Using Topic Sentiment Sentences to Recognize Sentiment Polarity in Chinese Reviews

Jiang Yang
School of Literature
Communication University of China
yangjiang@cuc.edu.cn

Min Hou
Broadcast Media Language Branch
Communication University of China
houminxx@263.net

Abstract

An approach to recognizing sentiment polarity in Chinese reviews based on topic sentiment sentences is presented. Considering the features of Chinese reviews, we firstly identify the topic of a review using an n-gram matching approach. To extract candidate topic sentiment sentences, we compute the semantic similarity between a given sentence and the ascertained topic and meanwhile determine whether the sentence is subjective. A certain number of these sentences are then selected as representatives according to their semantic similarity value with relation to the topic. The average value of the representative topic sentiment sentences is calculated and taken as the sentiment polarity of a review. Experiment results show that the proposed method is feasible and can achieve relatively high precision.

1 Introduction

Sentiment analysis, also known as "opinion mining", is the problem of analyzing the sentiment, opinion or any other subjectivity of written texts. With its potential applications to opinion search engine, public opinion analysis, product promotion, etc., sentiment analysis has been receiving increasing interest in recent years.

What sentiment analysis processes are texts with subjectivity which mainly describe the writers' (or on behalf of a group or an organization) private thoughts, attitudes or opinions on phenomena, persons, affairs and so on. Although various kinds of writings such as narration and exposition are possible to contain subjectivity, argumentation is the focus of sentiment analysis on which researchers put much strength at present. As a kind of argumentation and a typical and common subjective text, a review comments on some specific phenomenon, person or affair. Reviews, especially news reviews, have a certain degree of influence on public opinion in virtue of mass media. Domain-specific reviews like automobile, hotel, movie reviews have potential commercial value respectively. Therefore, recognizing sentiment polarity (SP thereafter) in reviews becomes necessary and practical.

Language is a hierarchical symbol system, which allows sentiment analysis to be conducted on different language levels. In general, most current studies concerning sentiment analysis are about determining the SP of words, phrases or sentences. Only a fraction of them addressed discourse level sentiment analysis. This paper, aiming at recognizing the overall SP of Chinese reviews, proposes a topic-sentiment-sentence based approach to carry out a discourse level sentiment analysis.

The remainder of this paper is organized as follows. Related works are presented in section 2. Section 3 is problem analysis and method description. Section 4 describes topic identification and topic sentiment sentence extraction. Section 5 is about recognizing SP in Chinese reviews using the extracted topic sentiment sentences. Section 6 is the experiment results and section 7 is the conclusion.

2 Related Works

The SP determination can be generally conducted on three language levels: the word level, the sentence level and the discourse level. The two main popular approaches, especially in real-world applications, have been based on

machine learning techniques and based on semantic analysis techniques. Research aiming at recognizing the overall SP of discourse is represented by Turney (2002), Pang et al. (2002) and Yi et al. (2003). Turney proposed an unsupervised learning algorithm to classify the sentiment orientation of reviews. The mutual information difference between the given word or phrase and the words "poor" and "excellent" was calculated respectively to measure its semantic orientation; then the average semantic orientation of all the words in a given text was regarded as the overall semantic orientation. Pang et al. employed such classification models as Naïve Bayesian model, Maximum Entropy model and Support Vector Machine model to classify the semantic orientation of movie reviews, in which the features of models selected included unigrams, bigrams, parts of speech, word position, feature frequency and feature presence. Yi et al. firstly analyzed the grammatical structure of sentences using NLP techniques. The semantic orientation of a sentence then is determined by referring to a sentiment lexicon and a sentiment pattern database. They applied the approach to classifying the overall SP of document.

