International Journal of
Information
and
Management
Sciences

International Journal of Information and Management Sciences
29 (2018), 171-189. DOI:10.6186/IJIMS.201806_29(2).0002

PRSentiMiner: A Product Review Mining System for Sentiment Analysis of Chinese Microblogging

Wei Shi

Huzhou University

Abstract

For massive product reviews in Chinese microblogging (Weibo), we propose an opinion aware framework- PRSentiMiner- to conduct sentiment analysis for product reviews in Chinese microblogging based on fuzzy opinion words ontology. Normalized TFIDF weighting scheme is applied to extract the most informative noun patterns to represent the product features, and fuzzy opinion words ontology is constructed based on HowNet. Then we use the established fuzzy opinion words ontology and BMI (Balanced Mutual Information) method to extract evaluation words for product features, we also build the influence analysis method of microblogging text and quantitative computing method of emotional semantic factors in microblogging text, and we finally construct the sentiment computing method of product reviews in Chinese microblogging and give the concrete applying steps of this method. The results show that the PRSentiMiner outperforms baseline methods in the parties and have a good application through experiments.

Keywords: Chinese Microblogging, product reviews, sentiment analysis, opinion words ontology.

1. Introduction

In China, the counterpart of Twitter is Weibo, which is the word for word translation of micro (wei) and blogging (bo) in Chinese. With the popularity of Chinese microblogging, Internet users will publish their love or disgust for the features of the popular products. Online product reviews begin to appear quietly on the Chinese microblogging platform, which were originally only in the electronic commerce website or professional review network. With its high popularity and real-time, Chinese microblogging platform expand the recognition of the products. Businesses begin to attach importance to the marketing based on Chinese microblogging platform and use micro groups or micro review to expand the impact of the products. However existing researches on sentiment analysis of Chinese microblogging mainly focus on public opinion and are lack of mining product reviews on Chinese microblogging. Moreover, the processing of feature - opinion and semantic analysis in microblogging product reviews are not deep enough. In view of this, this paper proposes a new analysis system of Chinese microblogging product review for improving the mining level of product reviews in Chinese microblogging.

Product reviews mining based on microblogging has several differences compared to product reviews mining from professional reviews platform: (1) Due to the certain reality relationship between microblogging owners and fans, the reviews are more pertinent and effective; (2) As microblogging owners have different grades, product reviews are influenced by different reviewers; (3) Chinese microblogging text is short text structure, the review mining has its particularity; (4) The colloquial of Chinese microblogging reviews is common, which makes the emotional semantic analysis and feature mining of Chinese microblogging text more complex.

2. Literature Reviews

2.1. Sentiment analysis for product reviews

Sentiment analysis, also known as review mining or opinion mining, is referred to find consumers' attitude towards the commodities through the automatic analysis of the textual content of product reviews(see Turney [23]). In recent years, some scholars start the research work on the sentiment analysis of product reviews from different perspectives: Yao and Lou [26] described how to determine the semantic orientation(polarity) of sentiment words in a Chinese sentence in opinion mining of car reviews, they considered not only the calculation of the prior polarity for sentiment words, but also the calculation of the dynamic polarity for sentiment words by analyzing the context of those, they improved the accuracy of sentiment analysis of car reviews. Meng et al. [8] presented an opinion analysis method of cross-domain product reviews based on feature transformation, this proposed method built the relevance of domain dependent words between source domain and target domain via domain independent words so that it can transfer acknowledge from the source domain to the target domain, the results were higher than Baseline algorithm. Shi [19] and Shi [17] studied the product reviews mining from the perspective of CRFs and syntax tree model, and obtained some accuracy. Morinaga et al. [9] collected sentences regarding target products using a general search engine, and they then extracted opinions from them and attached three labels to each opinion by using the rules as the product's reputation. Popescu and Etzioni [14] introduced OPINE, an unsupervised information extraction system which mined reviews in order to build a model of important product features, their evaluation by reviewers, and their relative quality across products. Compared to previous work, OPINE achieved 22% higher precision (with only 3% lower recall) on the feature extraction task.

Product feature extraction is the basis of sentiment analysis of product reviews. Most of the existing methods of feature extraction were based on statistical theory, the basic idea was to identify the frequency of the noun in the specific domain based on the predefined rules (see Zhan et al. [28]; Moghaddam and Ester [10]). The advantages of this method were efficient and simple, but it was easy to overlook these feature words which were low frequency. As product features in opinion mining were mainly extracted from a single corpus, some researchers extracted features from two corpora via intrinsic and extrinsic domain relevance, their method improved the accuracy of feature extraction (see Zhan et al.[27]).

2.2. Sentiment analysis for microblogging

There have been more and more scholars who are very interested in studying microblogging as a subject for sentiment analysis and opinion mining. Various researchers have conducted sentiment analyses using Twitter as a corpus from different perspectives (see Alec et al. [1]; Johan et al. [5];Thelwall et al. [22]; Pfeiffer and Tourte [13]; Svetlana et al. [20]). In China, since its appearance in late 2009 as a counterpart of Twitter in the name of weibo (the Chinese translation of microblogging), microblogging had quickly gained more and more popularity as a form of social media favorably used by Chinese online communities. After 2012, the sentiment analysis of Chinese microblogging has emerged, Liu and Liu [7] used three machine learning algorithms, three kinds of feature selection methods and three feature weight methods to study the sentiment classification for Chinese microblogging. Combining the three factors the conclusion could be drawn that the performance of a combination of SVM, IG, and TF-IDF was best. Wei and Wang [25] proposed a method of pruning the syntax tree to implement the topic-dependent sentiment analysis ·They used the convolution kernel of Support Vector Machine(SVM) to obtain the structured information from syntax tree and adopted the topic-dependent syntax pruning according to the domain ontology and syntactic paths library. Experimental results on two corpora with different topics showed that the accuracy could reach 86.6% and 86.0% ·

As most user reviews appear in the professional product review sites or well-known electronic commerce platform, and therefore, scholars take the product reviews in these platforms as the research object in previous studies. Research on product reviews based on Chinese microblogging platform has few research results: Tang and Wang [21] built a Chinese microblogging product review mining model, the model made the reviews as two classes to study by different polarity, and mined the users concerned advantages and disadvantages of product features with the corresponding view by statistical analysis methods. However, there are some shortcomings in this study: feature selection is completed by artificial screening; the method of evaluation words selection and feature - opinion extraction should also be further improved to achieve the better matching number.

