

Opinion Annotation in On-line Chinese Product Reviews

Ruifeng Xu^{1,2}, Yunqing Xia³, Kam-Fai Wong¹, Wenjie Li²

Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong¹

Department of Computing, The Hong Kong Polytechnic University²

Center for Speech and Language Technologies, Tsinghua University³

E-mail: rfxu@se.cuhk.edu.hk, yqxia@tsinghua.edu.cn, kfwong@se.cuhk.edu.hk, cswjli@comp.polyu.edu.hk

Abstract

This paper presents the design and construction of a Chinese opinion corpus. Based on the observation on the characteristics of opinion expression in Chinese online product reviews, which is quite different from in the formal texts such as news, an annotation framework is proposed to guide the construction of an opinion corpus based on online product reviews. The opinionated sentences are manually identified from the review text. Furthermore, for each comment in the opinionated sentences, its 13 describing elements are annotated including the expressions related to the target product attributes and user opinion expressions as well as the polarity and degree of the opinions. Currently, 12,724 comments are annotated in 10,935 sentences from product reviews. Through statistical observation on the opinion corpus, some interesting characteristics of Chinese opinion expression are presented. This corpus is helpful to support systematic research on Chinese opinion analysis.

1. Introduction

Aiming at identifying and analyzing the opinions in text, opinion analysis becomes an increasingly interesting research topic in information extraction and knowledge discovery areas. The discovered opinions are useful to many applications. For example, the opinions on products are helpful to customer purchase decision and manufactory quality improvement (Hu & Liu, 2004), while the opinions on specific policies from different sources are helpful to improve government management. Besides, as a fundamental natural language processing technology, the opinion analysis technique helps to promote research in information extraction and knowledge discovery such as automatic summarization (Hu & Liu, 2006) and question & answer (Yu & Hatzivassiloglou, 2003).

Many researches on opinion analysis have been reported in the recent decades (Hatzivassiloglou & McKeown 1997; Pang et al. 2002; Gordon et al. 2003; Wiebe et al. 2004; Xia et al. 2007]. Most of them focused on the subjective words extraction [Hatzivassiloglou & McKeown 1997] and opinion classification at the document [Yu & Hatzivassiloglou 2003] or sentence level [Riloff et al., 2003]. However, identifying only opinionated documents or sentences may not be sufficient [Wiebe et al., 2005]. Especially, a practical opinion analysis system for product reviews is expected to provide not only the positive or negative a comment is, but also the attributes targeted. Deep opinion analysis including the opinion holder identification and opinion target analysis therefore becomes a hot research topic [Seki, 2007]. Naturally, an opinion annotated corpus is a valuable resource to support the research in this area. However, there is few reported work on Chinese opinion corpus

construction (Ku et al. 2005). Especially, up to now, there is no published Chinese opinion corpus on product reviews. Furthermore, unlike the formal text such as news, on-line product reviews are always written in free style. The annotation scheme for this kind of informal text is naturally different from the one designed for formal text.

In this study, based on the observation on the characteristics of opinion expression in Chinese product reviews, an annotation scheme is proposed. In this scheme, the annotation granularity is set to each comment in the opinionated sentences. For each comment, both the opinion expression and its targeted product attributes are annotated. Domain opinion ontology, which maintains the concept nodes of the most discussed product attributes, is introduced to guide the annotation of target attributes. As for the user opinions, their expression segments, opinion keywords and polarity are annotated as well as negation or modifier if applicable.

Currently, 1,100 and 500 review documents on digital camera and mobile phone products are annotated, respectively. Two annotators identified 7,864 opinionated sentences and annotated 12,724 comments in these sentences to construct a Chinese product review opinion corpus (acronymed *CPro* corpus). The observation on the CPro corpus discovered some characteristics related to opinion expression in Chinese on-line product reviews. It is shown that CPro corpus is helpful to opinion analysis research.

