A Unified Framework for Emotional Elements Extraction Based on Finite State Matching Machine*

Yunzhi Tan, Yongfeng Zhang, Min Zhang, Yiqun Liu, and Shaoping Ma

State Key Laboratory of Intelligent Technology and Systems,
Tsinghua National Laboratory for Information Science and Technology,
Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China
cloudcompute09@gmail.com

Abstract. Traditional methods for sentiment analysis mainly focus on the construction of emotional resources based on the review corpus of specific areas, and use phrase matching technologies to build a list of product feature words and opinion words. These methods bring about the disadvantages of inadequate model scalability, low matching precision, and high redundancy. Besides, it is particularly difficult to deal with negative words. In this work, we designed a unified framework based on finite state matching machine to deal with the problems of emotional element extraction. The max-matching principal and negative words processing can be integrated into the framework naturally. In addition, the framework leverages rule-based methods to filter out illegitimate feature-opinion pairs. Compared with traditional methods, the framework achieves high accuracy and scalability in emotional element extraction. Experimental results show that the extracting accuracy is up to 84%, which has increased by 20% comparing with traditional phrase matching techniques.

Keywords: sentiment analysis, e-commerce, emotional elements extraction, finite state matching machine.

1 Introduction

With the rapid development of the Internet, E-commerce is becoming an increasingly popular network application. According to the 31th statistic report of China Internet Network Development State [1] from CNNIC (China Internet Network Information Center) in Jun 2013, there are 242 million internet users with online shopping experience up to the end of 2012. This is 20% of the total internet users. Compared with the year of 2011, the number has increased by 48.07 million with an increasing rate of 24.8%. At the same time, more and more internet users write reviews about product features and using experience. Sentiment analysis for customer reviews can help understand users' feedback about customers' attitude to each feature of a product timely and effectively. As a result, sentiment analysis for customer reviews is becoming a core technic that e-commerce websites are scrambling to develop.

* This work was supported by Natural Science Foundation (60903107, 61073071) and National High Technology Research and Development (863) Program (2011AA01A205) of China.

G. Zhou et al. (Eds.): NLPCC 2013, CCIS 400, pp. 60–71, 2013.

A practical review sentiment analysis system has two very important parts: one is to build emotional resources based on user comments (e.g. sentiment lexicon); the other is to analyze customer reviews and extract emotional elements based on the constructed emotional resource. There has been a lot of work [2-11] studying the issue of building a high-quality emotional resource, however there is little work focusing on how to use the constructed emotional resources to analyze user comments, extract emotional elements and summarize emotional tendencies of customers to product features. Traditional methods for emotional elements extraction are usually based on simple string matching techniques, which bring about the disadvantages of inadequate model scalability, low matching precision, and high redundancy. Besides, it is particularly difficult to deal with negative words. Our work is an attempt to fill this gap and improve the performance of emotional elements extraction using the emotional resource constructed.

In this work, we proposed a unified framework based on finite state matching machine for emotional elements extraction. We integrated max-matching principle and negative adverbs processing into the framework. Furthermore, the framework leverages rule-based methods to filter out illegitimate feature-opinion pairs. In our framework, emotional elements extraction is divided into three steps: emotional elements matching, extracting and filtering. Experimental results show that the accuracy of emotional elements extracted is up to 84%, which is an increase of more than 20% compared with traditional phrase matching techniques. On the other hand, the redundancy of emotional elements extraction is reduced a lot. In addition, the framework we proposed has a good scalability in that one can easily integrate new rules to deal with more complex customer reviews. Sophisticated evaluation and experiments verified both the effect and efficiency of our framework.

The main contribution of this work is that, we propose a unified framework based on finite state matching machine, which extracts emotional elements from customer reviews based on an emotional resource constructed. With this framework, the practicality of sentiment analysis is improved a lot.

