has good associations (e.g., "romantic ambience") and a negative semantic orientation when it has bad associations (e.g., "horrific events"). The third step is to assign the given review to a class, recommended or not recommended, based on the average semantic orientation of the phrases extracted from the review.

If the average is positive, the prediction is that the review recommends the item it discusses. Otherwise, the prediction is that the item is not recommended. The PMI-IR algorithm is employed to estimate the semantic orientation of a phrase [7]. PMI-IR uses Pointwise Mutual Information (PMI) and Information Retrieval (IR) to measure the similarity of pairs of words or phrases. The semantic orientation of a given phrase is calculated by comparing its similarity to a positive reference word (“excellent”) with its similarity to a negative reference word (“poor”). More specifically, a phrase is assigned a numerical rating by taking the mutual information between the given phrase and the word “excellent” and subtracting the mutual information between the given phrase and the word “poor”. In addition to determining the direction of the phrase’s semantic orientation (positive or negative, based on the sign of the rating), this numerical rating also indicates the strength of the semantic orientation (based on the magnitude of the number).

Up to now, most of the sentiment classification researches are focus on mining reviews written in English. No sentiment classification studies were reported for Chinese reviews.

According to a recent research by the CNNIC, there are about 94 million Internet users in China by the end of 2004, ranking number 2 in the world. There are also many product review web sites for various products from cell phones to movies. But no researches for the sentiment classification on Chinese reviews are reported by now.

Because the differences in characteristic between the two languages of Chinese and English, existing English oriented SO approaches were hard to imply directly on Chinese sentiment classification.

In this study, we want to make some improvement to the English semantic oriented approach to explore the sentiment classification method for Chinese text by SO approach on Chinese movie reviews.

2 Background

Several studies for sentiment classification are reported. Turney (2002) proposed the SO approach for sentiment classification and tested the SO approach on reviews about automobiles, banks, movies and travel destinations. Kushal Dave (2003) also researched the semantic classification of product reviews[11]. Chaovalit and Zhou compared the SO approach with machine learning approach by applying the SO approaches to movie reviews.

We summarized the process of typical Semantic Orientation SO approach for sentiment classification of Turney (2002) and Chaovalit(2005),etc, into 4 steps.

Step 1, parse and tag the part-of-speech to the

review documents by lexical analysis tools.

Step 2, Based on part-of-speech in the parsed output, only two-word phrases conforming certain patterns were selectively extracted.

Step 3, With two reference words pair (RWP) “excellent” and “poor” (presenting positive and negative opinions respectively), determine the semantic orientation of a phrase’ $SO(phrase)$ according to Equation (1)^{[3][5]}.

$$SO(phrase) = \log_2 [$$

$$\frac{hits(phrase NEAR "excellent") hits("poor")}{hits(phrase NEAR "poor") hits("excellent")}] \quad (1)$$

Here, $hits(\cdot)$ denotes the number of pages returned for a query consisting of phrase \cdot from a search engine. For example, $hits('poor')$ represents the number of pages returned for a search query ‘poor’. When there are both $phrase$ and ‘excellent’ (or ‘poor’) connected by NEAR operator in the parameter of $hits$ function, it defines the similarity between $phrase$ and ‘excellent’ (or ‘poor’). In other words, the similarities were measured with cooccurrences of the phrases and ‘excellent’ (or ‘poor’). This step could be automated by sending search queries to a search engine like Google and get the account number of return pages. A phrase’s semantic orientation would be positive if it is associated more strongly with “excellent” than “poor” and would be negative if it is associated more strongly with “poor” than “excellent”.

Step 4 a review’s semantic orientation was calculated by averaging the SO values of all the extracted phrases in it. The opinion will be positive if its average semantic orientation exceeds a threshold and is negative if otherwise. In Turney’s study the threshold was set to Zero^[3].