The rest of this paper is organized as follows. Section 2 presents the definitions and the annotation scheme. Section 3 describes the practical issues in the corpus annotation including raw corpus preparation, domain ontology preparation, annotation flow and quality assurance mechanisms. Section 4 gives current status of

C PRO corpus and evaluates the annotation agreement. Section 5 presents the characteristics of opinion expressions in Chinese product reviews based on the observation on C PRO corpus. Finally, Section 6 concludes this paper.

2. Definitions and Annotation Scheme

2.1 Definitions

In this study, an **opinion** is defined as a person's ideas and thoughts towards a product. It is an assessment, judgment or evaluation of a specific product or its targeted attributes. An opinion is not a fact, on the contrary, it is subjective assessment. Normally, an opinion carries the polarity of positive, negative or neutral. The sentence having opinions is defined as **opinionated sentence**. An opinionated sentence must have at least one comments, where a **comment** is defined as the smallest complete opinion expression having only one opinion and one targeted product attribute. If an opinionated sentence having two opinions for one target attribute or one opinions for two target attributes, they are regarded as two comments. For example, in a opinionated sentence,

Example 1.

屏幕大而明亮 (*The LCD screen is large and bright*)

The user thought the “LCD screen is large” and “LCD screen is bright”. Thus, this sentence has two comments. Both of the target attributes are “屏幕”(*LCD screen*) while the opinion expressions are “大”(*large*) and “明亮”(*bright*), respectively.

Similarly, in another example sentence,

Example 2.

按键和按钮很粗糙(*The key and button are rough*)

Two comments, “the key is rough” and “the button is rough” are identified. The target attributes are “按键”(*key*) and “按钮”(*button*), respectively, while the opinion expressions are both “粗糙”(*rough*).

2.2 Annotation Scheme

There are few reported work on the construction of Chinese opinion corpus. (Ku et al. 2005) developed a high-quality Chinese opinion corpus on the formal news text. This corpus annotates five key elements of an opinion, including:

opinion segment (the scope of one opinion in the text), **opinion source** (also named opinion holder. It is the governor of an opinion and normally refers to a person, a state or an organization.)

opinion operator (the keyword of expressing one opinion. It is always the verb indicating an opinion event)

sentiment keyword (also named opinion keyword, which is the keyword reflects the opinion polarity, i.e.

positive, neutral or negative)

sentence attitude (opinion polarity of a sentence) as well as one optional element, **negation** (negation expression in the sentence).

However, opinion expressions in product reviews have shown their own characteristics different from the ones in news text. Firstly, in forums, experienced customers always wrote down their reviews in structured Web tables. Thus, opinion sources (opinion holders) and opinion operators are always neglected in the review text. Secondly, opinion expressions in the forum reviews always lack strict syntax. Thirdly, it is common to find two or more opinions co-exist in a single sentence without obvious gap. Fourthly, the words used in the on-line reviews are flexible, especially the opinion keywords. The users used abundant opinion keywords including many speaking words. Furthermore, some opinion keywords have constant polarity irrespective of context and thus they are named **context-independent opinion word**, while much more opinion keywords may carry different polarity which is determined by the context and its targeted attributes (they are named as **context-dependent opinion words**). Finally, implicit metaphors are widely existed which increases the difficulty of opinion analysis.

An annotation scheme is proposed in this study to guide the corpus construction. Firstly, only provide the information of opinion polarity of one product is not enough to a practical opinion mining system because different kinds of users have their own interested attributes. For example, the female customers always care the weight of a digital camera while the male customers always do not care. Thus, both the opinion expression and its targeted attributes are expected. So, the annotation granularity in this study is set to each comment in the opinionated sentence, i.e. the pair of one opinion expression and one targeted product attribute.

Secondly, the explicit expression and implicit metaphors in the opinionated sentences are distinguished. Implicit metaphor is an interesting topic in opinion analysis. An example sentence is given below,

Example 3.

只要按下快门, 周围 20 米的人都知道你用的 Nikon D60.

