Feature-level Sentiment Analysis for Chinese Product Reviews

Haiping Zhang, Zhengang Yu, Ming Xu, Yueling Shi
Dept. of Computer Science
Hangzhou Dianzi University
Hangzhou, China, 310018
zhanghp@hdu.edu.cn, kingchdu@gmail.com, mxu@hdu.edu.cn, xiaoshi@hdu.edu.cn

Abstract—**The sentiment analysis for English product reviews has been widely researched in recent years, followed with many important achievements. Due to the special language traits of Chinese, the study on Chinese product reviews is much more difficult than the former. In this work, we focus on the finer-grained sentiment analysis for Chinese product reviews, that is feature-level based sentiment analysis. We propose a hybrid method which combines association rules and point-wise mutual information to extract the product features, and then take advantage of the sentiment dictionary—HowNet to analyze the opinion orientation expressed on the product features. The experiment result obtained shows the effectiveness and efficiency of our approach.**

Keywords- feature extracting; feature-level sentiment analysis; Chinese product reviews

I. INTRODUCTION

The dramatic development of Web 2.0 technology in the past few years has greatly changed people's life styles especially people's shopping patterns have experienced a significant change. People who want to purchase something get used to viewing a large number of network reviews about the relative products or services beforehand, so that they can make reliable decisions. Mining valuable information from these product reviews not only provides some necessary purchase information for the potential consumers but also helps producers track the feedbacks of users on time. The feedback information contributes producers to maintain the good characteristics of products and improve the inferior products timely, and finally make them gain competitiveness in the near future.

However, the huge number of network comments also makes mining useful information to be a new challenge. It is hard and unrealistic for human to tackle all the reviews and classify them to positive or negative manually. Under this situation, automatic sentiment analysis technology has significant meaning. Via this approach, we can automatically extract the opinion which the author expressed on the product or its features.

Although researchers have made some important achievements on English product reviews, studies on the Chinese product reviews are still in its infancy. The great language differences existing between English and Chinese make the results obtained under English area can not be applied to Chinese comments directly. In this research, we mainly focus on the technology of feature extraction and its sentiment tendency analysis based on Chinese product reviews.

As is known, document-level and sentence-level based sentiment analysis are usually used to judge the overall evaluations of the special product. In order to obtain much more finer-grained and detailed product information, feature-level sentiment analysis is essential. There are two major topics involved in feature-level sentiment analysis:

(1) Extracting the product features which the user concerned. For example, in the sentence "这款手机的音质不错" (*The tone quality of this mobile phone is good*), "音质"(*tone quality*) is the product's feature.
(2) Analyzing the sentiment orientation of the product features. That is, classifying the author's evaluations expressed on the product feature into positive emotion, negative emotion or neutral. In the above sentence, the evaluation on the feature "音质" is positive.

The remaining of the paper is arranged as follows. In section II, we introduce the previous related researches on feature-level sentiment analysis. In section III, we will illustrate our method in detail. The corpus used in our experiment and the experiments result will be presented in section IV. Section V draws the conclusion of our research and then proposes the future work.

II. RELATED WORK

Sentiment analysis or opinion mining has attracted many researchers' interests. One of the most important issue of sentiment analysis is how to extract the opinions on the product features in the review, which is feature-level sentiment analysis.

The purpose of feature-level sentiment analysis for network product reviews is obtaining the opinions expressed on the product features. By this process, users can focus on the product features they most concerned.

A crucial aspect involved in feature-level sentiment analysis is product feature extraction. Many important achievements have been made in this area.

Hu and Liu in [1] proposed a method based on association rule to mine the product features. It took the

978-1-61284-840-2/11/$26.00 ©2011 IEEE

advantage of the phenomenon that people always use the same opinion word to describe the same feature in his/her comments. The author drew the conclusion that the frequent noun and noun phrase ("item sets") are probably to be product feature words. Popescu and Etzioni in [2] extract features by computing pair-wise mutual information between noun phrases and a set of metonymy discriminators associated with the product category. Different from the algorithms in [1] and [2] that utilized word co-occurrence, [3] applied a language model approach with the assumption that product features are mentioned more often in a product review than they are mentioned in generic English. Kobayashi et al. in their research [4] used a pattern mining method to extract the features. The patterns which mined from a large corpus using pattern mining are the relations between feature and opinion pairs. [5] proposed a bootstrapping iterative learning strategy to identify product features and opinion words. Besides, they exploited linguistic rules to find low frequent features and opinion words.