Other related works are concerning the sentiment analysis of sentences and words which underlie recognizing the overall SP of a whole text. Wiebe et al. (2000, 2004) proved that the subjectivity of a sentence could be judged according to the adjectives in it. Kim & Hovy (2004) and Weibe & Riloff (2005) explored the classification of subjective and objective sentences. Yu et al. (2003) put forward an approach to extract opinionated sentences in order to serve an automatic question answering system. The extracted sentences were classified and the SP of each was determined. Hu & Liu (2004) took advantage of WordNet to obtain sentiment words and their orientations. The polarity of a sentence thus is judged according to the dominant semantic orientation of sentiment words.

For Chinese, Wang et al. (2005) proposed a hybrid approach to recognize the semantic orientations of sentences in reviews based on heuristic rules and Bayesian classification technique. Wang et al. (2007) applied a Multi-redundant-labeled CRFs method on sentence sentiment analysis. Experiments showed it solved ordinal regression problems effectively and obtained global optimal result over multiple cascaded subtasks. Meng et al. (2008) designed a recognition system of text valence based on key word template in which they proposed template matching arithmetic and text valence value arithmetic for the calculation of the valence of Chinese texts. Zheng et al. (2009) conducted a research on sentiment analysis to Chinese traveler reviews by SVM algorithm.

3 Problem Analysis and Method Description

3.1 Discourse Structure of Chinese Texts

The overall SP of a Chinese text is the sum of the SP of all its component parts. However, the importance of each component part in a given text varies. This is because no matter which writing style a text belongs to, it has a particular discourse structure which determines the importance of the component parts.

Discourse structure is the organization and constitution law of language units (greater than sentence) within a discourse. It formally indicates the hierarchy of discourse contents, semantically guarantees the integrity of discourse contents and logically reflects the coherence of discourse contents. In a word, discourse structure is the unity of discourse form, discourse meaning and discourse logic. A discourse consists of several semantic parts. The central meaning of a discourse is the aggregation of the central meaning of its semantic parts in a certain logic way. A semantic part is the set of paragraphs. It may be composed of as small as only a paragraph or as large as even a whole chapter. The basis for partitioning semantic parts depends on the writing styles, i.e., narration, description, argumentation and exposition. For argumentation, a typical argumentation may be divided into 4 parts as introduction, viewpoint presentation, demonstration and conclusion. Recognizing semantic parts has great significance in understanding the central idea of a text.

3.2 Features of Chinese Reviews

Chinese reviews are a kind of argumentation. According to what is reviewed, they can be categorized into finance reviews (e.g., stock review), literature reviews (e.g., book review),

product reviews (e.g., automobile review), current affairs reviews (e.g., news review), etc. Generally speaking, Chinese reviews bear the following features.

Firstly, the topic of a Chinese review is explicit. A Chinese review always comments on some specific phenomenon, person or affair. The object it deals with is very explicit.

Secondly, a Chinese review has generally only one topic. Thus, in a Chinese review, the reviewer always explicitly expresses his/her opinion towards the topic. The sentiment of the discussed topic is rather explicit. Some Chinese reviews may discuss subtopics and corresponding opinions on each subtopic may be shown. But it will not change or influence the reviewer's basic sentiment on the topic.

Thirdly, the topic of a Chinese review is closely related to its title. Chinese Reviews often use concise expressions in titles to show clearly the topics or the themes. Therefore, the topic of a review can generally be found in its title.

Fourthly, Chinese reviews have fixed expression patterns. A typical Chinese review consists of 4 semantic parts as is mentioned above. The reviewer's sentiment expressions towards the topic generally appears in the "viewpoint presentation" and "conclusion" part.

To prove the correctness of our knowledge of Chinese reviews, we conducted a survey on 560 Chinese reviews which were collected from newspapers and the Internet. The manually examined results, which are showed as follows, verify the above mentioned 4 features of Chinese reviews.

Table 1 A Survey on Features of Chinese Reviews

Features		Percent
Explicit Topic		100
One Topic		100
Title Reflects Topic		99.64
Discourse Structure	I-D-C[1]	40.17
	I-V-D-C	33.9
	I-V-D	18.75
	others	7.18

[1] "I" stands for introduction, "D" for demonstration, "C" for conclusion and "V" for viewpoint presentation.