2.3. Research gaps and questions

Although many efforts are being spent on sentiment analysis for product reviews, current studies still have limitations: (1) Most of them focus on sentiment orientation of a review as a whole, thus lacking sentiment analysis for each product's features in reviews and association analysis for product features and opinion words. (2) In sentiment analysis for Chinese product reviews, the semantic emotional factors are considered too little, and

only the influence of the negative words and the degree words are generally considered.
(3) Being lack of product reviews mining based on Chinese microblogging platform.

Based on literature review, we pose the following research questions:

(1) What is the framework of features extraction, feature-opinion extraction and updating from product reviews in Chinese microblogging?
(2) What are the text-related factors that possibly affect sentiment classification in Chinese microblogging? How can we evaluate the degree to which these factors influence sentiment classification?
(3) How to get the sentiment polarity and intensity of 'feature-opinion' and product reviews combined with the influence factors of Chinese microblogging?

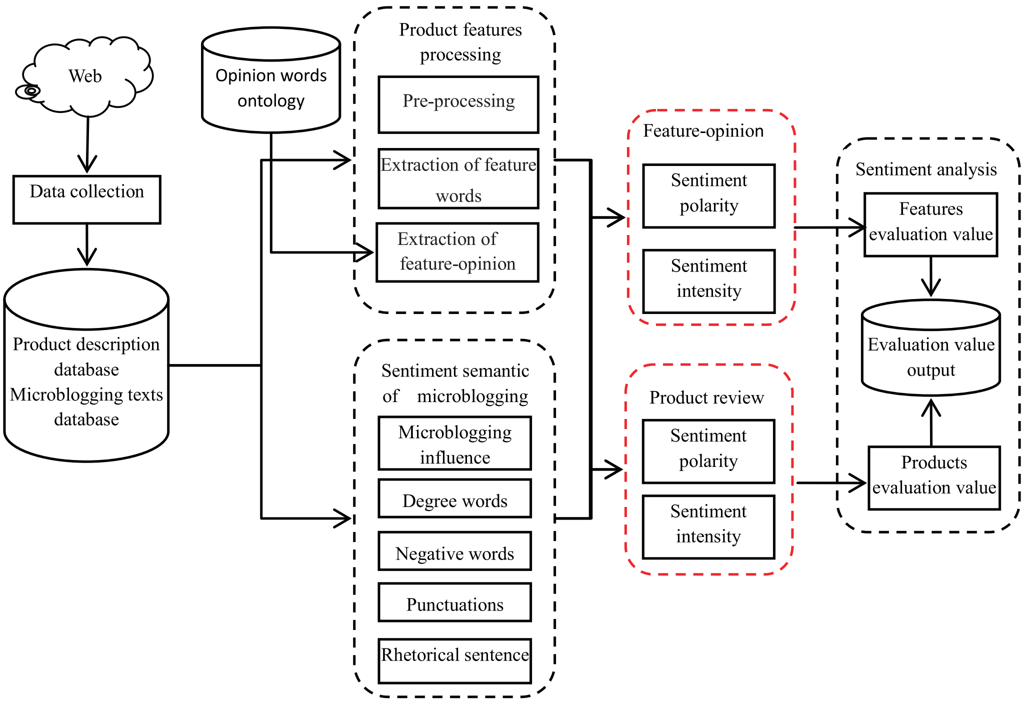

Figure 1: PRSentiMiner: A framework for sentiment analysis for product reviews in Chinese microblogging.

3. Research Design

We describe a framework, PRSentiMiner, illustrated as Figure 1 for analyzing sentiments of product reviews in Chinese microblogging, the steps are as follows: (1) Product features: feature words extraction based on TFIDF and feature - opinion extraction based on BMI; (2) Construction of opinion word thesaurus; (3) Quantitative analysis of Chinese microblogging emotional factors: including microblogging text influence, degree words and negative words, punctuation, rhetorical sentences; (4) Calculation of sentiment polarity and intensity of feature - opinion; (5) Calculation of sentiment polarity and intensity of product reviews; (6) Results output of product reviews Sentiment analysis.

3.1. Extracting product features based on TFIDF

Using the APIs provided by e-Commerce sites, our system can retrieve the product descriptions of a set of products under a product category. Crawler programs and the APIs of Google can also be used to collect product descriptions for each relevant product. Standard document-processing procedures (Such as word segmentation, remove disable words, POS tagging, POS filtering, etc.) are applied to each product description document retrieved from the Web. In this paper, we have prepared the crawler program to grab the product description and description from Tencent digital (http://digi.tech.qq.com/) and other products service web site. Normalized TFIDF weighting scheme (see Salton [16]) is applied to extract the most informative noun patterns to represent the product features of a particular product p_i. For each product description document d, the weight $w(f_i, d) \in [0, 1]$ of a product feature f_i is derived by:

$$w(f_i, d) = \frac{\left(0.5 + 0.5\frac{tf(f_i)}{\max tf(d)}\right) \times \log_2 \frac{N}{df(f_i)}}{\sqrt{\sum_{f_j \in d} \left[0.5 + 0.5\frac{tf(f_j)}{\max tf(d)} \times \log_2 \frac{N}{df(f_j)}\right]^2}} \tag{3.1}$$

where $tf(f_i)$ represents the word frequency of the product feature f_i in the document d, and $df(f_i)$ represents the document frequency of the f_i in all the product descriptions retrieved from the Web (i.e., the number of product feature f_i appearing in the product descriptions). The function $\mathrm{Max}\, tf(d)$ represents the maximum word frequency obtained from the product description d. $N = |D_{p_i}|$ is the number of documents retrieved for the product p_i. Finally, the fuzzy membership degree of a product feature for a product can be approximately represented by the average TFIDF weight of the product feature f_i in all product description documents D_{p_i}:

$$\mu_{R_{\mathrm{NTAX}}}(f_i, p_i) = \frac{\sum_{d \in D_{P_i}} w(f_i, d)}{|D_{P_i}|}. \tag{3.2}$$

In the opinion mining of user reviews, the product feature f_i is extracted from each review d by using the above method. A product feature weight vector $\vec{d}$ is extracted from the review d. After the opinion mining process, the product feature weight vector extracted from the user reviews D_{Rev} is applied to update $\mu_{R_{\mathrm{NTAX}}}(f_i, p_i)$ by using an approach similar to the Rocchio learning method (see Raymond et al. [15]):

$$\vec{F}_{i,i+1} = \alpha \times \vec{F}_{i,i} + \frac{\beta}{|D_{\mathrm{Rev}}|} \sum_{d \in D_{\mathrm{Rev}}} \frac{\vec{d}}{\|\vec{d}\|} \tag{3.3}$$

where
$$\vec{F}_{i,t+1} = \langle \mu_{R_{\mathrm{NTAX}}}(f_1, p_i), \mu_{R_{\mathrm{NTAX}}}(f_2, p_i), \ldots, \mu_{R_{\mathrm{NTAX}}}(f_n, p_i)\rangle$$
represents the original vector of product feature weight of product p_i. The parameter $\alpha = \beta = 0.5$ is applied to our experiment; $\|\vec{d}\|$ is the norm (length) of the product feature vector $\vec{d}$. After Rocchio learning, the updated product feature sets and their weights $\vec{F}_{i,t+1}$ are obtained. At the same time, new product features may be added, and the weakest product features will be removed after Rocchio learning. Parameters $\bar{w}_f = 50$ are used to control the number of retained product feature vectors $\vec{F}_{i,t+1}$.