In the most recent researches, Qiu et al. in [6] proposed a double propagation method, which took advantage of certain syntactic relations of opinion words and product features. The extraction rules are designed based on different relations between opinion words and features, and among opinion words and features themselves. The relations are described by dependency grammar. It runs well for medium–size corpora. But for large corpora, this method may extract many nouns/noun phrases which are not features. Thus the precision of the method drops. To improve the results in [6], Zhang et al. in their research [7] proposed a novel method to mine product features, which consisted of two steps: feature extraction and feature ranking. In the feature extraction step, they utilized part-whole relation patterns and a "no" pattern to enhance the performance. And in the feature ranking step, they ranked feature candidates by feature importance.

Similar to general information extraction, the traditional rule-based method and statistical-based method can also be adopted to feature extraction. The popular models of statistical methods are Conditional Random Fields (CRF) [8], Maximum Entropy Models (ME) [9] and so on.

Other related researches on feature extraction mainly utilize topic modeling or clustering to capture features in reviews ([10][11]). However, topic modeling or clustering is only able to find some general or rough features, it is difficult to extract finer-grained or precise features.

As is described in [12], textual sentiment analysis can be classified into three types: document-based, sentence-based and feature-based sentiment analysis. Feature-based sentiment analysis is to determine the opinion tendency on the features. As it is most related to the practical applications, we put the emphasis on the feature-based sentiment analysis. The approaches used in the document-based and sentence-based sentiment analysis can apply here too. An efficiency way to conduct feature sentiment analysis is called lexicon-based approach [13][14]. Lexicon-based methods are usually involved with sentiment lexicons and some linguistic features. The sentiment lexicons comprise of several seed words or just a big dictionary. In this research, we use the sentiment dictionary provided by HowNet [1] to conduct sentiment analysis.

III. Our Approach

In this section, we will present our approach in detail. Firstly, we use the Apriori association mining rules to extract the candidate product features, then adjust the orders of some candidate product feature words. Finally, we use point-wise mutual information (PMI) methods to filter feature words so as to obtain the meaningful product feature words. After get the feature words, we use the sentiment dictionary provided by HowNet to analyze the sentiments expressed on the features.

A. Feature Words Extacting Based on Apriori

The algorithm of Apriori association rules was first proposed by Agrawal and Srikant [15]. It is one of the most classical association rules used to mine frequent item sets.

The core part of the Apriori algorithm is the two-stage recursive process:

(1) Identify all the frequent item sets in the transaction database whose support are greater than or equal to the user defined threshold.

(2) Use the frequent item sets produced above to construct the rules that meet the minimum confidence.

Fig. 1 shows the Apriori algorithm in detail. Here, D is the transaction database, C_k is the candidate item sets and the function apriori_gen() is used to generate candidate item sets of length k from item sets of length k-1. Then pruning the candidates those who have infrequent sub pattern.

In this research, we adopt the Apriori algorithm to extract product feature words. The product feature words can be seemed as frequent item sets to some extends. It is no necessary to construct the associate rules further.

L_1 = {frequent 1-item sets};
for (k = 2; $L_{k-1} \neq \varnothing$; k++) **do begin**
    C_k = apriori_gen(L_{k-1});
    for any transaction $t \in D$ **do begin**
        C_t = subset (C_k, t);
        for any candidate item $c \in C_t$ **do**
            c.count++;
    end;
    $L_k = \{c \in C_k \mid c.count \geqslant \text{min\_supp}\}$
end;

Figure 1. The Apriori algorithm

B. Feature Words' Order Adjustment

In some cases, the candidate feature words extracted via Apriori algorithm may not have the normal sequence. This is because the procedure of merging previous item sets into new candidate feature words ignores the word's semantic.

[1] http://www.keenage.com/html/c_index.html

For example, the candidate feature word "分辨率 屏幕"(*Resolution Screen*) is generated from the frequent 1-item "分辨率"(*Resolution*) and "屏幕"(*Screen*). However, the normal sequence of these two words after combined should to be "屏幕分辨率" (*Screen Resolution*).

