WHICEB 2016 Proceedings

Wuhan International Conference on e-Business

Summer 5-27-2016

Mining Comparison Opinions from Chinese Online Reviews for Restaurant Competitive Analysis

Song Gao
School of Economics and Management, Tongji University, Shanghai 200092, China, hardygao@outlook.com

Hongwei Wang
School of Economics and Management, Tongji University, Shanghai 200092, China, hwwang@tongji.edu.cn

Yuan Song
School of Economics and Management, Tongji University, Shanghai 200092, China

Ting Lu
School Saic General Motors Sales Co., LTD, Shanghai 201206, China

Mining Comparison Opinions from Chinese Online Reviews for Restaurant Competitive Analysis

Song Gao [1], *Hongwei Wang*[1*], *Yuan Song* [1], *Ting Lu* [2]
[1]School of Economics and Management, Tongji University, Shanghai 200092, China
[2]School Saic General Motors Sales Co., LTD, Shanghai 201206, China

Abstract: Comparison is widely used by consumers during the process of product evaluation in order to emphasize their preference, which can contribute to a proxy for product competitiveness analysis. This paper proposes a novel method for mining comparative sentences based on the achievements of linguistic study. The definition of comparative sentence subcategory is put forward and a mixed rule pool containing both artificial rules and CSR is set up. Besides, an entity dictionary is used to re-check the identification result which can ensure precise identification and classification of comparative sentences. Real online comments are collected from Dianping.com as experimental data. The result shows that the proposed method outperforms baseline methods in terms of identification precision. Based on the result, features and opinions of comparative sentences are mined. We then conducted sentiment analysis to calculate the sentimental score of comparison relations. Finally, a feature competitive network of restaurants is constructed.

Keywords: comparative pattern, competitiveness analysis, pattern match, class sequence rule

1. INTRODUCTION

Comparison is a common way to express the relationship between two objects, especially in consumer reviews. While writing comments, consumers tend to compare the product or seller with their counterparts to express dissatisfaction or satisfaction. Comparative information can not only help consumers make better consumption decisions, but also help companies identify the potential gaps and discover valuable business information. It is observed that around 10% of online reviews are written in the form of comparative sentences. With the rapid development of electronic business platform like "*www.taobao.com*", the demand for mining comparative information out of massive Chinese reviews is rising. It is imperative to find effective methods of identifying, extracting and quantifying the comparative information to help consumers make better purchasing decisions, and to enable companies acquiring quick corresponsive ability for market changes.

Related researches mainly focus on mining comparative sentences from English text [1][2][3][4][5]. However, compared with English, Chinese expression is more flexible in word and sentence structures. Especially, Chinese comparison could even be ambiguous sometimes. To solve the problem, this paper focuses on Chinese comparative information mining, and proposes a combined method based on pattern matching and machine learning to realize comparative sentences recognition and classification.

2. RELATED WORK

2.1 Linguistic area

Comparison is an important means for human to understand things by comparing and discovering the similarities and differences among objectives. Researches on comparative relation are mainly from linguistics perspective. They focus primarily on defining the syntax and semantics of comparative constructs [6], but have not examined computational methods for extracting comparative relations [1]. Many Chinese scholars have defined the category of Chinese comparative sentences [7][8].

* Corresponding author. Email: hwwang@tongji.edu.cn(Hongwei Wang) , hardygao@outlook.com(Song Gao)

Research on comparative words and syntactical structures is another major field within linguistics area. Some scholars analyze and conclude grammatical features, typical sentence forms, and possible deformation for sentences belong to 'Equal' and 'Non-Equal' gradable category respectively [8]. Typical Chinese comparable words include "和", "跟", "与" or "同" which mean "with" or "and", while words like "一样", "相同", or "类似" are typical opinion words which express the meaning of similar or the same [7]. Besides, there are some special forms like "有...那么", "像...一样", "不比...差"[7][8].

2.2 Mining comparison opinions

Currently the task of comparison opinions mining is mainly divided into three parts: 1) identifying sentences that contain comparison information; 2) extracting elements of comparison relation; 3) calculating the sentimental score of the comparison relation[9] [10] [11].

