A Method of Fine-grained Emotion Analysis Based on Semantic Analysis Network

Bopeng Liu¹, Hong Yang¹, Zhengkui Lin¹, Menglu You¹ and Zhaohui Wang¹

¹School of Computer Science and Technology, Dalian Maritime University
dalianjx@163.com

Abstract. Sentiment Analysis also known as the Tendency Analysis, it is the emotional color of the subjective text analysis, processing, induction process. The traditional text emotion analysis research mainly faces the chapter and the sentence level, realizes the corresponding emotional polarity judgment. Based on the characteristics of Chinese language commentary language expression, this paper uses a semantic analysis based on the combination of attribute words and emotional words to identify the extraction method, and the extraction method and the method of attribute words combined with a method based on semantic analysis Network Review Fine-grained Emotion Analysis.

1. Introduction
With the increase of the number of comments on e-commerce sites and commodity attributes, the traditional commentary emotional analysis has been unable to meet the needs of shops or consumers. Buyers want to get products that match their own preferences from the product reviews. The merchant wants to get the commentary's comments on the product from the comments, that is, the buyer's comments show the degree of preference for the goods and improve their competitiveness.

2. related work and problems
A class of goods have a lot of attribute, fine-grained emotional polarity classification focuses on finding these attributes from the comments and judging the emotional tendencies of consumers for each attribute. In practice, fine-grained emotional polarity classification is significant for consumers to improve consumer behavior, product performance and service. The main research work focuses on extracting keywords and emotional words.

The product attributes are divided into explicit and implicit attributes through Mingqing Hu, Bing Liu's [1] method. Explicit phrases with explicit attributes as attribute words. The implicit attribute does not appear in the comment as a noun or noun phrase that describes the performance of the purchased item, which is obtained by way of a semantic understanding of the statement before or after it. Kobayash [2] build product attributes for cars and games, each of which is represented by a triplet <Subject, Attribute, Value> and then manually tagged it.

In general, the research of emotion analysis is mainly analysis of the emotional polarity of coarse-grained users, and the research on fine-grained emotional polarity classification for product or service attributes is relatively few. Therefore, this paper based on the literature [3] to further study the network assessment of fine-grained emotional analysis.
3. Fine-grained emotion analysis corpus pretreatment
Fine-grained emotional analysis corpus preprocessing mainly includes Chinese word segmentation and part of speech annotation. This paper adopts the Language Technology Platform (LTP) of Harbin Institute of Technology to carry out data preprocessing research. In terms of part-of-speech, LTP uses 863 parts. Examples of processing results are shown in Table 1 below. The English meaning is “Often to eat, It's super good, cost-effective, Complete variety, taste delicious. Everyone is having a good time”.

| Comments before corpus processing | Comments After the corpus preprocessing |
|----------------------------------|----------------------------------------|
| Often to eat, It's super good, cost-effective, Complete variety, taste delicious. Everyone is having a good time | eat/v | super/b | good/a | cost/d | effective/a |
|                                  | complete/n | taste/n | delicious/a | everyone/d |
|                                  | eat/v | have/d | good time/a |

4. A Method of Joint Extraction of Attribute Words and Emotional Words Based on Semantic Analysis
Emotional words in the modified attribute words expressed in a more flexible way, it's speech and syntax are very rich. In view of these characteristics of network commentary, this paper combines the semantic and grammatical extraction rules put forward by Jiang Tengjiao and Tang Ming-feng [4]-[6] to extract a new attribute word in combination with word frequency and word distribution ratio extraction method.

4.1 Candidate attribute words - emotional words on the extraction method
In this section, the attribute word - emotional word pair is defined as a set of <attribute words, emotional words>. The main goal of the semantic role annotation is to identify and mark the semantic roles associated with the predicate to find the attribute words that the predicate is modified. However, the role of the general emotional word in the comment sentence is not only as a predicate, but also may be other elements of the element, at this time, you must expand the semantic role label. The meaning of each label is shown in Table 2.

| Annotated | Explanation | Annotated | Explanation |
|-----------|-------------|-----------|-------------|
| A0        | agent       | A1        | Accept the matter |
| A2        | Indirect role object | TMP | time |
| DIR       | direction   | DIS       | Chapter mark |
| EXT       | range       | LOC       | position |
| MOD       | General modification | NEG | negative |
| PRP       | proposal    | ADV       | Adverbs |

The semantic analysis of the commentary sentences in the LTP system can be briefly described by the following examples.
example: “The taste of Bashu hot pot is getting better and better now, and the long-term development prospect is good.” This sentence is marked by LTP, as shown in Figure 1 below. The English meaning is “The flavor of the Batu hotpot is now getting better and the long-term development prospects are promising”
In Figure 1, the word is separated by words and words, and the result of the annotation is expressed below the word. The interdependence between the words is shown in the figure's arc; the semantic role annotation results are shown below the part of speech.