2 Related Work

Automatic sentiment analysis for customer reviews is an important domain in web data mining. Some previous work [2-4] discussed how to calculate sentiment score for a whole sentence for the purpose of review information retrieving and classifying. Generally speaking, a sentence may contain more than one features and more than one sentiment polarities (e.g. "电视画面清晰但是价格太高了" , translated as "The picture is clear but the price is too high", contains two different sentiment polarities for the two product features "Picture" and "Price"). Sentence-level sentiment analysis can only determine whether a whole sentence is positive, negative or neutral. It cannot reveal the customer attitude towards specific product features exactly. Kim et al. [5] proposed that sentiment can be represented within four aspects, which are topic, opinion-holder, sentiment-descriptor and sentiment-polarity. That is to say, the opinion-holder uses some descriptors containing sentiment polarity to express his or her opinion about a topic. As a result, various work attempts to construct emotional resources constituting the four aspects for further sentiment analysis. Liu et al. [6] manually annotated features of products based on part-of-speech tagging. Then they used association rules to mine the association of feature words and opinion words. Li et al. [7] proposed a method using semantic relations of feature and opinion words to mine the relationship between them, which achieved good performance. There is much similar work [8-11] discussing emotional resource construction based on customer reviews, which also achieved good results. It is noteworthy that there is little work focusing on systematic research about how to use the constructed emotional resource to analyze reviews, extract emotional elements and summarize customer attitudes to the features of products.

Based on the constructed emotional resource, traditional emotional elements extraction methods use string-matching techniques. They extract emotional elements directly if an element exists in the emotional resource (e.g. Sentiment Lexicon), which face the disadvantages of low accuracy, poor scalability, redundancy and the difficulty to deal with negative words, because customer reviews are usually very colloquial and flexible. These shortages lower the practicality of the traditional methods. In this work, we designed a unified framework for emotional elements extraction based on finite state matching machine to meet these challenges. We integrate the max-matching principal and negative words processing into the unified framework. In addition, we built a set of rules to determine whether a feature-opinion pair is reasonable and filter it out if not. With our framework, the precision of emotional elements extraction has been improved a lot, with a much lower extraction redundancy. Furthermore, the unified framework is flexible enough to add new rules to deal with more complex customer reviews, such as reviews containing comparative forms or adverbs of degree. In this work, we also evaluate the performance of our framework in detail over various aspects.

3 Emotional Resource Construction

In this work, we use the method proposed in [7] to construct emotional resources. Then with the constructed emotional resource, we implement our unified framework for emotional elements extraction and evaluate its performance.

In the method proposed in [7], the emotional resource constitutes of various feature-opinion pairs attached with the corresponding sentiment polarity. The method mainly consists of three steps: feature words extraction, opinion words extraction and sentiment polarity labeling of feature-opinion pairs. In the step of feature words extraction, the method extracts a set of feature word candidates based on part of speech rules. In this step, the method also takes the words out of vocabulary into consideration. It then filters noise feature words from the candidate set based on their co-occurrence with adjective words and their frequencies in background corpus. In the second step, the method extracts opinion words based on part of speech, statistics information and word dependencies. In the last step, namely the sentiment polarity labeling of feature-opinion pairs, the method takes context and the distance between tagged feature-opinion pairs with untagged feature-opinion words into consideration.

In addition, it also uses global information to calculate the sentiment score of each feature-opinion pair iteratively.

In this work, our main consideration is not the process of emotional resource construction. In fact, we used the method in [7] to construct an emotional resource directly, and further use it to implement our unified framework for emotional elements extraction. Our framework for emotional elements extraction is introduced in the following section.

4 The Unified Framework Based on Finite State Matching Machine

In our framework, the process of emotional elements extraction is divided into 3 steps: Matching, Extracting and Filtering. In the first step, emotional elements matching, each review is mapped into a list of feature words, opinion words and negative adverbs, which are sorted by the order in which they occur in the original review. In the second step, the list will be the input of the finite state machine. We will find out the feature-opinion pairs according to the context and the sentiment lexicon. In addition the sentiment polarity of each pair is also determined in this step. Then in the last step, our framework judges the legality of each feature-opinion pair.