3.2 Fuzzy representation of opinion words

As opinion words are often used in an irregular way, their sentiment class and membership are determined manually according to HowNet (see [3]). HowNet is an online common-sense knowledge base (www.keenage.com). From the sentiment vocabulary of HowNet, we extract opinion words including 3730 words in positive and 3116 words in negative, and then establish a fuzzy opinion words thesaurus. The sentiment orientation for opinion words would vary in some contexts. For example, '这款液晶电视屏幕很薄。' (English: This type of LCD TV screen is pretty thin.), and '这件外套太薄了, 不适合冬天穿。' (English: The coat is too thin to suitable for wearing in winter.) The opinion word '薄' (thin) has different sentiment orientations in the two contexts above. The former is clearly positive, while the latter is negative. For those opinion words that differ in sentiment orientations, we list all possible sentiment classes.

To solve some uncertainties in emotion expressions, we use fuzzy theory to quantify sentiment orientation of opinion words. Nine-point Likert items are adopted with nine categories of '-XL' , ' -L' , '-M' , '-S' , 'Z' , '+S' , '+M' , '+L' and '+XL' for measuring positive or negative sentiment orientation of opinion words. 'S' means small, 'M' for medium, 'L' for large, 'XL' for extra large and 'Z' for neutral. '+' and '-' denote positive and negative, respectively. We give each 9-point Likert item a values as follows: -1, -0.75, -0.50, -0.25, 0, 0.25, 0.5, 0.75 and 1. Sentiment class of opinion words includes good and bad denoted as G or B, each having an opinion value of no less than 0 or less than 0.Such as $MG = (G; 0.50)$.

Fuzzy representation of opinion words is described by a 3-triple, denoted as $FO = (B, R, E)$, where B is the basic information of vocabulary, including ID, term, the term in English, part of speech (POS), entry and version information; R represents the synonymous relationship between terms; E is the sentiment class and intensity of vocabulary.

An example of opinion word ontology is given as below:

$$FO = ((9;\ \text{不错};\ \text{not bad};\ \text{adv}),\ (\text{不坏};\ \text{不赖}), (G; 0.50))$$

To ensure the accuracy of annotating intensity of opinion words, we first ask three linguist experts to annotate the sentiment orientations for the same collection of opinion words. Only if more than two experts are in agreement about the sentiment orientations, the sentiment intensity will be accepted. Otherwise, another expert is required to make revisions until an agreement is achieved.

In this paper, we employ kappa to measure the agreements of manual annotation. The equation for kappa is:

$$k = \frac{P(A) - P(E)}{1 - P(E)} \tag{3.4}$$

where $P(A)$ is the relative observed agreement among raters, and $P(E)$ is the hypothetical probability of chance agreement, using the observed data to calculate the probabilities of each observer randomly saying each category. This paper measures the agreement among the manual annotation of opinion words from two sides: (1) sentiment class (G or B); (2) sentiment intensity (-XL, -L, -M, -S, Z, +S, +M, +L and +XL). We calculate Cohen's kappa by using formula (4) and obtain Table 1.

Table 1: The agreement of opinion words annotation.

	Sentiment class (1)	Sentiment intensity (2)
Opinion words	0.95	0.75

Table 2 gives an example of opinion words corpus.

Table 2: An example of annotation for the sentiment of opinion words.

Corpus example	Opinion base	
这款手机价位挺理想, 操作比较舒服, 显示屏一般, 性能有点缺陷, 待机时间无法忍受, 外观独一无二。 (For this type of cell phone, the price is relatively ideal. It is more convenient to use. The screen is not bad. performance has a little bit deficiencies. The standby time is intolerable, and the appearance is unique.)	Opinion words	Sentiment value
	理想(ideal)	LG
	舒服 (convenient)	MG
	一般 (not bad)	Z
	缺陷 (defective)	MB
	无法忍受 (intolerable)	XLB
	独一无二 (unique)	XLG

We removed those words that are not widely used in HowNet and then added some cyber words emerging over the Internet. The final fuzzy opinion words thesaurus includes 6862 opinion words, where 3715 words belong to G, and 3147 words belong to B. It could meet sentiment analysis for Chinese microblogging product reviews.

3.3. Extracting the pair 'feature-opinion' based on BMI

Similar to product feature extraction, a set of consumer reviews is used to build the non-taxonomic relations between opinion words and product features via an offline learning process. We have constructed the fuzzy opinion thesaurus, obtained 6862 opinion words. Here, we will associate the opinion words with the corresponding product features to form 'feature-opinion' for the particular product. Our system takes into account the simple feature sentence boundary when a sentence contains only one type of feature words, we call it the simple feature sentence. If a sentence contains more than two types of feature words, we need to find the nearest comma located before the current feature word and separate the sentence by the comma. The review thus is divided into two or more clauses, each of which is a simple feature sentence (see Pan and Lin [12]). The text window is set to 5, the meaning of 5 is that the opinion words which is 5 characters range from the product feature are extracted. Within the text window, opinion words which have the highest membership with the product feature will be detected. By means of Balanced Mutual Information (BMI) (see Lau et al. [6]), we can identify the opinion words that are highly associated with a given product feature, the advantage of the BMI measure is that it takes into account both term presence and term absence as the evidence of the implicit term association:

$$\begin{aligned}\mu_{R_{\mathrm{NTAX}}}(s_i, f_i) &\approx \mathrm{BMI}(t_i, t_j)\\ &= \bar{\omega}_{\mathrm{BMI}} \times \left[Pr(t_i, t_j) \log_2 \left(\frac{Pr(t_i, t_j)+1}{Pr(t_i)Pr(t_j)} \right) + Pr(\neg t_i, \neg t_j) \log_2 \left(\frac{Pr(\neg t_i, \neg t_j)+1}{Pr(\neg t_i)Pr(\neg t_j)} \right) \right] \\ &\quad -(1-\bar{\omega}_{\mathrm{BMI}}) \times \left[Pr(t_i, \neg t_j) \log_2 \left(\frac{Pr(t_i, \neg t_j)+1}{Pr(t_i)Pr(\neg t_j)} \right) + Pr(\neg t_i, t_j) \log_2 \left(\frac{Pr(\neg t_i, t_j)+1}{Pr(\neg t_i)Pr(t_j)} \right) \right].\end{aligned} \tag{3.5}$$