3.3 Topic Sentiment Sentence

According to the above analysis, the SP of a Chinese review is manifested by a certain expression pattern through several semantic parts, and its overall SP is generally expressed in the "viewpoint presentation" and "conclusion" part. Thus a straightforward idea to obtain the SP of a Chinese review is to: (1) partition the review into several semantic parts; (2) distinguish the viewpoint presentation part and the conclusion part; (3) analyze only the sentiment of the viewpoint presentation part and the conclusion part and take the result as the overall SP of the review. Intuitively, this seemingly simple method can achieve very good result.

However, to perform an automatic discourse structure analysis itself is actually a hard task and will lose precision during the processing; to distinguish different semantic parts by means of language cues without a discourse structure analysis can only solve some instead of all problems. Therefore, we introduce the concept of topic sentiment sentence.

A topic sentiment sentence is defined as a sentence bearing both the topic concept and sentiment towards that topic. The topic sentiment sentences in a Chinese review are the intersection of the topic sentences and sentiment sentences in it. Topic sentiment sentences are representative for sentiment analysis because, firstly, they are homogeneous in topic. And more importantly, the sentiment bearing in these sentences refer to the same topic. This makes sentiment in each sentence computable. Earlier works like Turney (2002) or Pang et al (2002) don't take into account the topic and the sentiment relating to that topic together as a whole, thus makes the result less reliable in that the sentiment words and phrases processed are not homogeneous in topic. Secondly, the degree of semantic similarity between topic sentiment sentences and the topic of the review reflects a potential relatedness between the topic sentiment sentences and their corresponding semantic parts. The more a topic sentiment sentence is similar in meaning to the topic, the more likely it appears in the viewpoint presentation part or conclusion part. This is just the reason we avoid an analysis of discourse structure of a review. We also try to avoid an automatic partition of semantic parts of a review since the topic sentiment sentences

themselves potentially point out the corresponding semantic parts they belong to. Thirdly, the distribution of the topic sentiment sentences, including density and extensity, reflects more or less the writer's intensity of attitude toward what is being discussed and can help with detailed sentiment analysis.

To summarize, with topic sentiment sentences, we can compute the SP of a Chinese review in a more simple and effective way without an automatic discourse structure analysis. Moreover, we can obtain a "shallow" structure since topic sentiment sentences potentially reflect the discourse structure of Chinese reviews.

3.4 Method

We thus propose a new method to recognize the sentiment polarity of Chinese reviews using topic sentiment sentences. It is described as follows. (1) Identify the topic of a review using an n-gram matching approach. (2) Extract candidate topic sentiment sentences, compute the semantic similarity between a given sentence and the ascertained topic and meanwhile determine whether the sentence is subjective. (3) A certain number of these sentences are selected as representatives according to their semantic similarity value with relation to the topic. The average value of the representative topic sentiment sentences is calculated and taken as the sentiment polarity of a review.

Experiment results show that the proposed method is feasible and can achieve relatively high precision.

4 Topic Identification and Topic Sentiment Sentence Extraction

4.1 Topic Identification of Chinese Reviews

The topic of a Chinese review is presented as a set of strings $T=\{Wn_1, Wn_2, \ldots, Wn_i\}$, in which Wn_i refers to a word or several continuous words and n indicates the number of words in a Wn_i. The evaluation of whether any candidate Wn_i belongs to T depends on its position and frequency. Wn_i's position reflects its distribution degree $D(Wn_i)$: the more extensive Wn_i distributes in a review, the more likely it relates to the topic. Wn_i's frequency reflects its importance degree $I(Wn_i)$: the more times Wn_i appears in a review, the more likely it relates to the topic. Thus the degree of Wn_i belongs to T is defined as membership degree $C(Wn_i)$ and is measured by the formula:

$$C(Wn_i)= \alpha \cdot D(Wn_i) + \beta \cdot I(Wn_i) \quad (1)$$

In (1), $D(Wn_i)$ is determined by the number of paragraphs in which $D(Wn_i)$ appears and the total number of paragraphs of a text, $I(Wn_i)$ is the binary logarithm of the frequency of Wn_i in a text, α and β are the weighted coefficients to adjust the weights of $D(Wn_i)$ and $I(Wn_i)$.