4.1 Emotional Elements Matching

Using traditional methods based on string matching for emotional elements extraction, there is much extracting redundancy in the results. For example, for the customer review “电视画面非常清楚(The TV picture is very clear)”, there are four feature-opinion pairs matched in the sentiment lexicon, which are “画面|清楚(picture | clear)”, “画面|清(picture | clear)”, “电视画面|清楚(TV picture | clear)” and “电视画面|清(TV picture | clear)”. Traditional methods based on string matching would extract all of the four feature-opinion pairs as emotional elements, which would bring about redundancy in feature-opinion pair extraction.

In order to solve the problem of extracting redundancy, we use the max-matching principle, which requires that the emotional elements would choose the max-length feature words and max-length opinion words if multiple feature or opinion words can be extracted. However, by taking advantage of the max-matching principle, we just extract “电视画面|清楚(TV picture | clear)” as the only emotional element.

Dealing with negative adverbs plays an important role in emotional elements extraction. A negative adverb could invert the sentiment of a review. In our framework, we build a list of words, which contains many popular negative adverbs, such as “不是(not)”, “没有(without)”, “不够(not enough)” and “不(no)”, etc. At the same time, we notice that there exist words that are not negative as a whole, but part of them may be negative adverbs. For example, “不是一般(not a little)” does not represent a negative sentiment, but “不是(not)”, which is a part of it, is exactly a negative adverb. As a result, we build a whitelist of words, which are not regarded as negative adverbs, such as “不是一般(not a little)”, “差不多(almost)” and so on.

By now, we have built three resources, which are sentiment lexicon, the list of negative adverbs and the negative adverbs whitelist. In the following parts, we will introduce the process of extracting emotional elements using these resources. In order to guarantee the precision of emotional elements extraction, we deal with these three resources in the order of negative adverbs whitelist, sentiment lexicon and the list of negative adverbs. For each customer review, the process of extracting emotion elements from it is like this:

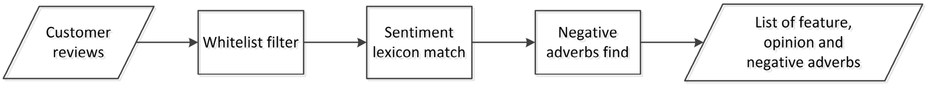

Fig. 1. Matching Process of Emotional elements

Each step in the figure above follows the max-matching principle. After the matching process of emotional elements, a review will be mapped into a list of feature words, opinion words and negative adverbs sorted by the order in the original review. For example, the review "画质不是一般的不清晰(The picture quality is not a little lack of clarity)" will be mapped into the list of words "画质(picture quality, feature word)", "不(not, negative adverb)" and "清晰(clarity, opinion word)", which is the input of the finite state machine in the next section.

4.2 Emotional Elements Extraction

For the purpose of judging whether a feature word and an opinion word is a pair according to the context, we design a finite state matching machine, which uses the resulting list of words in the previous step as input. The finite state matching machine is showed as follows:

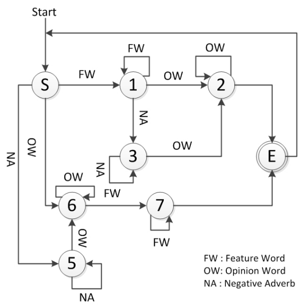

Fig. 2. The Finite State Matching Machine

When we design the finite state machine, we made two assumptions: 1) negative adverbs only occur in the front of opinion words or other correct negative adverbs; 2) when a user makes a review about more than one feature, he or she will put all the opinion words either in the front or in the back of the feature words. For example, a user may say "价格实惠画面也很清楚 (the price is reasonable and the picture is clear)", instead of "价格实惠且有清晰的画面 (The price is reasonable and the TV has a clear picture)". The two assumptions are in line with most of the user habits.