But the above approach for English text opinion classification can’t be applied onto the sentiment classification for Chinese movie reviews for at least two points because the different characteristic in these two language.

i. Word segmentation for Chinese text

It is easy to divide one word from other words in a text because there is usually a space sign between different English words. That makes possible for computer to automatic divide an English text into words. But, Chinese text have no this kind of signs to help computer to divide Chinese characters in a text into words. So a technique named Word Segmentation would be employed in Chinese review sentiment classification

ii. Determination of Reference Words Pair (RWP)

All the existing studies use “excellent” and “poor” as the RWP to present the extremely positive

and extremely negative opinion, because “excellent” and “poor” often represent two extreme attitude in 5 or 7 levels evaluation system. But it is difficult to decide that what Chinese words most match to the English words pairs of “excellent” and “poor” without a certain context. In Chinese, words to express extremely positive and extremely negative attitudes are very different for different products. For example, the common words for expressing extremely positive attitude to a craftwork, say a china teapot, would be “Jing Zhi” (delicate), but the same attitude to a movie would be “Jing Dian” (classical). So, further study should be conduct to optimize the process of RWP selection.

3 Methodology

3.1 Semantic Orientation approach for sentiment classification to Chinese Reviews

We established a SO approach with 6 steps for Chinese movie reviews.

Step 1 Chinese Words segmentation

We develop a tool for Chinese words segmentation. This job can also be conducted by ICTCLAS.

Step 2, Parse and tag the part-of-speech to the review documents by Chinese lexical analysis tools. We employed a Chinese POS tool of ICTCLAS (Institute of Computing Technology, Chinese Lexical Analysis System) from CAS (Chinese Academy of Sciences), (<http://mtgroup.ict.ac.cn/~zhp/ICTCLAS/>).

Step 3, Based on parts-of-speech in the parsed output, phrases with two Chinese words conforming for certain patterns were selectively extracted. Only two-word phrases conforming to certain patterns were extracted for further processing. We adopted English phrase patterns from Turney’s study [3]. Adjective or adverb in the patterns provides subjectivity, while the other word provides context. Table 1 summarizes five patterns for extracting Chinese phrases. We developed an extracting tool for extracting two-Chinese-word pairs based on java language. Knowing, one Chinese word sometimes consists of more than one Chinese Characters, so two words not always means having only two Chinese characters.

Table 1 two-word phrase patterns

ID	First word	Second word
1	Adjective	Noun
2	Adverb	Adjective
3	Adjective	Adjective
4	Noun	Adjective
5	Adverb	Verb

Step 4, Select optimal Reference Words Pair (RWP) to present the extremely positive and extremely negative opinions.

For a certain kind of product, get several of the most frequently used words to present the extremely positive opinion and the extremely negative opinion. Then, match these words in two-word pairs, and calculate the SO value of all the pairs according to formula (2) on a manually selected positive document and negative document. For each two-word pair, minus SO value of negative document from the SO value of positive document we got the value D. The two-word pair with the maximum D is the selected RWP.

To get the *hits(phrase)* data we develop a web based tool in Java, integrated with the interface of Google Web API.

Step 5, With the optimal RWP, *rw-p* and *rw-n*, (presenting positive and negative opinions respectively), determine the semantic orientation of a phrase’ *SO(phrase)* according to Equation (2)^{[3][5]}.

$$SO(phrase) = \log_2 \left[\frac{hits(phrase \text{ AND } rw-p)hits(rw-n)}{hits(phrase \text{ AND } rw-n)hits(rw-p)} \right] \quad (2)$$

Here, *rw-p* denotes the positive Reference Words Pair. A phrase’s semantic orientation would be positive if it is associated more strongly with *rw-p* than *rw-n* and would be negative if it is associated more strongly with *rw-n* than *rw-p*. Here we use the operator AND in Equation (2).

Step 6, Calculate the average SO values of all the extracted phrases in a review to determine its semantic orientation. The opinion will be judged as positive if its average semantic orientation exceeds a threshold and is negative if otherwise. The SO value threshold can be calculated from the result of step 5 by average the SO value of both the positive and negative document corresponding to the optimal RWP.