Usually, comparative sentences may contain some typical words with comparison meaning or suit certain patterns. Therefore, by concluding such features we can get some rules which can be used to identify comparative sentences. Typical rules include: keyword, pattern, and class sequence rule (CSR) etc. It is believed that patterns that the recognition precision of considering co-occurrence of words report is better than pure keywords, while CSR may have the best performance among the three [9] [10]. Classifier is also a common method for classification problem such as Support Vector Machine, Naïve Bayes etc. However, this kind of method is rarely used alone in this research area. Results of current researches show that the method combining rules with classifier can get better result [2] [9] [10].

The major task of comparative opinions extraction is to mine out the comparison entity, comparing point, as well as the opinion. For entity extraction, recent studies believe that syntactic analysis can help improve the precision. Semantic Role Labeling (SRL) is the most commonly used method. Some researchers also try to combine SRL with Condition Random Field (CRF) or syntactic dependency tree. The extracting of comparing point and opinion is a kind of feature and opinion mining and therefore can utilize current technologies. Previous researches mainly adopt the method of manually defining the dictionary of feature words and opinion words. Later, Hu and Liu [11] propose an association rule based on high frequency feature mining method which becomes quite popular within this field. Chang and Shi [12] point out that the connection between feature and opinion should be considered. But, some changes are needed for comparison information mining. For example, the type of comparison sentences should be considered while mining the comparative opinion word.

Comparative opinion calculation is the final step of comparison opinions mining. Essentially comparative opinion calculating is the calculating of emotion polarity and intensity, which is a major problem that study of sentiment analysis is dealing with. Sentiment analysis aims at discovering the attitude of reviewer towards certain topic or object by analyzing the text or document. Mainly three tasks are involved, classification of the objectivity or subjectivity of the emotion, emotion direction judging and emotion intensity calculating. According to the granularity analysis, related researches can also be divided into word-level, sentence-level and document-level [13] [14] [15]. Existing researches only focus on the emotion expressed upon one product or object while comparative opinion may deal with two or more objects. Therefore, comparative opinion calculation is a kind of sentiment analysis on sentence-level.

2.3 Research gaps and questions

So far, many efforts have been dedicated to comparison opinions mining, such as rule corpus setup, comparative sentence identifying, comparison opinion extracting, and so on. However, few studies contribute substantially to the multi-class classification problem of comparison sentences and visualization of comparative relations. In current studies, all sentences are simply divided into comparative or non-comparative sentences, ignoring the different categories. Furthermore, these studies are conducted in the way that does not distinguish comparative sentences' sub-category. And the visualization of comparative relation does not attract enough attentions from academics. This paper adopts the linguistic definition of comparative sentence sub-category and defines the recognition problem as a multi-class classification problem. The mixed rules based on comparative sentence identification approach are proposed and validated. In the experimental part, comparative information network construction is combined with the proposed method to realize a visualized restaurant competitiveness analysis.

Our contributions are expected as follows: 1) Construct mixed rules base to get higher precision with less supervision. 2) Combine the proposed method with sentiment analysis and realize the visualized restaurant competitiveness analysis based on mining comparative information.

3. METHOD FOR MINGING COMPARATIVE SENTENCES

3.1 Sub-category classification of comparative sentences

According to the definition of sub-category of comparison sentences in linguistics, the types of comparative sentences are divided into “Equative” and “Non-Equal” as shown in Table 1. A special type of comparison is “Superlative”, which usually may contain the word “最”(most). In this paper, it is deemed as the subdivision of “Non-equal” and all related keywords will be included in the keyword dictionary as well.