Our machine can deal with various complex conditions, such as the cases where one review contains more than one feature words, opinion words and negative adverbs. Using the list from step 1, the machine begins from state S and transforms its state according to the nature of words (feature, opinion or negative adverbs) in the list. When reaching the final state E, we get one or more than one feature-opinion pair(s). In the third step, we will further judge the legality of these feature-opinion pairs. We use the following three examples to make the process clearer:

1. The review "颜色鲜艳但是音效不是很好 (The color is bright but the sound effect is not very good)" contains feature words "颜色 (color)" and "音效 (sound effect)", opinion words "鲜艳 (bright)" and "好 (good)" and negative adverb "不是 (not)"; the process for this is S→1→ 2→E→1→3→2→E.
2. The review "不合理的价格 (It is not a reasonable price)" contains feature word "价格 (price)", opinion word "合理 (reasonable)" and negative adverb "不 (not)"; the process is S→5→6→7→E.
3. The review "时尚大方的外观 (Stylish and elegant appearance)" contains feature word "外观 (appearance)" and opinion words "时尚 (stylish)" and "大方 (elegant)"; the process is S→6→6→7→E.

4.3 Emotional Elements Filtering Based on Rules

There are still some errors in emotional elements extraction after the two steps introduced before, due to the fact that customer reviews are usually very colloquial and flexible. For example, "京东的售后服务真的很棒 (The service of JingDong is really great)" will be extracted into two feature-opinion pairs, which are "服务|真 (service | really)" and "服务|棒 (service | good)", where "服务|真 (services | really)" is not a correct pair. As a result, we build a set of rules to judge whether a feature-opinion pair is legal or not. We take four factors into consideration, which are the following:

1. The order of feature word and opinion word. There are some opinion words which could occur both in the front and in the back of a feature word, for example, "画面清楚 (the picture is clear)" an "清楚的画面 (clear picture)" are both legal. However, some opinion words can only occur either in the front or in the back, for example, people say "使用复杂 (complex usability)", and seldom do they say "复杂使用 (usability complex)".
2. The length of opinion words. The shorter an opinion word is, the more likely it's a part of another phrase. For example, "大 (big)" is a part of "强大(powerful)".
3. The distance between feature words and opinion words. Generally speaking, people tend to put opinion words close to feature words when writing a review. For example, people prefer saying "价格实惠 (The price is reasonable)" directly to inserting many other words between "价格 (price)" and "实惠 (reasonable)".

4. The probability that a feature word and an opinion word is a pair. We use the frequency of co-occurrence of a feature word and an opinion word to measure this probability, which is counted when constructing the sentiment lexicon.

Taking the four factors into consideration, the process of judging whether a feature-opinion pair is legal or not is shown in the following pseudo code:

```
Boolean Accept (FeatureWord, OpinionWord)
IF FeatureWord Occurs Before OpinionWord:
    IF (BeforeTime == 0) OR (Distance > ReviewLength/2 AND BeforeTime < β AND
    OpinionLength==1):
        RETURN False
ELSE:
    IF AfterTime==0 AND OpinionLength <3 AND (BeforeTime < γ OR Distance >
    ReviewLength/2):
        RETURN False
RETURN True
```

In the pseudo code, "BeforeTime" and "AfterTime" represent the times when the feature word occurs before and after the opinion word, correspondingly. "Distance" represents the distance between the feature and the opinion word. "ReviewLength" and "OpinionLength" indicate the length of the whole review and the opinion word. "β" and "γ" are two thresholds, which should be adjusted according to the size of the corpus used to construct the sentiment lexicon. We set β=10, γ=5 in our framework.

Now we give the whole procedure of our framework below (F-O: feature-opinion):

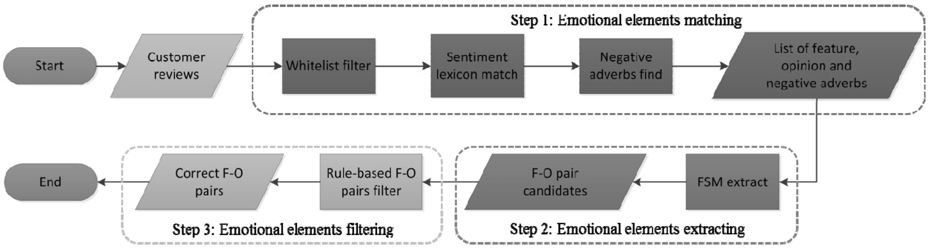

Fig. 3. Diagram of the whole framework for emotional elements extraction

We will process each customer review following the three steps above strictly. A feature-opinion pair is extracted if and only if the function "Accept" in the third step returns "True". If the function "Accept" returns "False" or the finite state machine fails to transform from one state to another, we regard these as wrong feature-opinion pairs and ignore them.