The membership function $\mu_{R_{\mathrm{NTAX}}}(s_i, f_i)$ is used to measure the degree of association between opinion word s_i and product feature f_i. The corresponding weight parameter $\bar{\omega}_{\mathrm{BMI}} \in [0.4, 0.7]$ was used to adjust the positive and negative evidence. $Pr(t_i, t_j)$ is the joint probability of words appearing in the text window, and $Pr(t_i)$ indicates the probability that the word t_i appears in the text window. The estimation of probability $Pr(t_i)$ is based on $\frac{|w_t|}{|w|}$, where $|w_t|$ is the number of windows containing the word t and $|w|$ is the total number of windows in the corpus. Similarly, $Pr(t_i, t_j)$ is the number of windows that contain both two words at the same time as the total number of upper windows. Through the calculation of BMI value, the $\bar{w}_s = 25$ opinion words with the largest correlation value with a product feature can be extracted. All BMI values are linear normalization processing (i.e., $\mathrm{BMI}_{\mathrm{normal}} = \frac{\mathrm{BMI} - \mathrm{BMI}_{\mathrm{Min}}}{\mathrm{BMI}_{\mathrm{Max}} - \mathrm{BMI}_{\mathrm{Min}}}$)), such that $\mu_{R_{\mathrm{NTAX}}}(s_i, f_i) \in [0, 1]$ is maintained. Similar to previous product feature updates, the degree of association between the pair (s_i, f_i) can be continually updated on the basis of $\mu^{t+1}_{R_{\mathrm{NTAX}}}(s_i, f_i) = \alpha \times \mu^{t}_{R_{\mathrm{NTAX}}}(s_i, f_i) + \beta \times \mathrm{BMI}_{\mathrm{normal}}(s_i, f_i)$ after scanning a new collection of consumer reviews.

3.4. Analysis of sentiment factors in Chinese microblogging texts

3.4.1. Calculation of text influence

The text influence of a microblogging is related to posters (see Hou and Zhao [2]), and they consider this from several factors: (1) whether the user is the one who has been authenticated by the microblogging platform, if the user is a social celebrity, he has a strong influence; (2) the number of followers for the user, more followers indicate that he is more influential; and (3) the number of the user's friends, if the number of the user's friends is excessive, then he is just the recipient of information, the text influence is very small. Considering these factors, they get the calculation methods of influence for text d, as follows:

$$w_d = x\Big(\frac{f_{d,1}}{f_{d,2}}\Big) \times \min\Big\{\frac{f_{d,1}}{\delta}, \frac{f_{d,1}}{f_{d,2}}\Big\} \times v \tag{3.6}$$

where the term $f_{d,1}$ is the number of followers of the user who publishes text d and $f_{d,2}$ is the number of friends of the user who publishes text d. The function $x(l)$ is the expansion scale coefficient of the influence; according to the characteristics of the microblogging platform it is defined as follows: when $l \geq 10$, $x(l) = 2$; when $1 < l \leq 10$, $x(l) = 1$; when $l \leq 1$, $x(l) = 0$. δ, ν are the adjustable constants, where δ is a set depending on the microblogging platform; ν denotes whether the user is authenticated by the microblogging platform, if he has been authenticated, $\nu > 1$, otherwise, $\nu = 1$.

3.4.2. Semantic analysis of degree words

In Chinese microblogging, degree words are widely used alongside opinion words to change the intensity of opinion words. To better analyze users' emotional intensity, we set a detection window for each opinion word. The window size is set to 5. If some degree word appears in such a 5-word-wide window of an opinion word, its emotional intensity will increase 0.8-1.5 times depending on types of degree words. We extract 60 degree words from HowNet and divide them into seven types as shown in(Shi et al. [18]).

Formula (3.7) calculates the sentiment value by combining degree words with opinion words

$$SO(phrase) = value_{deg} \times sensibility(s) \tag{3.7}$$

where SO(phrase) is the value of phrase; *sensibility*(s) is the sentiment value of opinion words s (membership of opinion words); $value_{deg}$ is the intensity value of degree words. For example, the value of '非常好(very good)'$= 1.3 * 1 = 1.3$.

3.4.3. Semantic analysis of negative words

Negative words are commonly used in Chinese microblogging. Considering that the influence of negative words may increase the accuracy of the study, in this paper, 22 negative words were extracted from HowNet. Formula (3.8) refines the sentiment value of a sentence where negative words appear. We set a detection window for each opinion word. If the negative word is within the window, the sentiment value of the phrase will get negated. The size of the window is set to 5 in our experiment.

$$SO(phrase) = (-1)^n \times sensibility(s) \tag{3.8}$$

where SO(phrase) is the sentiment value of extracted phrase; n is the occurrence number of negative words in the window of opinion words s. For example, the sentiment value of '不好 (not good)' $= (-1) * 1 = -1$.

3.4.4. Semantic analysis of punctuations

Punctuation marks convey different sentiment signals to readers. At the sentence level, we annotate punctuations with emotions to refine the sentiment value of online reviews. In our study, exclamation mark (!) and question mark (?) are considered as two main punctuations for influencing emotion intensity of sentence. We separately set the effect of "!" and "?" on sentiment value of sentences are 1.2 times and 0.8 times as much as that of original sentences (the value is determined by the questionnaire of language learners). Formula (3.9) calculates sentiment value for a sentence with the punctuation '!' or '?'.

$$SO(sentence) = value_{pun} \times sensibility(s) \tag{3.9}$$

where $SO(sentence)$ denotes sentiment value of a *sentence*. $Sensibility(s)$ is sentiment value of a sentence without consideration of punctuations. $value_{pun}$ is the weight of pun. For example,

(1) 这款手机太好了! (English: This phone is very good!)

(2) 这款手机好吗? (English: This phone is good?)

For sentence (1), $SO(sentence) = 1.4 * 1 * 1.2 = 1.68$.
For sentence (2), $SO(sentence) = 1 * 0.8 = 0.8$.

3.4.5. Semantic analysis of rhetorical sentence

In our study, rhetoric is considered as a kind of semantic factor to affect the expression of sentiment. To the best of our knowledge, however, sentiment analysis considering such influences remains a largely understudied area. To measure the influence of rhetoric on emotion intensity, we select four types of rhetoric in Chinese: 明喻 (simile), 对偶 (antithesis), 排比 (parallelism) and 重复 (repetition).