Jiang Tengjiao in the literature proposed predicate attribute words meet the pre-rules 1, pre-rules 2 and pre-rules 3, respectively, the corresponding conditions of use and attribute word priority:

1. non-psychological verb as a predicate, the priority of A0 → A1 → A2.
2. psychological verbs as predicates, the priority of A1 → A2 → A0.
3. the three in the same sentence, the attribute word for a number of A0, A1 or A2 combination.

In the commentary, the emotional words can be presented as many corners, including adverbs, predicates, adverbs of the object, and attributive elements. Thus, the rule of extract attribute are shown in Table 3.

| rule        | Predicate state                                             | Attribute word - emotional word pair                      |
|-------------|-------------------------------------------------------------|----------------------------------------------------------|
| rule 1      | When the emotional word is a predicate in a semantic role   | <The predicate word, the emotional word>                 |
| rule 2      | When the emotional word is a predicate object, adverbial, or complement | <The word of the predicate + predicate                    |
| rule 3      | An adverbial adverb or adverb                               | <Predicate A0 + predicate + object, emotional word>      |
| rule 4      | The above rules have not yet found the emotional word of the word | The latest term of emotional words, emotional words      |

In order to facilitate the use of LTP platform, the system proposed the definition of 12 kinds of dependencies, used to analyze the sentence structure and the word collocation relationship, as shown in Table 4.

| Annotated Structural relationship | Annotated Structural relationship |
|----------------------------------|----------------------------------|
| ATT Fixed relations             | DE “A”structure                  |
| QUN Quantitative relationship   | DI “the”structure                |
| COO Constellation               | DEI “get”structure               |
| APP Cohabitation                | BA “hold”structure               |
| LAD Pre-attached relationship    | BEI “cover”structure              |
| HED core                        | SBV s-v relation                 |
4.2 Attributes and emotional words of the characteristics of the calculation method

With the candidate word pair, we can calculate the alignment probability of the attribute word to the emotional word according to the formula (1).

\[
P(o_a | o_e) = \frac{\text{count}(o_a, o_e)}{\text{count}(o_e)}
\]

(1)

Similarly, according to formula (2) to calculate the emotional word between the words to the probability of alignment.

\[
P(o_e | o_a) = \frac{\text{count}(o_e, o_a)}{\text{count}(o_a)}
\]

(2)

The number of times expressed and multiplied in the formulas (1) and (2) above. Through the probability of alignment and the noun / noun phrases and adjectives / verbs within the existence of an emotional relationship between the specific out, the formula as shown in (3)

\[
\text{Association}(w_e, w_a) = (\mu * p(w_e | w_a) + (1-\mu) * p(w_a | w_e))^{-1}
\]

(3)

The formula (3) is a harmonic factor that adjusts the two alignment probabilities that are set in this article.

In the commentary, the attribute word is the product itself; in addition, the most likely to be the attribute word in the commentary corpus is the evenly distributed candidate word with high frequency. To this end, we have selected five commentary corpus (sports, politics, humanities, science and technology, news commentary) in the field that is not the same as the topic "hot pot" in this paper.

The corpus of the comment in front of the text is written as C. In C, each comment can be treated as a separate class, and C contains n useful comments, that is, by calculating the information entropy of the comment, we can get the complexity of the whole system, but also can show the attributes The distribution of words in the document. The formula is as shown in (4):

\[
\text{IE}(t_i) = -\sum_{j=1}^{n} p(d_j, t_i) \log p(d_j, t_i)
\]

(4)

Equation (4) shows the probability that the candidate word appears in the comment, as shown in equation (5):

\[
p(d_j, t_i) = \frac{tf_i}{\sum_{j=1}^{n} f_i}
\]

(5)

Which means that the word frequency of the candidate attribute word in the jth comment, if only in a comment, then, then, so that, in order to simplify the calculation, the probability of inaccuracy is 0, so in equation (5) Add the constant factor and sum in the denominator with the previous denominator. The new probability formula is (6):

\[
p(d_j, t_i) = \frac{tf_i}{\sum_{j=1}^{n} f_i + 0.0001}
\]

(6)

"Person", "things" and other frequent nouns do not belong to the members of the word. That is, the frequency of occurrence of the attribute word is not necessarily high. On the contrary, the frequency of low vocabulary may also be a property word. In order to dig out more of the attribute words in the comments as much as possible, this paper is calculated by the corpus of different fields, combined with the document frequency. The specific formula is (7):

\[
D_s(t_i) = \begin{cases} 
\frac{\log(1 + df_{out})}{\log(1 + df_{in})} & \text{if } df_{out} \neq 0 \\
\alpha \times \log(1 + df_{out}) & \text{otherwise}
\end{cases}
\]