Table 1. Sub-category of comparative sentences

Category		Meaning	Comparing object
Equative		The comparing entities are the same or alike	One/more objects
Non-Equal	Better/worse/ different	The comparing entities are different	One/more objects
	Superlative	The best/worst among all objectects	More than 2 objects

Comparative keyword is an important symbol for comparative sentences, especially for English texts. However, for Chinese text, the sentence structure is more complicated. To make a complete comparison, the comparative keyword need to pair with the words that show the comparative result, such as predicate verb, adjective, preposition, etc. Some of the improved methods on pattern rule base have considered this feature and try to list all possible combinations of comparing keywords with result words. In fact, it is feasible for “Equative” type sentences as the result words of this type, all of which have the meaning of the same or similar and therefore can be listed. Although there may also exist some special cases that only use a verb to stand for “Equative” comparison, and listing most of these words is possible. However, for “Non-Equal” sentences, this method may require a tremendous amount of work. Even when a rule base is constructed, it is not easy to cover all possible patterns. “Non-Equal” sentences tend to be more flexible in forms and the possibility of result words is infinite. In fact, almost all adjectives can act as a result word and different areas may require different patterns or keywords. To simplify the work, using CSR may be a more suitable way to recognize “Non-Equal” sentences. By using Part-of-Speech (POS) to replace the result word, we do not need to worry about what the result word may be, but just focus on the possible sequence pattern. This can greatly reduce the workload and at the same time achieve satisfied result. Therefore, this paper proposes a method combining artificial rules and CSR, then constructs a mixed rule base for recognizing comparative sentences.

3.2 Class sequence rule

Using CSR for comparative sentences is firstly proposed by Jindal and Liu [1][2]. This paper will make some improvements based on their method. The construction of keyword dictionary will reference the existing patterns that have been concluded in previous research and take the fruits of linguistic study into consideration. By observing the real online comments, some newly emerging words will also be included.

Although we have proposed a co-occurrence keyword dictionary for identifying possible CSRs, there may still be some cases where only one keyword existing in a sequence. As the CSR only extract the POS tag and keywords, it is possible that a sequence, especially that with only one keyword, may cover CSRs with different class factor. Therefore, not all candidate CSRs can be used to identify whether a sentence is “Non-Equal” or not. It is necessary to set the minimum support and confidence threshold to screen rules and ensure their quality. Support and confidence are the most commonly used indicators to measure quality of rules. The support of a rule means the percentage of instances that satisfy the rule. The confidence refers to the ratio of instances that satisfy the rule divided by that cover the rule. For Chinese comparative sentences, keywords may differ in frequency, therefore using a single support threshold for screening rules is improper. To tackle this problem, Jindal and Liu [2] propose a multiple minimum support method. By this method, only CSRs with support value $\sup(r)$ which is bigger than $\lambda \cdot \min(f_i)$ can be extracted into the final rule base. Huang et al. [16] point out this method may report poor performance if a sequence contains an item with extremely low frequency. The support of this sequence may turn out to be less than $1/N$ where N is the size of the whole sequence database and thus leading to significant increase in computing complexity. Therefore, Huang et al. [16] propose a method to improve support threshold value as shown in Equation (1). This paper will follow this method and filter out CSRs that cannot meet this constraint.

$$\sup(r) > \max(\lambda \cdot \min(f_i), s) \qquad (1)$$

Here, f_i is the frequency of the i_{th} item of rule r in the whole dataset, λ is a value in a range of 0 and 1, s is a support threshold greater than $1/N$. All parameter setting will follow Huang’s study where λ is 0.1 and s is $2/N$. As for the confidence setting, a comparative experiment is conducted and the confidence threshold is set as 0.7 finally. Compared with [16], we use a slightly higher confidence threshold. Because it can ensure the accuracy of all selected CSRs to identify “Non-Equal” sentences, and avoid the over-fitting of rules.

The length of sequence is also an important factor that may affect the method for sequence pattern mining. Jindal and Liu [1] use a fixed length strategy for English text which will only extract the keyword and 3 items before and after it respectively. However, Chinese text is special in that the words composing a comparative relation may not be close to each other. Therefore, this paper will take Huang’s strategy to use a clause as a sequence. By observing the real online comments, we find out that most comparative structures are within one clause instead of cross clauses. This also proves that using a clause as a sequence is feasible.