The rules of judging rhetorical types are defined as follows:

(1) A simile is a rhetorical figure expressing comparison or likeness that directly compares two objects through some connective words, such as 像 (like), 似 (seem), 仿佛 (as), 犹如 (just as), 似的 (as if), 一样 (same as), 如 (as), 宛如 (just like) or a verb such as resembles. We can judge simile by finding whether or not connective words appear in a sentence.

(2) Antithesis is used when two opposites are introduced in the same sentence, for contrasting effect. Antithesis can be identified by judging if or not two adjacent sentences have the same amount of words.

(3) In rhetoric, parallelism means giving two or more parts of the sentences a similar form so as to give the passage a definite pattern. Parallelism can be identified if it is true of two adjacent sentences with a similar amount of words. In this experiment, the number difference of words in two sentences is set to be less than 2.

(4) Repetition is the simple repeating of a word, within a sentence or a poetical line, with no particular placement of the words, in order to provide emphasis. It is considered as a repeat when some words appear more than one times in a single sentence.

For the four rhetorical types, values of their influence on emotion intensity are set 1.1, 1.2, 1.3 and 1.4, respectively. Formula (3.10) is the sentimental value considering the influence of rhetorical types.

$$SO(sentence) = value_{rhe} \times sensibility(s) \tag{3.10}$$

where $SO(sentence)$ is the sentence's sentiment value; $Sensibility(s)$ represents sentence's sentiment value without considering the rhetoric; $value_{rhe}$ is the intensity value of rhetoric.

The semantic influence factors of Chinese microblogging text are obtained:

$$Value_{sem} = (-1)^n \times value_{deg} \times value_{pun} \times value_{rhe} \tag{3.11}$$

where $Value_{sem}$ is the semantic influence factor around opinion words, $value_{deg}$ is the intensity value of degree words, $value_{pun}$ is the intensity of punctuation pun, $value_{rhe}$ is the intensity of rhetoric rhe.

3.5. The sentiment polarity and intensity of the pair 'feature-opinion'

The sentiment value of opinion words has been introduced in the fuzzy representation of opinion words. Based on the pair 'feature-opinion' , it is easy to determine the intensity and the polarity for each pair 'feature-opinion'. The polarity score $polarity_{doc}(d)$ for a review d is derived by:

$$polarity_{doc}(d) = \frac{\sum\limits_{(s_i,f_i)\in d} polarity(s_i, f_i)}{|d|}. \tag{3.12}$$

The opinion value of the product feature f_j in a microblogging review d:

$$polarity_{doc}(f_j) = \frac{\sum\limits_{i=1}^{|f_j|} \mu_{R_{\text{NTAX}}}(s_i, f_j) \times polarity(s_i, f_j) \times value_{sem}(s_i, f_j)}{|f_j|}. \tag{3.13}$$

where $polarity(s_i, f_j)$ represents the sentiment value of the pair 'feature-opinion' (that is the membership of opinion word s_i in fuzzy opinion words thesaurus), $|d|$ is the number of the pair 'feature-opinion' in a microblogging review, $|f_j|$ is the number of opinion for the product feature f_j in the review d, $\mu_{R_{\text{NTAX}}}(s_i, f_j)$ is the fuzzy membership of the relation (s_i, f_j), $Value_{sem}(s_i, f_j)$ is the semantic factor around opinion word s_i.

The final evaluation value of the feature f_j for the product p_i is:

$$polarity(f_j) = \frac{\sum\limits_{d\in D_{Rev}} w_d \times polarity_{doc}(f_j)}{|D_{Rev}|} \tag{3.14}$$

where w_d is the influence factor of a microblogging d, $|D_{Rev}|$ is the number of microblogging reviews.

3.6. Sentiment value calculation of product reviews

Before sentiment analysis is applied to each consumer review, standard document pre-processing procedures (e.g., stop word removal, POS tagging, and stemming) are applied to the review document, As illustrated in the previous section, normalized TFIDF weighting is applied to extract the most representative product features from each review. In addition, low-frequency candidate product features are identified by matching the target tokens with the common product features, candidate opinion words which are close to the product features are then identified and selected to constitute the pair 'feature-opinion' according to the normalized BMI scores. As mentioned in the previous section, various semantic factors such as negative words and degree words which could change the sentiment value of the pair 'feature-opinion' are taken into consideration by our system in the same text window, Finally, the polarity score *polarity* (p_i) for a product pi is derived by:

$$polarity(p_i) = \frac{\sum\limits_{f_i \in F(p_i)} \sum\limits_{d \in D_{Rev}(s_i, f_i)} w_d \times \mu_{R_{\mathrm{NTAX}}}(f_i, p_i) \times \mu_{R_{\mathrm{NTAX}}}(s_i, f_i) \times polarity(s_i, f_i) \times value_{sem}(s_i, f_j)}{|D_{Rev}| \times |d|} \tag{3.15}$$

where $\mu_{R_{\mathrm{NTAX}}}(f_i, p_i)$ is the product and product feature association weight defined in the previous section and $\mu_{R_{\mathrm{NTAX}}}(s_i, f_i)$ is the fuzzy membership of the relation (s_i, f_i), w_d represents the influence factor of microblogging d(formula 3.6), $Value_{sem}(s_i, f_j)$ is the semantic factor around opinion word s_i, $|D_{Rev}|$ is the number of microblogging review, $|d|$ is the number of the pair 'feature-opinion' in a microblogging review. The polarity score $polarity(s_i, f_i)$ of a (s_i, f_i) pair is computed based on the fuzzy opinion words thesaurus. A certain threshold is established empirically to divide the value of *polarity* (p_i) into several levels, consumers could see the reputation of the product from the level of results.

4. Experiments and Results

In China, the counterpart of Twitter is Weibo, which is the word for word translation of micro (wei) and blogging (bo) in Chinese. There are four popular microblogging services in China: Tencent Weibo, Sina Weibo, Sohu Weibo, and Netease Weibo. In our study, we took the mobile phone reviews in Tencent Weibo (t.qq.com/) as the object, input "#mobile phone micro review #" in the search area of the platform, and got 69784 mobile phone reviews in the time interval from 2012/5/1 to 2013/5/1. And then three mobile phones which were Weibo users most interested in were selected by "heat" ranking: "# Nokia iumia 800#", "# Apple iPhone 4S 16GB#" and "# Motorola atrix 4G#". After microblogging texts cleaning and resolution (remove the nontext symbols, removed the forwarding content and other preprocessing) (see Shi, W et al.[18]), we got the specific statistics as shown in Table 3.

Table 3: Microblogging reviews statistics.