In order to quickly obtain T, an n-gram matching based approach is applied according to the following algorithm.

(1) Strings separated by punctuations in the title and the main text are segmented and then stored respectively in queue T_q and B_q.

(2) For n=1 to m ($1 \leqslant m \leqslant$ the maximum length of T_q), take out a Wn_i from T_q successively and search it in B_q. If there is a Wn_i in B_q, then insert it into the index table $G=\{Wn_i$, position, frequency$\}$. When n=1, which means there is only one word in Wn_i, Wn_i should be a content word.

(3) Calculate the value of $C(Wn_i)$ for every Wn_i and add Wn_i to T if its $C(Wn_i)$ is greater than the threshold L_c. In this paper, we choose α=0.25, β=1, and L_c=0.8 according to our experience and experiment results.

4.2 The Extraction of Topic Sentiment Sentences

Topic sentiment sentences are essential in the analysis of the SP of reviews. Sentiment analysis based on topic sentiment sentences excludes unrelated sentiment and makes "homogeneous" sentiment computable. Topic sentiment sentences are extracted by 2 steps.

(1) Extract topic sentences from a review. Given a definite T, to extract topic sentences is actually the computing of semantic similarity of candidate sentences and the topic T. Factors that influence the similarity degree are the amount of identical words and strings, the length of identical words and strings, the position of a candidate sentence, semantic similarity of non-identical words.

The amount of identical words and strings. The more identical words or strings a candidate sentence has with T, the more likely they are similar in topic.

The length of identical words and strings. The longer an identical string (counted by word) shared by a candidate sentence and T, the more likely they are similar in topic.

The position of a candidate sentence. We hold that sentences in a paragraph are not in the same importance. As is the general common knowledge, the beginning and ending sentence in a paragraph are often more important than other sentences and thus receive more weights.

We use HowNet, a Chinese ontology, to compute the semantic similarity and assign each candidate sentence a value of similarity. If the similarity value of a sentence is greater than the threshold Ls, it is taken as a topic sentence.

(2) Extract sentiment sentences from topic sentences. We use a precompiled sentiment lexicon to roughly judge whether a sentence expresses sentiment or not.

Through the above procedures, the topic sentiment sentences in a Chinese review, each with a value indicating the distance in similarity with the topic, are extracted and arranged into order by value. We call them the set of candidate topic sentiment sentences.

5 Recognizing the Sentiment Polarity Based on Topic Sentiment Sentences

Based on section 3.3, in Chinese reviews, the higher similarity degree a topic sentiment sentence gets, the more likely it is a key sentence expressing the writers' basic sentiment orientation. But meanwhile, to avoid excessively relying on too few candidate topic sentiment sentences, more sentences are required to be analyzed to assure precision. Therefore, the number of sentences selected from the set of candidate topic sentiment sentences for final sentiment analysis is quite a question worth careful consideration.

Different Chinese reviews have different numbers of topic sentiment sentences. How many topic sentiment sentences a review has is determined by various factors. We find out, after an investigation of 560 Chinese reviews, that generally a Chinese review has not more than 7 topic sentiment sentences and the average number of that is about 4. Besides, long reviews tend to have rather more topic sentiment sentence. Thus we define that for any review the number of topic sentiment sentences which are needed to be analyzed as:

$$N(tss)=4\pm\gamma \tag{2}$$

γ in the above formula is an adjustable parameter which is determined by the ratio of the length of the analyzing review and the average length of a set of reference reviews.