In this research, we use a statistical method to adjust the candidate feature words' unreasonable order.

Given the candidate item set f, and f is composed of 1-item sets: $w_1, w_2, \ldots, w_k$. The position of w_i in the comment sentence is noted as s_i. Then re-arrange the word sequence in f by ascending order based on w_i's value. Symbol f' means the new sequence after re-arranging, and the position of w_i in f' is noted as s_i'. Finally, we identify all product reviews which containing the feature words of candidate item set f, then sum each feature's s_i' value so as to get the result $pos(w_i)$. The value of $pos(w_i)$ determines the candidate features' final sequence.

For instance, suppose the candidate item set f = "分辨率 屏幕", thus w_1= "分辨率", w_2 = "屏幕". Given the product review: "这个手机的屏幕很大分辨率也很高"(*This mobile phone's screen is big and its resolution is also very high*). After Chinese segmentation, it turns to be: "这个/r 手机/n 的/u 屏幕/n 很/d 大/a 分辨率/n 也/d 很/d 高/a". As can be seen, w_1 and w_2 appears in this review, and $s_1 = 7$, $s_2 = 4$. Then according to the value of s_i to adjust the sequence of candidate item set f. We can get the new feature set f' = "屏幕 分辨率" as $s_2 = 4 < s_1 = 7$. Therefore, the position of word w_1 in f' is $s_1' = 2$, and w_2 in f' is $s_2' = 1$.

Suppose there are 4 reviews contain the item set f. Among these reviews, "分辨率" appears before the word "屏幕" only in one review, which means $s_1' = 1$, $s_2' = 2$. In the left three reviews, "分辨率" appears behind the word "屏幕", thus $s_1' = 2$, $s_2' = 1$. As a result, $pos(w_1$= "分辨率") $= \sum_4 s_1' =$ 1+2+2+2 = 7 and $pos(w_2$ = "屏幕") $= \sum_4 s_2' =$ 2+1+1+1 = 5. Hence, after processing all of the reviews which contain the feature f, the final sequence of f is adjusted to be "屏幕分辨率" as 5 < 7.

C. Feature Words' Filtering Based on PMI

Since the algorithm is calculated based on the noun or noun phrase's frequency, the "product feature" mined in this way may have no practical meanings. The senseless features will become noises and they will affect the efficiency and effect of the extracting procedure. Therefore, it is necessary to do further job to filter senseless product feature words extracted by the Apriori algorithm after the order adjusting.

To solve this problem, we take advantage of the Google search engine combined with the approach which used point-wise mutual information (PMI) proposed by Turney [16]. PMI is used to mine the semantic correlations between the candidate product feature words and product words. By calculating the PMI value, then a proper threshold will be obtained. Thus, some candidate product feature words will be abandoned if their PMI value are below the threshold.

The definition of PMI is given in Equation 1:

$$PMI(word_1, word_2) = \log_2 \frac{p(word_1 \,\&\, word_2)}{p(word_1)\,p(word_2)} \tag{1}$$

Here, $p(word_1$ & $word_2)$ is the co-occurrence probability of $word_1$ and $word_2$, and $p(word_1)p(word_2)$ gives the probability that the two words co-occurring if they are statistically independent. The ratio between $p(word_1$ & $word_2)$ and $p(word_1)p(word_2)$ is thus a measure of the degree of statistical dependence between the words. The log of this ratio is the amount of information that we acquire about the presence of one of the words when we observe the other.

Based on the Equation (1), we improved the formulation to calculate the PMI value between the product feature and product itself:

$$PMI(product, feature) = \log_2 \frac{hit(product \,\&\, feature)}{\sqrt{hit(product)hit(feature)}} \tag{2}$$

In (2), $hit(x)$ is the number of relative links returned by the search engine when query some keywords. For example, *hit*("*product A*") is the number of links returned when search the keyword "product A". *hit*(*product* & *feature*) stands for the link numbers returned when use the "product" and "feature" together as the query keywords. Due to the actual value of *hit*(*product*)*hit*(*feature*) may be very large, here we extract the square root value of the denominator so as to be suitable for computer handling.