3.2 Test data

There is no ready-to-use Chinese movie reviews data on the web. To get the test data, what we can do is to search on the web and collect the data we need. There are many product review websites or review forum in Chinese. Finally, we keep our eyes on to a website which specially offers movie reviews (<http://www.mov8.com/>). We retrieved 550 pieces of movie reviews. All the movie reviews were manually examined one by one. The introduction to the movies and the reviews with uncertain attitude were all excluded. At last, we got 170 pieces qualified movie reviews, and labeled

them into two parts, 40 pieces of them positive and 130 pieces of them negative.

3.3 Evaluation of the Performance

To evaluate the performance of Semantic Orientation Approach, we adopted three indexes that are generally used while in text categorization: Recall, Precision, Accuracy.

Table 2 Contingency Table for performance evaluation

	actual positive reviews	actual negative reviews
predict positive	<i>a</i>	<i>b</i>
predict negative	<i>c</i>	<i>d</i>

The indexes can be calculated according the figures in table 2 according the following formulas respectively.

$$Accuracy = \frac{a + d}{a + b + c + d}, \text{ Recall}(p) = \frac{a}{a + c},$$

$$Precision(p) = \frac{a}{a + b}, \text{ Recall}(n) = \frac{d}{b + d},$$

$$Precision(n) = \frac{d}{c + d}$$

Here, Recall(p) and Precision(p) is the recall ratio and precision ratio for actual positive reviews, Recall(n) and Precision(n) is are the recall ratio and precision ratio for actual negative reviews and Accuracy is the whole Accuracy of the SO approach.

4 Experiment result

4.1 Optimal RWP

In order to select optimal RWP, we design a questionnaire, which required the person being inquired to list 6 words used to describe their attitude to a movie, three words for extremely positive opinion and 3 words for extremely negative opinion. We collected 70 copies of such questionnaire result from the college student in HIT. And then we count the numbers of appearance for the positive words and sort the words according to their appearance frequency, the three most frequently appeared words are selected to by candidate positive represents. They are “Cheng Gong” (Successful), “Jing Dian” (Classical) and “Jing Cai” (Wonderful). We did so for the words used to describe negative opinion to movies and got “Shi Bai” (failing), “Shi Wang” (disappoint) and “Wu Liao” (boring) as the candidate negative represents. The positive candidates and negative

candidates were matched respectively being 9 candidate RWPs. Then according to formula (2) calculate the average SO value on a manually selected positive document and negative document, the result is shown in table 3.

Table 3 Average SO Value for Candidate RWPs

ID	Candidate RWP	P	N	D=P-N
1	Successful/failing	-1.6338	-2.0581	0.4243
2	Successful/disappoint	-1.3843	-2.2721	0.8878
3	Successful/boring	-2.5628	-3.3106	0.7478
4	Classical/failing	0.0044	-0.8864	0.8908
5	Classical/disappoint	0.4111	-0.9126	1.3237*
6	Classical/boring	-0.7157	-1.7070	0.9913
7	Wonderful/failing	-1.6339	-2.3449	0.7111
8	Wonderful/disappoint	-1.2275	-2.0475	0.8200
9	Wonderful/boring	-1.7909	-2.2770	0.4861

P : Average SO value of positive movies

N : Average SO value of negative movies

The RWP “Classical”/ “disappoint” have the highest D value, it was selected as the RWP.

4.2 Obtain the SO Threshold

We adopt the average SO of the selected RWP to both the positive and negative review in table 3 as the SO threshold. That is -0.2508. Here, we can also find that different RWP would bring very different threshold figures. It shows the great influence of RWP on the SO approach.