N(tss) topic sentiment sentences with most weights are drawn from the set of candidate topic sentiment sentences and then are computed by a sentence-level sentiment analyzer. The average score of them is taken as O(r), i.e. the overall SP of a review.

$$O(r)=\frac{1}{N(tss)}\sum_{i=1}^{N(tss)}SP(tss_i) \tag{3}$$

We use a semantic approach in the sentence-level sentiment analyzer. For each sentence, a Chinese dependency parser is used to distinguish the dependency relations between language units, especially the probable relations between the topic words and the sentiment expressions, and the relations between the sentiment expressions and their modifiers. Making use of the syntactic information, the sentiment of a sentence is determined mainly by the sentiment expressions in it according to a precompiled sentiment lexicon. Meanwhile, the following factors are considered.

Negatives. Negatives inverse the sentiment of a sentence.

Connectors. Some connectors strengthen the original sentiment while others inverse the original sentiment.

Intensifiers. Intensifiers make the original sentiment more forcefully.

Discourse makers. In linguistics, a discourse marker is a word or phrase that is relatively syntax-independent and does not change the meaning of the sentence. However, discourse marker itself has certain semantic orientation: some of them are positive, some are negative and others are neutral. Thus discourse marker help recognize the SP in a sentence.

Punctuations. We pay special attention to question mark and exclamatory mark, especially when there is a negative in a question sentence.

6 Experiments and Results

6.1 Data

The data used in the experiment are Chinese current affairs reviews. They are originally collected from the website http://opinion.people.com.cn/ and then cleansed and stored as text. 400 texts are randomly selected from the reviews set. 3 annotators are trained and then instructed to annotate the topic sentiment sentences and judge the SP the 400 reviews individually. The following table shows the general information of the annotation result.

Table 2 General Information of the Annotation Results

Annotator	Pos. texts	Neg. texts	Other texts
1	87	302	11
2	93	298	9
3	88	288	14

Finally we get 370 texts (86 positive and 284 negative) totally agreed by the 3 annotators. We use them as the test reviews.

6.2 Resources

In order to perform an SP analysis, the following resources are required to use.

Sentiment Lexicon. We manually build up the sentiment lexicon. The words and phrases in the lexicon are mainly from three dictionaries: Positive Word Dictionary, Negative Word Dictionary and A Student's Positive and Negative Word Dictionary. We also get some words from HowNet Sentiment Dictionary and NTUSD. For each word or phrase, we give its part of speech, positive value and negative value. The positive and negative values of words and phrases are manually assigned by annotators according to human intuition.

Other lexicons. We collect as many negatives, connectors, intensifiers and discourse markers as we can and make them into different lexicons.

HowNet. As a Chinese ontology, HowNet is used to compute the semantic similarity of words.

LTP. LTP (Language Technology Platform developed by HIT) is a package of tools to process Chinese text, with a Chinese dependency parser in it. We use the dependency parser to perform a syntactic analysis of sentences.

CUCSeg. CUCSeg is a Chinese pos tagger. We use it to segment Chinese words.

6.3 Results of the Extraction of Topic Sentiment Sentences Experiment

The extraction of topic sentiment sentences is a vital task in this research. Annotators judge in the test reviews which sentences are topic sentiment sentences firstly and method described in 4.2 is applied and the result of which is evaluated. We adopt the commonly used precision, recall and F-measure to measure the result. It shows as follows.

Table 3 Result of the Extraction of topic Sentiment Sentences

Threshold	Precision	Recall	F_1
L_s=0.64	89.9	82.3	86.1
L_s=0.55	86.1	90.6	88.3
L_s=0.37	77.8	98.4	88.1

The above result shows we get a rather high precision and recall when L_s=0.55.

6.4 Results of Recognizing the SP of Chinese Reviews Experiment

We use precision to measure the result. Comparison is made among Turney's method (2002), Pang's SVM method (2002) and our method.

Table 4 Result of the SP of Chinese reviews

Method	Precision
Turney's	74.39
Pang's SVM	82.9
Ours	86.8

Compared to reports in earlier works, our approach achieves a relatively high precision.