Compared with the traditional method based on string matching, our framework not only achieves a high precision, but also lowers the redundancy a lot in emotional elements extraction. In the next section, we will evaluate the effect in detail.

5 Evaluation for the Unified Framework

5.1 Data Preparation

We crawled 65549 customer reviews of 340 television products from TaoBao and JingDong. These reviews are mainly about five TV brands, which are Samsung, LG, Hisense, TCL and Skyworth. There are 263 smart television products and 77 non-smart television products. We randomly selected 80% of these reviews and used them to construct our emotional resource, resulting in a sentiment lexicon with 3800 entries in smart television domain. Each entry contains a feature-opinion pair and a the corresponding sentiment polarity, for example, "{1} 画面|细腻({1} Picture | smooth)" indicating that feature word "画面(Picture)" has a positive sentiment when described by the opinion word "细腻(smooth)". The remaining 20% of the user reviews is used to test the performance of our framework for emotional element extraction.

5.2 Accuracy of Emotional Elements Extraction

Based on the sentiment lexicon and the remaining 20% of customer reviews, we evaluated the performance of emotional elements extraction of our framework. In this section, we make a comparison among the following three methods:

1. **TSM (Traditional String Matching):** this is the traditional method for emotional element extraction, which extracts all the feature-opinion pairs that occur in both the sentiment lexicon and the reviews.
2. **FSMM (Finite State Matching Machine):** this method is our framework without the third step of rule-based feature-opinion pair filtering. A feature-opinion pair is extracted as long as the finite state matching machine reaches the final state.
3. **TUF (The Unified Framework):** this method is our final unified framework. Compared with the method of FSMM, rule-based feature-opinion pairs judging and filtering technic is added into the method of TUF.

We run the experiment three times independently for the purpose of more accurate evaluation. In each experiment, we randomly selected 10% of customer reviews from the remaining 20% user reviews. Then we use the three methods above to extract emotional elements. At last, we manually annotated each emotional element into three categories, which are explained below (M=1 means the feature-opinion pair is matched correctly and P=1 means the polarity labeling is conducted correctly):

- **M=1, P=1:** Both feature-opinion pair matching and polarity labeling is correct. For example, "价格|实惠 (price | reasonable)" is extracted from "电视价格实惠 (Price of the TV is reasonable)", and the polarity is labeled as positive.
- **M=1, P=0:** The feature-opinion pair matching is correct, but the polarity is labeled incorrectly. For example, "使用|满意 (usability | satisfactory)" is extracted form "电视使用满意 (usability of the TV is satisfactory)", which is a correct feature-opinion pair, but the sentiment polarity is labeled as negative.

- **M=0:** The feature-opinion pair matching is incorrect, no matter whether the polarity is correct or incorrect. For example, "服务|真 (service | really)" is extracted from "服务真不错 (the service is really good)".

We annotated 14932 emotional elements in total, and calculated the proportion of each category in each run. The results are shown below:

Table 1. Accuracy of emotional elements extraction

Run No.	Method	M=1, P=1	M=1, P=0	M=0
1	TSM	0.5880	0.0619	0.3501
	FSMM	0.8088	0.0416	0.1496
	TUF	**0.8392**	**0.0407**	**0.1201**
2	TSM	0.6080	0.0654	0.3266
	FSMM	0.8238	**0.0421**	0.1341
	TUF	**0.8404**	0.0436	**0.1160**
3	TSM	0.6062	0.0629	0.3309
	FSMM	0.8241	**0.0453**	0.1306
	TUF	**0.8405**	0.0465	**0.1130**

In our experiments, we regard the category of M=1, P=1 as correct extractions. We can see that FSMM and TUF can achieve much higher accuracies in this category compared with TSM. Compared with TSM, FSMM increases the extracting accuracy by 23.4% at most and lowers the proportion of incorrect matching by about 20%. Compared with FSMM, we see that in the unified framework (TUF), the rule-based feature-opinion pairs filtering technique can further reduce the matching error rate, which contributes to the improvement of accuracy.