(*Once you press the shutter, everyone within 20 meters will know your camera is Nikon D60*)

In this sentence, the customer uses implicit metaphor to criticize the shutter of Nikon D60 is too noisy in spite of the surface meaning of this sentence never involve the noise or shutter. The annotation of implicit metaphor is helpful to linguistic study.

Thirdly, the visibility of opinion expressions and target attributes are included in the opinion annotation.

Normally, a comment has one opinion expression and one target attribute. However, sometimes the target attribute is not given in the text and it can be recognized by observing the opinion expression. In the following example sentence,

Example 4.

Canon 30D 太贵了

(*Canon 30D is too expensive*)

the target attribute “price” is not given in text, but this attribute can be recognized since “贵” (*expensive*) carries the information of price.

Fourthly, the domain opinion ontology, which maintains the most discussed product attributes, is introduced in the annotation of the target attributes. In practice, a user may use different expressions to describe one product attribute. For example, one may use the words of “LCD”, “LCD screen”, “屏”, “屏幕”, “彩屏”, “芒”, and “彩芒” to express one attribute “LCD Screen”. All of these words can be assigned one attribute concept “LCD Screen” in the domain ontology. Ontology-based opinion attribute annotation is helpful to group the relevant attribute expression and reduce the data sparseness.

Fifthly, the annotation of opinion expressions are determined. For each opinion expression, its expression segment, opinion keyword, polarity and degree are annotated. Furthermore, the negations and modifiers relevant to the opinion expression are annotated (if applicable).

Lastly, XML format is designed to record the annotations.

Based on the above discussion, 13 elements are scoped to describe a comment in the reviews, including:

opinion_source (also **opinion holder**, key element, i.e. compulsory, the holder of one opinion),

opinionated_sentence (key element, the sentence is opinionated or not)

comment_segment (element, the scope of one comment in the opinionated sentence)

target_attribute (key element, the target product attribute of a comment which is a leaf node in the domain opinion ontology)

visibility_of_target_attribute (key element, whether the target attribute is explicitly given in the comment segment)

attribute_segment (key element, the scope of the attribute description)

visibility_of_opinion_expression (key element, implicit metaphors or explicit expression)

opinion_expression_segment (key element, the scope of one opinion expression in a comment)

opinion_keyword (key element, the keyword reflects the polarity)

negation (optional, the negation in one comment)

modifier (optional, the modifier in one comment)

comment_polarity (key element, the sentimental polarity of one opinion)

polarity_degree (element, the polarity degree of one opinion)

Section 3 will illustrate the identification and annotation of these elements.

3. Corpus Annotation

3.1 Raw Data Preparation

The opinion corpus is constructed based on the reviews for two kinds of hot products, namely digital camera (DC) and mobile phone (MP). The on-line reviews are downloaded from two hot profession review sites, respectively, namely:

Digital Camera: www.xitek.com

Mobile Phone: www.soit.com.cn

A xml-based wrapper is developed to extract review documents from these web-pages. The raw corpus for DC and MP reviews have 7,868 and 622,662 documents, respectively.

3.2 Domain Opinion Ontology

Two opinion ontology are manually complied for DC and MP products, respectively. A part of DC opinion ontology is given below,

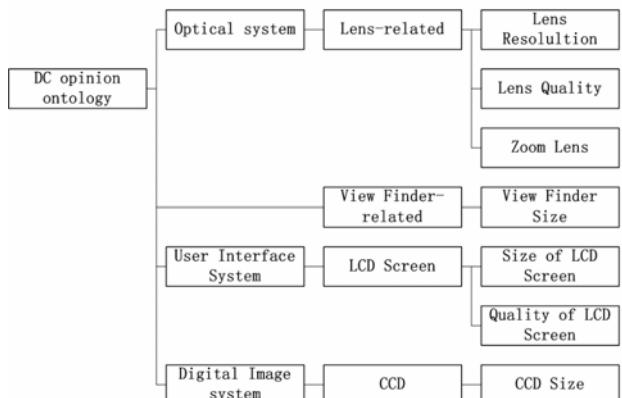


Figure 1. Illustration of DC Ontology

It is seen that each opinion ontology has three level nodes and each leaf node is a target product attribute. Table 1 gives the node information of DC opinion ontology and MP opinion ontology, respectively.