It should be noted that, considering the complexity of Chinese, even if the multiple supports are set, we still cannot avoid the circumstance that the extracted rules may apply both “Non-Equal” and “Equative”, sometimes may even suit “Non-comparison”. Therefore, CSRs cannot be used alone to identify “Non-Equal” sentences as it may cause some misjudgments. Most of the researches use CSR as an input feature for classifiers and then use the trained classifier to identify comparative sentences. This is a kind of method based on statistical calculation. However, online reviews tend to be more casual in forms and may not always follow certain language structures. In response to these features, this article will use the way proposed by He [17] in his study. An entity (restaurant name in this paper) dictionary is set up to scan all candidate comparative sentences recognized by the mixed rule base. After the second time scanning, all misclassified sentences can be excluded, therefore improves the accuracy of the proposed method.

3.3 Mixed rule base

In this paper, the mixed rule base mainly contains two types of rules, artificial rules for identifying “Equative” sentences and CSRs for identifying “Non-Equal” sentences. The comparative keyword dictionary and the co-occurrence keyword patterns are manually constructed while CSRs are extracted automatically from training database and filtered by multiple minimum supports and minimum confidence.

According to the linguistic study, most “Equative” sentences tend to have the structure of one comparing word paired with one result word. Although there may be a few cases that one special verb can represent the whole comparing structure, usually these words and patterns can be listed manually. However, the situation of “Non-Equal” sentences is more complicated as there are much more special structures of this type and the result word can have a lot of possibility when the application field changes. Therefore, method of manual pattern does not suit especially when new words or new application fields are emerging. For this kind of result words, the only thing can be determined is their POS tag, either verb or adjective. Therefore, CSR is a more suitable method which can save effort of listing all possible words, and avoid too much human supervision during the procedure of constructing the whole mixed rule base.

For establishing artificial dictionary, we also try to make it less time consuming and control the degree of human effort. Based on researches [1][16], the Chinese dictionary as well as massing real online texts is included to expand current dictionary.

The first step of establishing artificial rule base is to construct a keyword list, which mainly contains comparing keywords that represent the relation of comparison and esult words that can express the meaning of “Equative” comparison. Besides, some common words, synonyms, slangs and network terminologies are also included in the keyword list. For those words specifically expressing the meaning of self-comparison, a mark will be labeled as shown in table 2. Chinese comparative sentences are more complex than English ones, therefore cannot depend on one single word like “than”. Usually, a complete Chinese comparative structure is formed by 2 or more words. These words may be far away from each other, sometimes even not in the same clause. Therefore, a co-occurrence pattern is set up as shown in table 3. Based on the keyword list, all keywords that can pair with each other are saved as patterns. Here, we assume that a structure may contain at most 2 keywords. If a comparison only has one keyword, this pattern will be saved as the single word pattern.

Table 2. Example of keywords

No.	Keyword	Type
1	and(和)/c	Equative，Non-Equal
2	almost(差不多)/l	Equative
3	with(跟)/p	Equative，Non-Equal
4	still(一如既往)/i	Equative，Self
5	more and more(越来越)/d	Non-Equal，Self
6	than(比)/p	Non-Equal

Table 3. Example of keyword co-occurrence patterns

Type	Co-occurrence pattern	
Equative	2 almost(差不多)/l	Equative
Equative	2 with(跟)/p	Equative
Non-Equal	1 and(和)/c	Non-Equal
Equative（Self）	4 still(一如既往)/i	Equative（Self）
Non-Equal（Self）	5 more and more(越来越)/d	Non-Equal（Self）

Use the co-occurrence pattern to scan all sentences that may be “Non-Equal”. Extracting the current clause that contains the keywords and transferring the clause into a CSR by the method introduced in 3.3. Firstly filter all CSRs by the specified confidence and then calculate the minimum support for each CSR. Finally, only the remained CSRs can be added into the mixed rule base.

For “Equative” type, a pattern base is set up. Both the word and its POS tag are extracted and saved. All those marked as “Equative” in the co-occurrence pattern dictionary can be used directly here. If a clause contains an “Equative” pattern, then the current sentence will be classified into “Equative”. The final mixed rule base for comparative sentences recognition is shown as table 4.