In addition, the unified framework we proposed reduces the emotional elements extraction redundancy, and has a good performance in negative adverbs processing. We evaluate the performance in these two aspects respectively next.

5.3 Reduce the Redundancy of Emotional Element Extraction

Many redundant emotional elements may be extracted using traditional methods based on string matching. The unified framework we proposed has a good effect in reducing redundancy, some typical examples are shown as follows:

1. **Example 1:** If we use traditional methods, the customer review "功能强大 (the function is powerful)" will be extracted as "功能|强大 (function | powerful)", "功能|强 (function | strong)" and "功能|大 (function | big)"; but in our framework, only "功能|强大 (function | powerful)" will be extracted because it meets the max-matching principle.
2. **Example 2:** For the review "亮度好价格也很低 (brightness is good and price is low too)", four feature-opinion pairs are extracted using traditional method, which are "亮度|好 (brightness |good)", "价格|低 (price | low)", "亮度|低 (brightness |

low)” and “价格|好 (price | good)”, which contains two redundant pairs. However, only “亮度|好 (brightness | good)” and “价格|低 (price | low)” are extracted using our unified framework, because in our finite state matching machine, a feature-opinion pair will not be re-extracted as long as it reaches the final state.

As a result, the unified framework for emotional elements extraction based on finite state matching machine can not only reduce redundancy (Example 1), but also reduce the number of incorrect extractions (Example 2).

Furthermore, we counted the number of emotional elements extracted using different methods in each experiment. To compare with the traditional method TSM, we calculated the *Redundancy Reduction Rate* for method FSMM and TUF. The *Redundancy Reduction Rate* is defined as $R = (\#\mathrm{E} - \#\mathrm{E}^{'}) / \# E$, where #E represents the number of emotional elements extracted using TSM, and $\#\mathrm{E}^{'}$ represents the number of emotional elements using FSMM or TUF. The result is shown in Table 2 below:

Table 2. Redundancy Reduction Rate using different methods

	Experiment 1			Experiment 2			Experiment 3		
	TSM	FSMM	TUF	TSM	FSMM	TUF	TSM	FSMM	TUF
#E/#E’	2211	1517	1449	2033	1447	1397	2034	1455	1389
R	--	31.4%	34.5%	--	28.8%	31.3%	--	28.5%	31.3%

From the result above, we see that redundant emotional elements extracted can be reduced a lot using the unified framework we proposed. Compared with TSM, FSMM reduced the redundancy by 31.4% at most and TUF reduced by 34.5% at most. The underlying reason is that the max-matching principle made it possible to extract only the feature-opinion pairs of maximum length, which are most likely to be correct.

5.4 Evaluation of Negative Adverbs Processing

The traditional methods for emotional elements extraction based on string matching cannot deal with negative adverbs well, as they usually cannot take advantage of the rich context information of negative adverbs in customer reviews. In our framework, negative adverbs processing can be easily integrated into the finite state matching machine. With the context information provided by the finite state machine, we achieved good performance in negative adverbs processing. For example, “功能确实不少 (The functions are not few indeed)” can be relabeled as positive from negative by considering the negative adverb “不(not)”.

In addition, the number of feature-opinion pairs whose polarities are changed by taking negative adverbs into consideration in our unified framework is 2290; and the accuracy of these changes is 88.6%.

Furthermore, we noticed that the polarities of some entries in the sentiment lexicon are labeled incorrectly, although the feature-opinion pairs extracted are correct. In order to evaluate the pure performance of negative adverbs processing, we made an evaluation under the hypothesis that the polarities are always correct, although it is not true for some entries. Following this hypothesis, if the negative adverbs processing gives a correct sentiment invert, we annotate it as M=1 and P=1, regardless of whether the polarity itself is correct or not. We randomly selected 10% of the reviews containing negative adverbs at each time and conducted three times of evaluation independently, the results are shown in the table below:

Table 3. Evaluation results of negative adverbs processing

Experiment No.	Use the hypothesis?	M=1,P=1	M=1,P=0	M=0
No.1 (394 reviews)	No	0.8219	0.0585	0.1196
	Yes	0.8626	0.0178	0.1196
No.2 (351 reviews)	No	0.8429	0.0429	0.1142
	Yes	0.8629	0.0229	0.1142
No.3 (347 reviews)	No	0.8353	0.0434	0.1213
	Yes	0.8671	0.0116	0.1213

The performance of negative adverbs processing without the hypothesis is satisfactory, which is similar to the overall performance in Table 1. This indicates that our framework performs well in terms of stability. On the other hand, under the hypothesis that the sentiment polarities are always correct in the sentiment lexicon, the precision of negative adverbs processing is more than 86%, which is about a 2% increase from that without the hypothesis. It indicates that if the quality of the sentiment lexicon were better, our unified framework would also achieve better performance, which is in good consistency with sentiment lexicon construction techniques used.

Note that, although we used many methods to improve the performance of emotional elements extraction, the proportion of emotional elements extracted where M=0, namely the incorrect extractions, is still about 10%. After analyzing the bad cases, we find that most of them are in need of leveraging semantic and knowledge information to 'understand' the meaning of a user review correctly, such as "价格当当最便宜 (The price of DangDang is the lowest)". As a result, future attempt would be devoted to the use of semantic and knowledge information in our framework to further enhance the performance of emotional elements extraction.

6 Conclusion

In this work, we proposed a unified framework based on finite state matching machine for emotional elements extraction. We integrated max-matching principle and negative adverbs processing into the framework. Furthermore, the framework leverages rule-based methods to filter out illegitimate feature-opinion pairs. Compared with traditional methods, the framework achieves high accuracy and scalability in emotional elements extraction. In addition, the framework has a good scalability in which one can easily integrate new rules to deal with more complex customer reviews. Sophisticated evaluation and experiments verified both the effect and efficiency of our framework.

In the future work, we would add more rules to our emotional elements extraction framework in order to deal with more complex reviews, such as reviews containing comparative forms or adverbs of degree. Furthermore, we would introduce semantic and knowledge information into our framework to further enhance the performance of our framework.

References

1. 中国互联网络信息中心(CNNIC). 第31次中国互联网络发展状况统计报告 1 (2013), http://www.cnnic.cn/hlwfzyj/hlwxzbg/hlwtjbg/201301/P020130122600399530412.pdf
2. Yu, H., Hatzivassiloglou, V.: Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In: Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing, pp. 129–136. Association for Computational Linguistics (2003)
3. Ku, L.W., Liang, Y.T., Chen, H.H.: Opinion extraction, summarization and tracking in news and blog corpora. In: Proceedings of AAAI-2006 Spring Symposium on Computational Approaches to Analyzing Weblogs (2001, 2006)
4. Dave, K., Lawrence, S., Pennock, D.M.: Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In: Proceedings of the 12th International Conference on World Wide Web, pp. 519–528. ACM (2003)
5. Kim, S.M., Hovy, E.: Determining the sentiment of opinions. In: Proceedings of the 20th International Conference on Computational Linguistics, p. 1367. Association for Computational Linguistics (2004)
6. Liu, B., Hu, M., Cheng, J.: Opinion observer: analyzing and comparing opinions on the Web. In: Proceedings of the 14th International Conference on World Wide Web, pp. 342–351. ACM (2005)
7. 李智超，面向互联网评论的情感资源构建及应用研究. 北京：清华大学计算机科学与技术系， 4 (2011)
8. Hu, M., Liu, B.: Mining opinion features in customer reviews. In: Proceedings of the National Conference on Artificial Intelligence, pp. 755–760. AAAI Press, MIT Press, Menlo Park, Cambridge (1999, 2004)
9. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 168–177. ACM (2004)
10. Popescu, A.M., Etzioni, O.: Extracting product features and opinions from reviews. In: Natural Language Processing and Text Mining, pp. 9–28. Springer, Heidelberg (2007)
11. Hiroshi, K., Tetsuya, N., Hideo, W.: Deeper sentiment analysis using machine translation technology. In: Proceedings of the 20th International Conference on Computational Linguistics, p. 494. Association for Computational Linguistics (2004)