Equation (2) illustrates that the larger the PMI value is, which means the more sufficient mutual information of candidate product features and product itself, the more likely these candidate product features to be the real features of the product. Suppose the threshold is α and if $PMI(product, feature) \geq \alpha$, then we can draw the conclusion in statistically that the "*feature*" is the real part of the "*product*". Otherwise, the "*feature*" will be removed from the candidate feature words.

D. Feature-level Sentiment Analysis Based on HowNet Sentiment Dictionary

The feature-level sentiment analysis task can be divided into two sub tasks:

(1) Identifying product features;
(2) Determining whether the opinions on the features are positive, negative or neutral.

In this section, we focus on the second sub task as the product features have been identified through above steps. In our methodology, we utilize unsupervised learning method based on sentiment dictionary to identify the opinions expressed on the features. The sentiment dictionary used is provided by the HowNet[2]. In this research, we only make use of the Chinese evaluation and sentiment words in this dictionary.

The processing procedure is shown in Fig. 2. Detailed process of feature sentiment analysis is described as follows:

[2] http://www.keenage.com/html/c_index.html

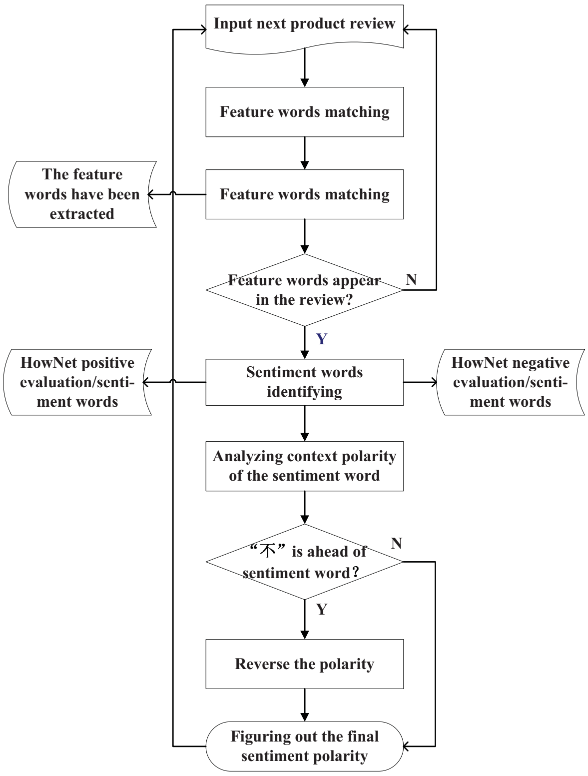

Figure 2. The processing steps in determing feature's opinion

(1) Word segmentation and POS tagging to the reviews. The tool used here is ICTCLAS system[3] which is developed by the Institute of Computing Technology of Chinese Academy of Sciences.
(2) Feature words matching. It takes advantage of the product feature words obtained through the Apriori association rule mining algorithm and PMI filtering algorithm to identify the product feature words appeared in the review. Potential feature mining is not be taken into account in this research. Hence, if there is no obvious product feature word found in the review, then go to step (1) to process the next review.
(3) Classifying sentiment word's polarity. To the reviews which contains any product feature word, extracting a adjective word before or behind it as the opinion word using the result of word segmentation and POS tagging. Then determining the opinion word appears in which category of the sentiment dictionary provided by HowNet. That is, the opinion word will be classified into positive if it falls in the scope of positive evaluation words or positive sentiment words. Otherwise, it will be grouped into negative.
(4) Analyzing sentiment word's context polarity. As is known, the sentiment orientation will be opposite to the original one in most cases if negative word appears ahead. We utilize the word "不"(no) as the prefix negative word which appears frequently in Chinese product reviews.
(5) Figuring out the final sentiment polarity of the product feature. Considering there may be different opinions expressed on the same feature, it is necessary to sum up the various views.

IV. EXPERIMENTS

A. Data Sets

We select 6 different products reviews which are collected from the website IT168[4] as the experiment corpus. In the corpus, there are 3 different brands, and each brand has 2 similar type products. The 6 different products are including: *Lenovo G450M-TFO* laptop, *Lenovo IdeaPad* laptop, *Panasonic EZ35* digital camera, *Panasonic LX3* digital camera, *Nokia 5230* mobile phone and *Nokia 5800 XM* mobile phone. The number of each type product review is 100.