4.3 Opinion Classification by Semantic Orientation approach

We selected 72 reviews, 36 positive and 36 negative to conduct opinion classification. The result was shown in table 4.

$$Accuracy = \frac{27 + 28}{27 + 8 + 9 + 28} = 73.61\%,$$

$$Recall(p) = \frac{27}{27 + 9} = 75.00\% ,$$

$$Precision(p) = \frac{27}{27 + 10} = 72.97\% ,$$

$$Recall(n) = \frac{26}{26 + 10} = 72.22\% ,$$

$$Precision(n) = \frac{26}{26 + 9} = 74.29\%$$

Table 4 Contingency Table for performance evaluation

	actual positive reviews	actual negative reviews	total
predict positive	27	10	35
predict negative	9	26	37
total	36	36	72

Comparing with the same figures in existing studies for English review mining, the accuracies were commonly between 60% to 80%, the proposed SO approach for Chinese reviews performs not bad.

5 Summary and conclusion

According to the result, we concluded that the performance of proposed semantic orientation approach for Chinese review sentiment classification is acceptable compare with the figures of previous English movie review classification studies. Data test also indicated that the selection of both Reference-Word-Pair (RWP) and the SO value threshold have great influence on the performance of SO approach for opinion classification.

Word segmentation method was introduced into the sentiment classification process. That is one of the main differences between Chinese and English languages in sentiment classification. This study also introduced an optimal Reference-Word-Pair (RWP) selecting process and improved the selection of optimal SO value threshold. The research process implied that when SO approach is applied to customers' sentiment analysis in different fields to different products, different RWP and different threshold are required.

During the phrase extracting process we simply accept the two word patterns from approaches for English review classification [3] [5]. But the language structures of English and Chinese have many differences, so we assumed if more specific words patterns for Chinese reviews mining could be found, it will improve the performance of opinion classification greatly. That will be studied in the future research.

6. Acknowledgement

This research was sponsored by National Natural

Science Foundation of China Project (70501009), Heilongjiang Natural Science Foundation project (G0304), and Social Science Foundation of Harbin Institute of Technology project (No 050417) .

References

- [1] Peter D. Turney and Michael L. Littman, Measuring Praise and Criticism: Inference of Semantic Orientation from Association, ACM Transactions on Information Systems, vol.21, pp.315-346, 2003
- [2] V N Vapnik. The Nature of Statistical Learning Theory [M]. New York: Springer,1998
- [3] Peter D. Turney, Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews, presented at the Association for Computational Linguistics 40th Anniversary Meeting, New Brunswick, N.J., 2002
- [4] V N Vapnik. Statistical Learning Theory [M]. New York: John Wiley and Sons Inc, 1998
- [5] Pimwadee Chaovalit, Lina Zhou. Movie Review Mining: a Comparison between Supervised and Unsupervised Classification Approaches, Proceedings of the 38th Hawaii International Conference on System Sciences, 2005
- [6] V. Hatzivassiloglou and K. R. McKeown, Predicting the semantic orientation of adjectives, presented at Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and the 8th Conference of the European Chapter of the ACL, 1997
- [7] Turney, P.D. Mining the Web for synonyms: PMI-IR versus LSA on TOEFL. *Proceedings of the Twelfth European Conference on Machine Learning* (pp. Berlin: Springer-Verlag, 2001:491-502
- [8] Janyce M. Wiebe, Learning Subjective Adjectives from Corpora, presented at the 17th National Conference on Artificial Intelligence, Menlo Park, California, 2000
- [9] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan, Thumbs up? Sentiment Classification using Machine Learning Techniques, presented at the 2002 Conference on Empirical Methods in Natural Language Processing(EMNLP2002), 2002
- [10] Brill, E. Some advances in transformation-based part of speech tagging. Proceedings of the Twelfth National Conference on Artificial Intelligence (pp. 722-727). Menlo Park, CA: AAAI Press. ,1994
- [11] Kushal Dave, Steve Lawrence, and David M. Pennock.. Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. 12th international conference on World Wide Web, Budapest, Hungary, 2003.