	DC	MP
Level-1 Nodes	12	13
Level-2 Nodes	41	36
Level-3 Nodes	60	80

Table 1: The Concept Codes in DC and MP Opinion Ontology

3.3 Opinion Annotation

The opinion annotation has seven steps which is illustrated as follows.

Step 1. Index the sentence by giving its *review_id*, *paragraph_id* and *sentence_id*.

Step 2. Determine the opinion source. In the online review forums, the user who posts the review documents contributes to the opinion source.

Step 3. Analyze the sentences in the review document and determine whether the sentence is opinionated. For a factual sentence without opinions, such as Example 5,

Example 5.

使用 CF 卡，支持 MicroDrive
(Use CF card and support MicroDrive)

it is annotated as illustrated below,

<Sentence>使用 CF 卡，支持 MicroDriver
<Comment no="0"> </Comment>
</Sentence>

where, *<Sentence>* indicates a full sentence in the review and **Comment no** indicates the number of comments in this sentence. If current sentence has no opinions, the value of Comment no is assigned 0.

Step 4. For an opinionated sentence, recognize all of the comments in this sentence.

Example 6. Canon 30D 的屏幕很大但不精细
(The LCD screen of Canon 30D is very large, but the resolution is low)

In this example sentence, there are two comments identified. The first one is that “Canon 30D 的屏幕很大” (*The LCD screen of Canon 30D is very large*) and the second one is that “Canon 30D 的屏幕 不精细”. (*The resolution of the LCD screen of Canon 30D is low*). Each comment is given a *comment_id*.

Step 5. For each comment, determine the maximum scope of the comment, i.e. *comment segment*, covering both the *attribute segment* and *opinion expression segment*. Corresponding to Example 6, the two comment segments are respectively,

1: *comment_segment* = “Canon 30D 的屏幕很大”
(*The LCD screen of Canon 30D is very large*)
2: *comment_segment* = “Canon 30D 的屏幕很大但不精细”(*The LCD screen of Canon 30D is very large, but the resolution is low*)

The *comment_segment* gives the boundary of a comment. Attribute to the fact that some sentences have more than one comments, the segments for different comments may be overlapped.

Step 6. Annotate the elements related to the target attribute.

6a. Determine the attribute nodes in the domain opinion ontology corresponding to the target attribute of each comment. For the Example 6, the targeted attribute of its two comments are,

1: *target attribute* = “Size of LCD Screen”
2: *target attribute* = “Quality of LCD Screen”

6b. Determine whether the target attribute is presented in the opinion segment explicitly. If yes, annotate the *attribute segment*, otherwise, mark the target attribute is implicate. For the two comments in Example 6, the target attributes are both presented explicitly, and, they are annotated as below,

1: *Visibility_of_target_attribute*= TURE
attribute segment = Canon 30D 的屏幕 (*The screen of Canon 30D*)
2: *Visibility_of_target_attribute*= TURE
attribute segment = Canon 30D 的屏幕 (*The screen of Canon 30D*)

For another example, Example 4, “Canon 30D 太贵了 (*30D is too expensive*)”, since the target attribute “price” is not given in the opinion segment, the value of *interested attribute* is assigned FALSE.

Visibility_of_interested_attribute= FALSE
attribute segment= ””

Step 7. Annotate the elements related to the opinion expression.

7a. Determine an opinion is explicit expression or implicit metaphor. For the two comments in Example sentence 6, both of them are explicit expressions and thus the values of their *visibility_of_opinion_expression* element are TRUE. On the contrary, Example 3 use implicit metaphor, and thus, its value of *visibility_of_opinion_expression* element is assigned FALSE.