B. Experiment of Feature Words Extacting Based on Apriori

In this experiment, suppose one product type to be the training data, then the other product with the same brand is taken as the testing data. The arrangement of training data and testing data is shown in Table I.

TABLE I. THE TRAINING DATA AND TESTING DATA

The reviews in training sets	The reviews in testing sets
Nokia 5800XM mobile phone	Nokia 5230 mobile phone
Lenovo IdeaPad laptop	Lenovo G450M-TFO laptop
Panasonic LX3 digital camera	Panasonic EZ35 digital camera

Generally speaking, the product feature words are noun or noun phrases. To avoid many senseless words and improve the efficiency of the Apriori algorithm, only the noun or noun phrase are left in the experiment.

Through times of verification in the training sets, we find that the feature words in 3-item set or above 3 extracted based on the Apriori algorithm are almost nonsense words. Therefore, only the 1-item set and the 2-item set are extracted so as to decrease the runtime consuming of the Apriori algorithm.

TABLE II. THE RESULTS OF APRIORI EXTRACTIONG

Product	Recall	Precision	F1 value
Panasonic EZ35 digital camera	62.5%	71.4%	66.7%
Lenovo G450M-TFO laptop	70.4%	73.1%	71.7%
Nokia 5230 mobile phone	80.8%	72.4%	76.4%

[3] http://ictclas.org/Download.html

[4] http://www.it168.com

The performance of the experiment depends greatly on the support value of the Apriori. We adopt the support value as the threshold which makes the F1 value gain the optimum in the training set. Here, the support threshold in *Panasonic EZ35* is 5, in *Lenovo G450M-TFO* is 6, and in *Nokia 5230* is 7. The result is shown in Table II. It shows that the average F1 value is 71.6%.

C. Experiment of Feature Words' Order Adjustment

The consequence of the feature words' order adjustment is listed in Table III. It can be seen that the average F1 value of the 3 product after adjustment is 75.1% comparing to 71.6% before adjustment. The adjustment step improves the overall performance.

TABLE III. THE RESULTS OF FEATURE WORD'S ORDER ADJUSTMENT

Product	Recall	Precision	F1 value
Panasonic EZ35 digital camera	68.8%	78.6%	73.4%
Lenovo G450M-TFO laptop	74.1%	76.9%	75.4%
Nokia 5230 mobile phone	80.8%	72.4%	76.4%

D. Experiment of Feature Words' Filtering Based on PMI

The purpose of this experiment is to identify the PMI threshold by use of the Google search engine so that filtering the nonsense feature words more effectively. The PMI threshold value is obtained through times of experiments in the training sets. Here, the PMI threshold in *Panasonic EZ35* is -3.02, in *Lenovo G450M-TFO* is -4.61, and in *Nokia 5230* is -3.57. The performance after PMI filtering is illustrated in Table IV. As is shown in the table, the average F1 value of the 3 product is 75.4%. Comparing to the initial result 71.6% gained by the Apriori algorithm, the performance after series processing has been enhanced obviously.

TABLE IV. THE RESULTS OF FEATURE WORD'S FILTERING

Product	Recall	Precision	F1 value
Panasonic EZ35 digital camera	68.8%	81.5%	74.6%
Lenovo G450M-TFO laptop	70.4%	82.6%	76.0%
Nokia 5230 mobile phone	76.9%	74.1%	75.5%

E. Experiment of Feature-level Sentiment Analysis Based on HowNet Sentiment Dictionary

In this experiment, we take advantage of the sentiment dictionary provided by HowNet to analysis the opinion orientation of the product feature. The mining result is presented in the form of ("product feature", "semantic polarity"), such as (*Cannon lens*, *positive*). Table V gives the average performance of the experiment.

TABLE V. THE RESULTS OF FEATURE-LEVEL SENTIMENT ANALYSIS

Product	Recall	Precision	F1 value
Panasonic EZ35 digital camera	67.1%	79.2%	72.6%
Lenovo G450M-TFO laptop	70.4%	81.3%	75.5%
Nokia 5230 mobile phone	69.2%	83.5%	75.7%

The experiment results demonstrate that the simplicity approach proposed in this paper performs well on the Chinese product reviews. But the recall is low. The reasons lead to this phenomenon are mainly due to:

- Some reviews contain the feature words but have no opinion words modifying the features.
- Some opinion words are ambiguity in different situations. This will make the approach in this paper miss the right sentiment words expressed on the product feature.
- Lacking of semantic analysis. A case in point is that we ignored the potential feature word mining such as in the review "*The G2 mobile phone is so dear*", the potential feature is "price".
- The network words appeared in the product reviews changes so quickly that the sentiment dictionary can hardly cover them completely and timely.