Table 4. Example of mixed rules

Equative	{still(一如既往)/i} （self）
	{and(和)/c，almost(差不多)/l}
	{with(跟)/p，almost(差不多)/l}
Non-Equal Equative	{/n more and more(越来越)/d /a，“Non-Equal”，“self”}
	{/n than(比)/p /ns /a，“Non-Equal”}
	{ best(最好)/, “Superlative”}

4. EXPERIMENT AND RESULT

4.1 Data prepare

Experimental data are from *Dianping.com*, a famous restaurant review management website in China. Take the region of Shanghai, extract comments from 100 restaurants, which account for around 26.58% of all comments. Randomly pick 20000 comments and exclude those which are not qualified for experiment like too short or written in English. Finally, we get 14872 as the raw data and divide them into training set and test data, 12872 and 2000 respectively. The restaurant’s names in comment are also extracted to set up the comparison entity dictionary, which also contains the short names and some pronouns.

Some preprocessing is done including word segmentation, typos correction, synonyms merger and stop words filtration. 10 testers are invited and trained about the definition of comparison sentence category. According to the method proposed in section 3, testers will help label all data into three categories. If a tester is not sure about the comparison category of a comment, other testers will join to help and work out a category that everyone agrees with. The mixed rule base is set up based on the training set and the training process is realized by a VC program.

After getting the rule base, the method is tested using the test data. Firstly, pick out all candidate comparative sentences that contain words matching those in the keyword dictionary. Secondly, identify all “Non-Equal” sentences using the CSRs. Thirdly, for the rest of candidate comparative sentences, we use the patterns rule to identify the “Equative” ones. Lastly, scan all the “Equative” and “Non-Equal” and eliminate those without comparison entity, thus get the final classification result.

4.2 Results and analysis

Based on the training data, a mixed rule base is set up. For “Equative” pattern, 291 rules are set up, among which 27 belong to single word pattern and all of the rest are co-occurrence pattern. For “Non-Equal” pattern, 372 rules are extracted, 322 of which are selected using the support and confidence threshold defined in Section 3 while the rest 50 are manually added rules.

A comparative experiment is conducted between the proposed method and other methods. The evaluation indicators being used include the precision, the recall, the overall accuracy and the F-Measure. All methods will be denoted in short as below:

1) T, keyword
2) SVM, using keyword and its nearby as input feature of SVM
3) SVM+N, based on 2), add the entity name information as the input feature
4) TP+CSR, Keyword co-occurrence dictionary combined with CSR
5) TP+CSR+N，Proposed method, the entity name dictionary is added for second time recognition

The experiment result is as shown in table 5 and table 6. Compared with all the other baseline methods, the proposed method reports a better performance in precision, reaching 85.9%. Although the recall of this method is a bit lower than that of pure keyword strategy, it has the best overall performance with F measure of 91.44%. By using a mixed rule base, the identification of comparison sentences is greatly improved in term of precision and at the same time, the subcategory classification also achieves a satisfactory result.

Table 5. Experiment result of different approaches for this problem

Method	Precision		Recall		Overall Accuracy	F Measure
	Comparison	Non comparison	Comparison	Non comparison		Comparison
SVM	56.59%	81.63%	72.62%	68.59%	70.04%	56.59%
SVM+N	72.73%	90.01%	83.81%	82.28%	82.83%	72.73%
T	63.55%	97.46%	99.03%	94.44%	94.85%	63.55%
TP+CSR	76.71%	99.91%	97.17%	97.11%	97.12%	76.71%

Table 6. Sub-category discrimination result of the proposed method

	Precision	Recall
Equative	90.56%	85.98%
Non-Equal	90.13%	93.51%

It can be detected that few cases of Non-Comparison sentences are classified as Comparison sentences. This is mainly because of the use of entity dictionary. Most of the errors are coming from the misjudging between “Non-Equal” and “Equative” classes. This is because that the impact of negative word has not been considered. For example, “涨价” (Price hikes) is a keyword for “Non-Equal” sentences, so all sentences with this keyword will be classified as “Non-Equal”. But if there is a negative word right in front of it, then the meaning of the sentence may change into “the price remains the same” which should be an “Equative” sentence. To avoid this kind of error and improve the performance of the mixed rule base, the impact of negative words can be considered during the comparative sentence recognition.