(7)
In equation (7), the frequency of the document in the commentary corpus is the frequency of the document in the corpus of the jth (5 in total) domain. (4) and the formula (7) to find the characteristics of the word word Indicator formula (8) as shown in the formula:

\[ I(t_j) = \frac{Ds(t_j) \times IE(t_j)}{m} \]  

In formula (8) \( Ds(t_j) = \sum_{j=1}^{m} Ds(t_j) / m \)

For the recognition of emotional words, this paper combines the word frequency of words and the distribution of words in the corpus to calculate the characteristic Indicator of candidate emotion words. The formula is (9):

\[ I(o_i) = \log(1 + df_i) \times D_i \]  

\[ D_i = \frac{D_i - \bar{D}_i}{\sqrt{\frac{\sum(D_i - \bar{D}_i)^2}{n-1}}} \]  

In the formula (9), we show the frequency of the candidate words in the corpus. The formula (10) indicates the distribution of the candidate words in the corpus. \( D_i \) is the frequency of the candidate word in the jth comment, indicating the average word frequency in the commentary corpus.

4.3 The Method of Exact Extraction of Attribute Words and Emotional Words

We can extract the emotional words and attribute words accurately and combine with the random walk algorithm of undirected bimodal and restart to construct a probabilistic model. The probability of extraction is higher than that of the former. Set the words of the threshold as emotional words or attribute words. The algorithmic algorithm is described as follows:

Suppose that s is the starting node of random walk, q is the current node, node q either returns to the starting node with probability c (ie, restart probability), or moves along probability (1-c) along the edge of the graph. Adjacent nodes, A is a standardized adjacency matrix, to calculate the relationship between the nodes in the graph and the node s (11):

\[ \bar{r} = c\bar{A}\bar{r} + (1-c)e \]  

Equation (11) is the starting vector, when the initial node of the walk is s, otherwise.

The bipartite graph G (V, E, A) is shown in Figure 2. V denotes the set of points in the bimodal, representing the emotional word, representing the attribute word. E represents the edge set in graph G, and if there is an emotional relationship between the attribute word in the graph and the emotional word, it is represented by an edge.

**Candidate emotional words**

**Candidate evaluation object words**

**Figure. 2** An example of a dichotomous graph of the relationship between the modeled emotional word and the attribute word.
A represents the emotional relationship between the two, calculated by the formula (3). The formula (12) represents the process of calculating the affective word and the attribute word, and iterates the formula (12) until the energy value state is stable.

\[
\begin{align*}
E(t) &= \lambda \times R \times E(o) + (1 - \lambda) \times I_t, \\
E(o) &= \lambda \times R \times E(t) + (1 - \lambda) \times I_o
\end{align*}
\]  

(12)

In the formula (12), the energy values of the attribute word and the emotional word are expressed by the sum. R denotes the relationship weight between the relation matrix, the i-th candidate attribute word and the j-th candidate emotion word. The characteristic vector representing the candidate attribute word is calculated by the formula (8). The characteristic vector representing the candidate emotion word is calculated by the formula (9). Is a harmonic parameter.

4.4 Analysis of experimental and experimental results

In this paper, the sample data set is marked by manual annotation to verify the accuracy of the extracted results, and compared with other extraction methods. The parameters in equation (12) determine the relationship between the characteristics of words and the emotional relationship between words. The exact rate, recall rate, and F-measure value are shown in Figure 3.

![Figure. 3 The graph different \( \lambda \) in the three values on the dataset](image)

Through the curve of Figure 3 can be seen when the effect of about 0.4 when the best. After the parameter values are determined, we compare our experimental results with Hu, IDR, EDR, IEDR methods (Our method). The experimental results are shown in Table 5.

| Method | Accuracy | Recall rate | F1 value |
|--------|----------|-------------|----------|
| Liu    | 0.60     | 0.85        | 0.71     |
| IDR    | 0.51     | 0.77        | 0.61     |
| EDR    | 0.53     | 0.63        | 0.58     |
| IEDR   | 0.55     | 0.81        | 0.66     |
| Our    | 0.63     | 0.85        | 0.72     |
From the comparison of the results of Table 5, with the hot pot shop comments as experimental corpus, respectively, the five groups of experiments, we can see in this method the highest accuracy. We found that the Hu method does not consider the distribution of words, domain correlation, candidate emotional words and other information, but only consider the importance of candidate attribute words, resulting in low accuracy. Therefore, the three important factors influencing the extraction of attribute words include the distribution of words, the information of field related information and the word information of emotional words.