Mobile phone type	Review number number	Review reading	Average length of these reviews (words number)	Post number	Authenticated users
Nokia lumia 800	4665	5451	36.5	3983	75
Apple iPhone 4S 16GB	3562	7562	42	3023	58
Motorola atrix 4G	1651	3542	56	1527	19

Four evaluation indexes (precision, recall, accuracy, and F-measure) are introduced (see van Rijsbergen [24]):

$$\text{precision} = \frac{a}{a+b} \tag{4.1}$$

$$\text{recall} = \frac{a}{a+c} \tag{4.2}$$

$$\text{accuracy} = \frac{a+d}{a+b+c+d} \tag{4.3}$$

$$F_{\beta-1} = \frac{(1+\beta^2)\times \text{precision} \times \text{recall}}{\beta^2 \times \text{precision} + \text{recall}} = \frac{2a}{2a+b+c} \tag{4.4}$$

where

$a =$ the number of correctly classified positive (negative) microblogging reviews,

$b =$ the number of classified non-positive (negative) microblogging reviews,

$c =$ the number of non-classified positive (negative) microblogging reviews,

$d =$ the number of non-classified non-positive (negative) microblogging reviews.

Experiment 1: Detection window size selection

Previous studies have shown that the representation of the product and the proximity of the sentiment recognition (detection window size of opinion words) have an impact on the accuracy of opinion mining (see Hu and Liu [4]). The first experiment is to examine the differential impact of product feature expressions (different length noun patterns) and sentiment recognition (different proximity to product feature words) in the whole opinion mining. Try to change the detection window of opinion words from 1 (before and after the word of product features) to 10, and use the 'Noun (N)', 'Noun Noun (NN)', or 'Noun Noun Noun (NNN)' to represent the product features. Figure 2 describes the average accuracy of the PRSentiMiner system with varying detection window size and product feature length. The results show that the accuracy of opinion mining is the best when the feature expression form is a unigram (single noun) and the detection window size are 5-6 words. Consistent with earlier studies (see Hu and Liu [4]), if the window size was set to 5, and the extraction efficiency of opinion words related to product features was better.

With a small detection window, many candidate sentiments were missed. On the other hand, a large detection window might introduce too many irrelevant tokens (noises) to the process of sentiment extraction. However, unlike the previous studies, we found that unigram (a single noun) was more effective than bigram (two nouns) or trigram (three nouns) for the representation of product features. According to our in-depth analysis, it was revealed that standard names of product features such as "mobile screen" for mobile phones were not frequently used in consumer reviews. Indeed, in more than a half of the reviews related to mobile phones, the word "screen" alone (e.g., "The screen is nicely clear." 屏幕很清晰) was referred to. As a result, quite a number of important product features were not extracted from the reviews if we used the "Noun Noun" pattern to represent product features. Similarly, because most of the product features names only contain 1 to 2 nouns if we use "Noun Noun Noun" pattern to represent product features names, some product features will not be de detected, it will lead to a very poor recall.

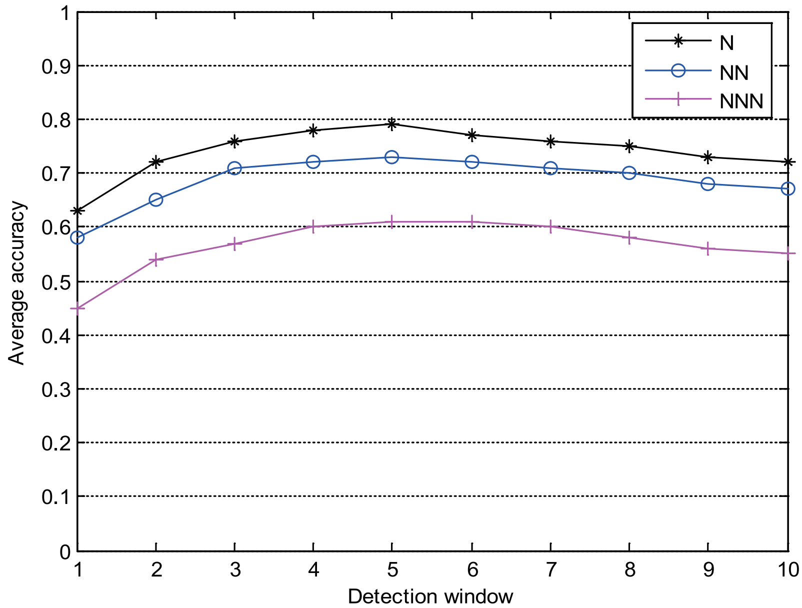

Figure 2: The impact of feature representation and window size on accuracy.

Experiment 2: Comparison of system results

Based on the established benchmark data set, we tried to evaluate the performance of the PRSentiMiner system. Benchmark dataset are typical microblogging product reviews: mobile phone reviews. Each sentence of a review was manually tagged by two researchers. We select 920 simple feature sentences (only one feature in the sentence) from 400 reviews (t.qq.com/) about cell phone as the experimental corpus. In order to ensure the accuracy of annotation, here also using kappa statistical method to measure the agreements of manually tagged by two researchers, the statistical formula is as follows,

$$k = \frac{P(A) - P(E)}{1 - P(E)} \tag{4.5}$$

$P(A)$ is the agreements proportion of two researchers in the result of annotation, $P(E)$ is the coincidence probability of consistent annotation, here we use consistency probability of the first intuitive tagged by the researchers to estimate. If the researchers are in complete agreement then $k = 1$. If there is no agreement among the researchers other than what would be expected by chance, $k = 0$. The annotation consistency is 95% in our experiment.

If the polarity (e.g., positive or negative) classified by our system was the same as the manually tagged polarity in the benchmark dataset, it was considered a correct classification. The performance of our system and that of the published results are tabulated in Table 4. It is shown that the precision, recall, and F-measure scores of our system outperform that of baseline methods (only NLP and sentiment vocabulary). In the baseline method, the sentiment intensity of an entire sentence is the sum of the emotional intensity for simple feature sentences with class G minus the sum of the emotional intensity for those with B.

Table 4: system performance based on benchmark data sets.

Systems	precision	recall	F-Measure	accuracy
Baseline method	0.68	0.72	0.79	0.69
PRSentiMiner	0.91	0.90	0.92	0.80

According to published results (see Pang et al. [11]), the accuracy of Support Vector Machine(SVM) based sentiment polarity classification falls in the range of $[0.8, 0.9]$. Although a direct comparison between our work and early studies is not possible due to varying experimental settings, to some extent, it reflects the effectiveness of the method in this paper.