4.3 Visualized analysis of restaurant competitiveness

Based on the experimental result, the sentiment analysis technology [18] is used to realize a visualized analysis. Still take “*Dianping.com*” as an example. The comparative entities are extracted for each comparative sentence that have been recognized and classified. The opinion words are mined with the method proposed in [19]. We adopt sentimental score for each comparative relation considering the position information and negative words.

The calculation formula of emotional score on sentence level is as shown in equation (2). The total count of negative words is saved as n, the sentiment polarity of the comparing keyword will be saved as c. If the keyword has obvious tendency of negative sentiment, then c=-1, otherwise, c=1. The comparative type is denoted as f. If the type is “Equative”, f=0, which is not relevant for sentimental calculation. If the type is “Non-Equal”, f will be 1. l represents the position information of comparative entities and the default value is 1. If major entity appears after the keyword or the baseline entity appears in front of the keyword, l will be -1. The emotional score of opinion word is saved as score. The direction of score is determined using a dictionary combined with similarity computing. Here the dictionary is provided by *HowNet* which lists some possible Chinese sentiment words. If the pattern match fails for a word, the similarity of the word with all typical positive and negative opinion words will be calculated. By comparing the weighted average of the two types, the direction can be determined. As for the value of score, that is the sentiment intensity, is determined by the sentiment score of the modifier words. All opinion words will be assigned the same sentiment value, 3 or -3. Then for those have modifier words nearby, the sentiment intensity is adjusted with the aid of a dictionary. Finally, by considering all the factors together, we can get the sentiment score of each comparison relation.

$$\text{Score} = (-1)^{n} \cdot c \cdot f \cdot l \cdot score \qquad (2)$$

Each comparative relation can be quantified and saved as a quad mode <E_1, E_2, F, $Score$>. Here, E_1 represents the current commented object, E_2 is the comparing baseline, F is the comparative point and $Score$ is the overall emotional score of relation. Each feature as a unit, construct the comparing network between various entities. The network structure and weight setting method all follow paper [20]. After all the network are constructed, the restaurant competitiveness analysis can be realized which shows more business insight. For example, the restaurant can compare itself with its major competitors on some key indicators to see the gap, which can be realized using the visualized result shown in figure 2. For a chain stores manager, he can also utilize the analysis tool to check the performance differences of all the branches. Restaurant 1 and 2 are marked as 4.5 on the website, but the visualized analysis result tells us restaurant 1 enjoys better reputation on taste and diversity of dishes while restaurant 2 mainly earns the reputation by cheaper price and better service.

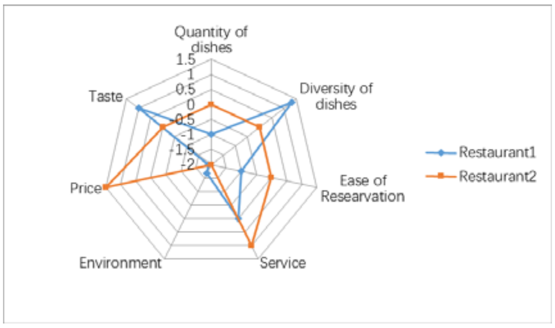

Figure 2. Example of restaurant competitiveness

5. CONCLUSIONS

This paper proposes an improved method about comparative sentence recognition. We construct the mixed rule base combined with entity name dictionary. Compared with previous study, the proposed method can achieve better recognition precision with less artificial work and supervision. As shown in the experimental result, this method outperforms many methods. By classifying sentences directly into ‘Equative’, ‘Non-Equal’ and ‘Non-Comparison’, the recognition result is good, then the mining work can be simplfied. Finally, based on above method, a visualized restaurant competitiveness analysis is made which testifies the practical value of this paper. To realize real business intellegence, the technology of comparative sentence recognition can be combined with product feature mining, sentiment analysis, and comparative network construction etc. As a result, the potential useful business value can be extracted and assist the decision makers to make correct decision or judgement. What should be noted is that the accurancy of comparative sentence recognition will directly affect the reliability and credibility of the results gained from the following mining work. Therefore, the accuracy and applicability of the identificaiton method is the focus of future research.