We reexamine the 49 texts which are judged wrong, together with the 4 extracted representative topic sentiment sentences of each text. Error analysis shows that about 35% of errors are made by the topic identification step, about 49% of errors are made by the sentence-level sentiment analysis, about 4% of errors are made due to the faultiness of the sentiment lexicon. And the causes of other errors are to be explored.

7 Conclusion

We have presented a topic sentiment sentence-based approach to explore the overall sentiment polarity of Chinese reviews. Considering the features of Chinese reviews, we identify the topic of a review using an n-gram approach. To extract topic sentiment sentences, we compute the semantic similarity of a candidate sentence and the ascertained topic and meanwhile determine whether the sentence is subjective. A certain number of these sentences are selected as representatives according to their semantic similarity value with relation to the topic. The average value of the representative topic sentiment sentences is calculated and taken as the sentiment polarity of a review.

Error analysis indicates that to enhance the identification of topic, to build up a better sentence-level sentiment analyzer and to compile a better sentiment lexicon will help improve the final result.

Acknowledgements

We thank Prof. Ting Liu at Harbin Institute of Technology for sharing the LTP package and Mr. Zhendong Dong for sharing the HowNet. We highly appreciate that Hanbing Zhao, Nan Li and Yifu Xue have done a lot of heavy and tedious annotation work for this research.

References

Hu, M. and Liu, Bing. 2004. Mining and summarizing customer reviews. In Proceedings of the 10th ACM SIGKDD.168- 177.

Kim, S., Hovy E. 2004. Determining the Sentiment of Opinions. In Proceedings of COLING-04: the 20th International Conference on Computational Linguistics.

Lun-Wei Ku and Hsin-Hsi Chen 2007. Mining Opinions from the Web: Beyond Relevance Retrieval. Journal of American Society for Information Science and Technology, Special Issue on Mining Web Resources for Enhancing Information Retrieval, 58(12): 1838-1850.

Meng, F., L. Cai, B. Chen, and P. Wu. 2008. Research on the recognition of text valence. Journal of Chinese Computer Systems, 28(2007): 1-4.

Pang Bo, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of EMNLP, pages 79-86.

Turney, P. 2002. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. 417-424.

Wang, Gen and Jun Zhao. 2007. Sentence Sentiment Analysis Based on Multi-redundant-labeled CRFs. Journal of Chinese Information Processing, 21(5): 51-56.

Wang, C., Lu, J., Zhang, G. 2005. A semantic classification approach for online Product reviews. In Proceedings of the 2005 IEEE/WIC/ACM International Conference on web intelligence (Wl'5).

Wang, G. and Zhao, J. 2007. Sentence Sentiment Analysis Based on Multi-redundant-labeled CRFs. Journal of Chinese Information Processing. 5, 51-56.

Wang,C. J. Lu and G. Zhang. 2005. A semantic classification approach for online product reviews, In Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence. 276-279.

Wiebe J. 2000. Learning subjective adjectives from corpora. In Proceeding of the 17th National Conference on Artificial intelligence. Menlo Park, Calif. AAAI Press, 735-740.

Wiebe J., Riloff E.2005. Creating Subjective and Objective Sentence Classifiers from Unannotated Text. In: Proceedings of CICLING.

Wiebe J., Wilson T., BrueeR., Bell M. and Martin M.2004. Learning subjective language, Computational Linguistics, 30(3):277-308.

Yi J., Nasukawa T., Bunescu R., Niblack W.2003.Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In Proceeding of the Third IEEE International Conference on Data Mining.

Yu, H. and Hatzivassiloglou Vasileios.2003. Towards answering opinion questions. In Proceeding of EMNLP. 2003.

Zheng, W. and Q. Ye. 2005. Sentiment classification of Chinese traveler reviews by support vector machine algorithm. In The Third International Symposium on Intelligent Information Technology Application, 335-338.