Experiment 3: Application results and comparison

Finally, we tried to test the practical application of the PRSentiMiner system. For each product in users' reviews, we used crawlers to retrieve the product feature or product description from appropriate product page in Tencent Digital (http://digi.tech.qq.com/). And then the 50 most representative features (ie) were selected by the formula $(1)-(2)$, each feature was represented by a simple noun. Equation $(13)-(14)$ were used to determine the opinion values of product features, and finally, comprehensive opinion values of products on the microblogging platform were calculated by the formula (15) for reference. In our experiment, we took "# Nokia Lumia 800 #", "# Apple iPhone 4S 16GB #" and "# Motorola Atrix 4G #" as experimental subjects. For continuously updated learning function in our system, we filtered out 10 most representative product features (most users are willing to review) from 2012/5/1 to 2013/5/1: appearance, screen, function, speed, performance, pixel, price, cpu, resolution, power consumption.

Figure 3 plots the comparison of opinion values for ten features in three mobile phones. We can find that Apple iPhone 4S 16GB are better than the other two models

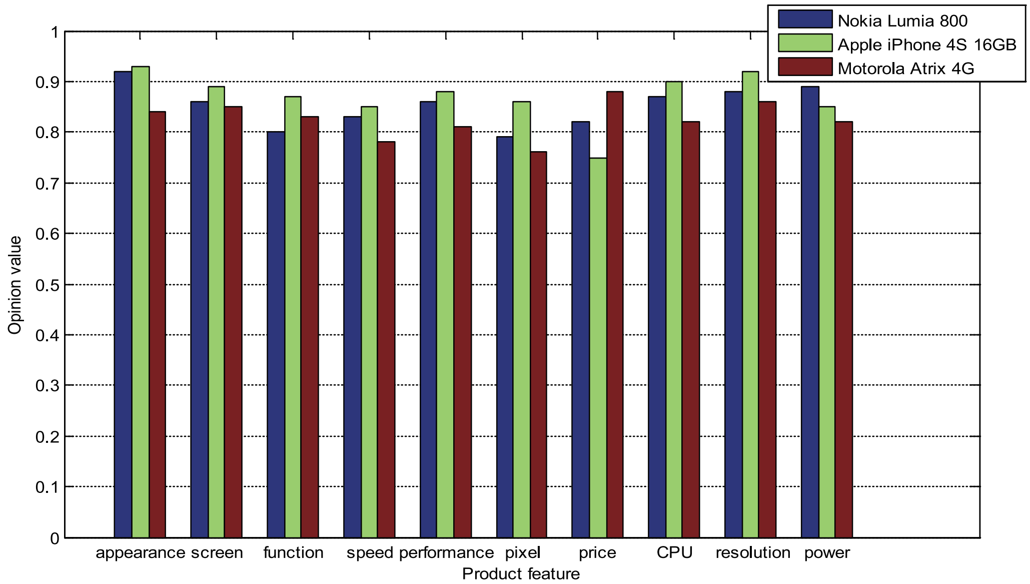

Figure 3: The comparison of opinion values for ten features in three phones.

in most of the features evaluation, but the users' opinion values are low for the price and power consumption. The performance of Nokia Lumia 800 and Motorola 4G Atrix are similar in many features. However, in general, Nokia Lumia 800 is a little better than the Motorola 4G Atrix in Chinese microblogging reputation. These information can be used as a reference for businesses and users. There are users' scoring columns for the corresponding products in Tencent digital(http://digi.tech.qq.com/), evaluation results of our system are basically the same as the results of scoring system for the three mobile phones, the total score of Apple iPhone 4S 16GB is 8.6, the total score of Nokia lumia 800 is 8.5, the total score of Motorola atrix 4G is 8.4.

Figure 4 is the comprehensive opinion values for the three mobile phones according to formula (15). The results show that Apple iPhone 4S 16GB has the best Word of mouth among the three mobile phones in Chinese microblogging, which can also explain that the sales of Apple iPhone 4S 16GB are better than the two other mobile phones. The results of Experiment 3 show that our PRSentiMiner system has very good application value. Output results can be as the basis for businesses to adjust the product structure or sales strategy, they can also be an important reference for users to buy products.

5. Conclusions

We illustrate the design, development, and evaluation of a novel product review mining system based on Chinese microblogging in this paper. The research contributions of the paper as follows:

(1) We construct the framework of features extraction, feature-opinion extraction and updating from product reviews in Chinese microblogging. Specifically, we use the

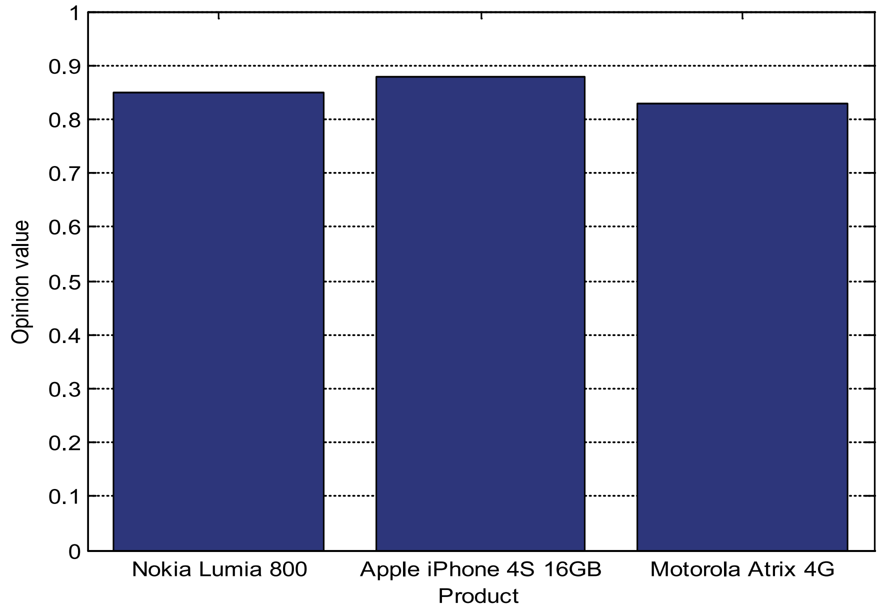

Figure 4: The comparison of comprehensive opinion values for the three phones.

traditional TFIDF formula to extract the noun model which represents product features, the product feature set is automatically updated by Rocchio learning method, and the most representative product features are retained by parameter control for the further study. Then the opinion words set which have the highest membership degree with given product features is determined by the method of BMI, the most representative pair 'feature- opinion' is selected by setting parameter for product evaluation, and similar to product features updating way, the pair 'feature- opinion' could also be updated automatically with the number of product reviews increasing.

(2) Based on fuzzy opinion words thesaurus, the sentiment value of pair 'feature- opinion' is determined by calculation of association degree between product features and opinion words. We also consider the various emotional semantic factors in Chinese microblogging and finally determine the calculation method of sentiment value for product reviews.