V. CONCLUSIONS AND FURTURE WORK

The approach we proposed in this paper takes advantage of the Apriori association rule algorithm to extract candidate product feature words. After adjusting the order of the candidate features and then filtering nonsense features based on the improved PMI, the performance of products feature extraction has been improved significantly.

In most cases, customers need more detailed information about the products or their features. This research utilize the sentiment dictionary of HowNet to conduct feature-level analysis since the feature words have been obtained. The approach is simplicity and efficiency in Chinese product processing. Due to lacking of deeply semantic analysis of on the reviews, the overall performance of the method needs to improve.

In the near future, we will put more emphasis on the feature extracting. Some classical models such as Conditional Random Field(CRF), Maximum Entropy(ME) and other machining methods will be utilized comprehensively. In the feature-level analysis phase, we will employ the semi-supervise learning method which combines the advantages of supervise learning and unsupervised learning methods. This is extremely adapted to the situation when there are less annotated data and lots of un-annotated data. Besides, the semantic and sentence rules will also be taken into consideration.

ACKNOWLEDGMENT

This paper is supported by the State Key Program of Major Science and Technology (Priority Topics) of Zhejiang Province, China under Grant No 2010C11050.

REFERENCES

[1] M. Hu and B. Liu. Mining and summarizing customer reviews. Proceedings of SIGKDD. 2004.

[2] A. M. Popescu and O. Etzioni. Extracting product features and opinions from reviews. Proceedings of EMNLP. 2005.

[3] C. Scaffidi, K. Bierhoff, E. Chang, M. Felker, H. Ng and C. Jin. Red opal: Product-feature Scoring from Reviews. Proceedings of EC. 2007.

[4] N. Kobayashi, I. Kentaro and Y. Matsumoto. Extracting Aspect-Evaluation and Aspect-of Relations in Opinion Mining. In Proceedings of EMNLP. 2007.

[5] B. Wang, and H. F. Wang. Bootstrapping both Product Features and Opinion Words from Chinese Customer Reviews with Cross-Inducing. Proceedings of IJCNLP. 2008.

[6] G. Qiu, B. Liu, J. Bu and C. Chen. Expanding Domain Sentiment Lexicon through Double Propagation. Proceedings of IJCAI. 2009.

[7] L. Zhang, S.H. Lim, B. Liu, and E. O'Brien-Strain. Extracting and Ranking Product Features in Opinion Documents. Proceedings of COLING. 2010.

[8] J. Lafferty, A. McCallum and F. Pereira. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. Proceedings of ICML. 2001.

[9] H. L. Chieu, and H.-T. Ng. Name Entity Recognition: a Maximum Entropy Approach Using Global Information. Proceedings of the 6th Workshop on Very Large Corpora. 2002.

[10] Q. Z. Mei, X. Ling, M. Wondra, H. Su and C.X. Zhai. Topic Sentiment Mixture: Modeling Facets and Opinions in Weblogs. Proceedings of WWW. 2007.

[11] Q. Su, X. Y. Xu., H. L. Guo, Z.L. Guo, X. Wu, X.X. Zhang, B. Swen and Z. Su. Hidden Sentiment Association in Chinese Web Opinion Mining. Proceedings of WWW. 2008.

[12] B. Liu. Sentiment analysis and subjectivity. Handbook of Natural Language Processing, second edition. 2010.

[13] M. Hu and B. Liu. Mining and summarizing customer reviews. Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD). 2004.

[14] X. Ding, B. Liu, and P. S. Yu. A holistic lexicon-based approach to opinion mining. Proceedings of the Conference on Web Search and Web Data Mining (WSDM). 2008.

[15] R.Agrawal, and R.Srikant. Fast Algorithms for Mining Association Rules. In Proceedings of VLDB.1994.

[16] P. D. Turney. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. Proceedings of ACL. 2002.