Future study can be conducted in the following aspects: 1) Improve the method of mixed rules by adding the negative words, which can optimize the classification precision between ‘Equative’ and ‘Non-Equal’. 3) Improve the method for recognition of implicit comparison sentences.

ACKNOWLEDGEMENT

This work is partially supported by the NSFC Grant 70971099 and 71371144, and Shanghai philosophy and social science planning projects (2013BGL004).

REFERENCES

[1] Jindal N, Liu B. (2006). Identifying comparative sentences in text documents. In Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval, 244-251.

[2] Jindal N, Liu B. (2006). Mining Comparative Sentences and Relations. AAAI American Association for Artificial Intelligence, vol.22: 1331-1336.

[3] Ganapathibhotla M, Liu B. (2008). Mining opinions in comparative sentences. In Proceedings of the 22nd International Conference on Computational Linguistics, vol.1: 241-248.

[4] Xu K, Liao S S, Li J, Song Y. (2011). Mining comparative opinions from customer reviews for Competitive Intelligence. Decision support systems, 50(4): 743-754.

[5] Zhang Z, Guo C, Goes P. (2013). Product comparison networks for competitive analysis of online word-of-mouth. ACM Transactions on Management Information Systems (TMIS), 3(4): 20.

[6] Moltmann, F. (1992). Coordination and comparatives. PhD Thesis. Massachusetts Inst. of Technology Cambridge.

[7] Xia Qun. (2009). A Review on Studies of Comparative Sentence of Chinese. Chinese Language Learning, 6(2):59-64(in Chinese).

[8] Li C N, Thompson S A. (1989). Mandarin Chinese: A functional reference grammar. Univ of California Press.

[9] Li S, Lin C Y, Song Y I, Li Z. (2013). Comparable entity mining from comparative questions. Knowledge and Data Engineering, IEEE Transactions on, 25(7): 1498-1509.

[10] Hu M, Liu B. (2006). Opinion Feature Extraction Using Class Sequential Rules. In AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs, pp: 61-66.

[11] Hu M, Liu B. (2004). Mining opinion features in customer reviews. In American Association for Artificial Intelligence, 4(4): 755-760.

[12] Shi B, Chang K. (2006). Mining chinese reviews. In Sixth IEEE International Conference on Data Mining-Workshops, pp: 585-589.

[13] Vinodhini G, Chandrasekaran R M. (2012). Sentiment analysis and opinion mining: a survey. International Journal, 2(6): 282-292.

[14] Moraes R, Valiati J F, Neto W P G. (2013). Document-level sentiment classification: An empirical comparison between SVM and ANN. Expert Systems with Applications, 40(2): 621-633.

[15] Pang B, Lee L. (2008). Opinion mining and sentiment analysis. Foundations and trends in information retrieval, 2(1-2), 1-135.

[16] Huang X, Wan X, Yang J, Xiao J. (2008). Learning to identify comparative sentences in Chinese text. In PRICAI 2008: Trends in Artificial Intelligence, pp: 187-198.

[17] He S, Yuan F, Wang Y. (2012). Extracting the comparative relations for mobile reviews. In IEEE on Consumer Electronics, Communications and Networks, pp: 3247-3250.

[18] Maks I, Vossen P. (2012). A lexicon model for deep sentiment analysis and opinion mining applications. Decision Support Systems, 53(4): 680-688.

[19] Cambria E, Schuller B, Xia Y, Havasi C. (2013). New avenues in opinion mining and sentiment analysis. IEEE Intelligent Systems, (2): 15-21.

[20] Wang H W, Wang W. (2014). Product weakness finder: an opinion-aware system through sentiment analysis. Industrial Management & Data Systems, 114(8): 1301-1320.