5. Fine-grained Emotional Tendency Calculation Method for Attribute Words
In this paper, we find that if we have both the attribute words and the emotional words related to it in the sentence in the commentary corpus, then there is a specific relationship between the two. The commentary in the commentary is an attribute word or a category to which it belongs.

5.1 Emotional Review Unit Construction
After the pretreatment of the test corpus, the attribute words affect the word pairs, the degree adverbs and the negative words, and construct the corresponding emotional evaluation unit according to the form given below:

Emotional comment unit = <comment unit ID, store name, attribute word, emotional word, degree word, negative word>

(1) The extraction of degree adverbs
An adverb is an adverb that limits or modifies an adjective or adverb. In this paper, 219 degree words are selected in the knowledge network. Divided into six levels, the weight added to the degree adverbs. The degree of adverb and its weight setting examples and extraction algorithms are simplified. Table 6 and Table 7.

**Table.6** Degree adverb and its weight setting example
| Degree adverb         | Extremely weight setting instance |
|----------------------|----------------------------------|
| exceed(1.5), very(1.25), extremely(2), most(2), owe(0.5)……, slightly(0.8) |

**Table.7** Degree Adverb Extraction Algorithm

```
Algorithm illustrate: Look up the degree adverb in a given sentence
Algorithm input: CString
Algorithm output: CString
Algorithm description:

For each Paragraph in Document:
    For each Line Paragraph:
        For each Group in Line:
            For each Word in Group:
                if word in senDict:
                    senWord = (Position in sentence, Emotional tendencies, Emotional intensity)
                    LastSenWordPosition = 0 # The position of an emotional word in a sentence
                    for i in range(senWord[0], LastSenWordPosition+1):
                        if Group[1] in degreeDict:
                            DegreeWord = ((Position in sentence, Modification strength))
                            LastSenWordPosition = senWord[0]
```
5.2 Emotional Aggregation and Emotional Tendency Classification
In the process of analysis and research, this paper divides the comment sentence into seven categories: price, service, side dish, soup, seasoning, environment and location, and returns the extracted attribute words to the corresponding classification. The tendency of calculating the corpus in all the emotional tendencies of aggregated corpus is the analysis of coarse-grained emotion. The emotional tendencies of the unit or sentence are emotional analysis. Among them, the emotional unit of the emotional value of the formula is:

\[
\text{Emotional unit emotional value} = \text{negative word} - 1 \times \text{degree word weight} \times \text{emotional word weight}
\]

Because the negative words and the degree of adverbs in the sentence in the local differences for the emotional unit expressed by the degree of emotional will have different effects, when the negative word is located before the adverb adverb, the negative feelings of the negative word 0.5, when the negative word in the degree adverb after its Emotional weight take -1.

5.3 Analysis of experimental and experimental results
According to the fine-grained emotion analysis process, combined with the semantic analysis of the combination of attribute words and emotional words extraction method, the construction of emotional units, while the emotional value of the calculation, the data visualization of the results shown in Figure 4 and Figure 5 shows.

![Figure 4 Results of emotional analysis](image)
6. Summary and Prospect
In this paper, based on the semantic analysis of attribute words and emotional words jointly extract candidate attribute words and emotional word pairs, and calculate the word frequency and document ratio characteristics of attribute words and emotional words and phrases to further accurate calculation, and then the extraction method and orientation. The method of attribute word combination puts forward a semantic analysis based on fine-grained sentiment analysis method for web commentary. Finally, it verifies the effectiveness and superiority of this method over traditional methods. However, due to the diversity of natural language expressions, sometimes subjective views of critics do not appear clearly in sentences, but are concealed in the comments with hidden language. Further mining of implicit attribute words and emotions helps to more fully understand the author's subjective evaluation of goods or topics.

References
[1] Hu M, Liu B. Mining opinion features in customer reviews[C]. AAAI. 2004: 755-760.
[2] Kobayashi N, Inui K, Matsumoto Y. Collecting evaluative expressions for opinion extraction[M]. Lecture Notes in Artificial Intelligence, 2004(3248):596-605.
[3] ShiHanXiao. Research on Fine-grained Sentiment Analysis[D]: (Doctoral Dissertation). Suzhou: Soochow University, 2013.
[4] JIANU Tend Jiao WAN Chang-Xuan LIU De-Xi LIU Xi-Ping LIAO Uuo-Qiong
[5] Extracting Target-Opinion Pairs Based on Semantic Analysis[J]. Journal of computer science, 2017, 03: 617-633.
[6] TangMingShuang. Sentiment Analysis of Online Review Based on Machine Translation Model[D]. University of Electronic Science and technology, 2016.