(3) Our system can accurately predict the polarities of sentiments without requiring extra human effort to annotate training examples. Based on real consumer reviews collected from Tencent Weibo (t.qq.com/), the effectiveness of our PRSentiMiner system is empirically tested, the proposed system performs significantly better than a baseline system. The business implication of our research is tremendous, our opinion mining methodology assists organizations to analyze a large number of consumer reviews efficiently based on Chinese microblogging. As a result, organizations can develop effective business strategies related to marketing, customer support, and product design functions in a timely fashion. Our work also facilitates a large number of microblog users' comparison shopping processes.

Future research involves examining the correlation of the product sentiment scores generated by our system in microblogging and the actual sales ranks or sales volumes of various products and so on, these researches will be a supplement for sentiment analysis based on Chinese microblogging.

Acknowledgements

This research was supported by Zhejiang Provincial Natural Science Foundation of China under Grant No. LY15G030023; The Results of Social Science Planning Project in Zhejiang Province under Grant No.16NDJC079YB; Project supported by key humanities and social science projects in Zhejiang province under Grant No. 2016QN003.

References

[1] Alec, G., Richa, B., and Lei, H. (2009). *Twitter sentiment classi-fication using distant supervision*, CS224N Project Report, Stanford.

[2] Hou, S. and Zhao, Z. (2011). *Research on product market analysis model for microbloggings platform*, Microcomputer Application, Vol.27, 4-7.

[3] HowNet. (2007). HowNet's home page. Retrieved from http://www.keenage.com

[4] Hu, M. and Liu, B. (2004). *Mining and summarizing customer reviews*, In Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, Washington, 168-177.

[5] Johan, B., Alberto, P., and Huina, M. (2010). *Modeling public mood and emotion: Twitter sentiment and socioeconomic phenomena*, WWW 2010, April 2630, Raleigh, North Carolina,235-242.

[6] Lau, R. Y. K., Song, D., Li, Y., Cheung, C.H. and Hao, J. X. (2009). *Towards A Fuzzy Domain Ontology Extraction Method for Adaptive e-Learning*, IEEE Transactions on Knowledge and Data Engineering, Vol.21, 800-813.

[7] Liu, Z. M. and Liu, L. (2012). *Empirical study of sentiment classification for Chinese microblog based on machine learning*, Computer Engineering and Applications, Vol.48, 1-4.

[8] Meng, J., Duan, X., and Yang, L. (2013). *Opinion Analysis of Cross-domain Product Review Based on Feature Transformation*, Computer Engineering, Vol.202, 168-171.

[9] Morinaga, S., Yamanishi, K., and Tateishi, K. (2002). Mining product reputations on the web. In Proceedings of the 8th ACM SIGKDD international conference on knowledge discovery and data mining (KDD), ACM, New York, 341-349.

[10] Moghaddam, S. and Ester, M. (2010). *Opinion digger: an unsupervised opinion miner from unstructured product reviews*, Proceedings of the 19th ACM international conference on Information and knowledge management. ACM, New York, 1825-1828.

[11] Pang, B., Lee, L., and Vaithyanathan, S. (2002). *Thumbs up? Sentiment Classification using Machine Learning Techniques*. In Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Philadelphia, 79-86.

[12] Pan, Y. and Lin, H. (2008). *Restaurant reviews mining based on semantic polarity analysis*, Computer Engineering, Vol.34, 208-210.

[13] Pfeiffer, H. D. and Tourte, G. (2012). *Twitter, information sharing and the London Riots?* Bulletin of the American Society for Information Science and Technology, Vol.38, 49-57.

[14] Popescu, A. M. and Etzioni, O. (2005). *Extracting product features and opinions from reviews*, Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing (HLT/EMNLP), Association for Computational Linguistics, Morristown, 339-346.

[15] Raymond, Y. K., Lau, Chapmann, C.L., Lai, J. M., and Yuefeng, L.(2009). Automatic Domain Ontology Extraction for Context-Sensitive Opinion Mining. ICIS 2009 Proceedings, Phoenix, paper 35.

[16] Salton, G. (1990). *Full text information processing using the smart system*, Database Engineering Bulletin, Vol.13, 2-9.
[17] Shi, N. S. (2013). *Research on opinion mining of product reviews based on syntactic tree pattern*, Donghua University master's thesis.
[18] Shi, W., Wang, H., and He, S. (2013). *Sentiment analysis of Chinese microblogging based on sentiment ontology: A case study of '7.23 Wenzhou Train Collision'*, Connection Science, Vol.25, 161-178.
[19] Shi, P. Z. (2012). *Sentiment classification of product reviews based on CRFs*, Shanghai Normal University master's thesis.
[20] Svetlana, K., Zhu, X. D., and Saif, M. M. (2014). *Sentiment analysis of short informal texts*, Journal of Artificial Intelligence Research (JAIR), Vol.50,723-762.
[21] Tang, X. B. and Wang, H.Y. (2013). *Research on microblogging product reviews mining model*, Journal of intelligence, Vol.32, 107-111.
[22] Thelwall, M., Buckley, K., and Paltoglou, G. (2011). *Sentiment in Twitter events*, Journal of the American Society for Information Science and Technology, Vol.62, 406-418.
[23] Turney, P. (2002). *Thumbs up or Thumbs down? Semantic orientation applied to unsupervised classification of reviews*, Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), USA: Association for Computational Linguistics, Morristown, 417-424.
[24] van Rijsbergen, C. (1979). Information Retrieval. London: Butterworths.
[25] Wei, H. and Wang, Y. H. (2015). *Sentiment Analysis of Chinese Micro-blog Based on Topic*, Computer Engineering, Vol.41, 238-244.
[26] Yao, T. and Lou, D. (2007). *Research on semantic orientation distinction for Chinese sentiment words*, In Proceedings of the 7th International Conference on Chinese information processing, BEW, Wu Han, 222-225.
[27] Zhan, H., Chang, G.Y. and Jung-Jae, K. (2014). *Identifying Features in Opinion Mining via Intrinsic and Extrinsic Domain Relevance*, IEEE Transactions on Knowledge and Data Engineering, Vol.26, 623-634.
[28] Zhan, J., Loh, H. T. and Liu, Y. (2009). *Gather customer concerns from online product reviews - A text summarization approach*, Expert Systems with Applications, Vol.36, 2107-2115.

Business School, Huzhou University, Hu Zhou 313000, P.R. China.
E-mail: shiwei@zjhu.edu.cn

Major area(s): Sentiment analysis, Text mining, Business intelligence.

(Received May 2017; accepted April 2018)