7b. Annotate the opinion expression segment in the comment. Note that, the *opinion_expression_segment* for different comments do not overlap. The annotation results corresponding to Example 6 are given below,

1: *opinion_expression_segment*= “很大” (*very large*)
2: *opinion_expression_segment*= “不精细” (*not fine*)

7c. Annotate the opinion keyword, negation and modifier. In the first comment in Example 6, the identified opinion keyword is “大” (*large*) and a modifier “很” (*very*) is identified. As for the second one, the opinion keyword is “精细” (*fine*) which is positive. However, a negation “不” (*not*) is found which leads the polarity of this comment to negative. Example 6 is further annotated as,

```

1: opinion_keyword= “大”(large) negation=""
modifier=“很”(very)
2: opinion_keyword= “ 精 细 ”(fine) negation
=“不”(not) modifier=""

```

7d. Determine the opinion polarity and the polarity degree. The opinion polarity has three categories, i.e. *positive*, *neutral* and *negative*. For positive and negative polarities, the two degrees, i.e. *normal* and *strong*, are used to describe the strength of the opinion polarity. The polarity degree of the first comment in Example 6 is strengthened by a degree modifier “很”(very), thus the polarity degree of this comment is *strong*. As for the second one, its degree is *normal*. They are annotated as,

```

1: comment_polarity= “positive”, polarity_degree
=“strong”
2: comment_polarity = “negative”, polarity_degree
=“normal”

```

The annotation output for the example sentence 1 is given below,

```

<Sentence sentence_opinionated="TRUE">
Canon 30D 屏幕很大但不精细

<Comment comment_id="1" target_attribue="Size of
LCD Screen" visibility_of_target_attribue ="TURE"
attribute_segment="Canon 30D 的屏幕" visibility_of_
opinion_expression="TRUE" opinion_expression_
segment="很大" opinion_keyword="大" negation=""
modifier=" 很 " comment_polairty="positive"
polarity_degree="strong"> Canon 30D 的屏幕很大
</Comment>

<Comment comment_id="2" target_attribue="Quality
of LCD Screen" visibility_of_target_attribue
="TURE" attribute_segment="Canon 30D 的屏幕"
isiblity_of _opinion_expression="TRUE" opinion_
expression_segment=" 但不精细" opinion_keyword
=" 精 细 " negation=" 不 " modifier=""
comment_polairty="negative"
polarity_degree="normal"> Canon 30D 屏幕很大但不
精细 </Comment>
</Sentence>

```

More examples are given in the Appendix.

3.4 Quality Assurance

The annotators are two students majoring in linguistics. To ensure annotation quality, the annotation is done in two phases. In the first annotation phase, 50 DC review documents was annotated by two annotators in duplicates. Their outputs were then checked. The inconsistencies between different annotators were then discussed to clear any misunderstanding in order to come up with the most appropriate annotations and proper understanding by all annotators. In the second phase, the rest review

documents was then divided for annotation by different annotators in which 100 DC review documents are duplicate distributed to the two annotators so that the annotation agreement between different annotators can be estimated.

4. Current Status and Evaluations

Up to now, the annotation on DC and MP review documents has been finished. In the DC topic, 1,100 review documents, consisting of 7,538 sentences, are annotated in which 100 review documents are annotated doubly. The annotators identified 5,628 opinionated sentences and annotated 8,990 comments correspondingly. On the average, 1.60 comments are recognized in each opinionated sentence. As for the MP reviews, 500 review documents, consisting of 3,397 sentences, are annotated. Within the 2,236 opinionated sentences, 3,734 comments are annotated. On the average, 1.67 comments are recognized in each opinioned sentence.

The annotation agreement is estimated by comparing the annotation results of 100 doubly annotated DC reviews. Annotator1 annotated 844 comments and Annotator2 annotated 848 comments in these review documents. In their outputs, the annotations of 542 comments are completely same. 203 comments have matched comment segments (in which 37 ones have same key elements and same comment polarity; 154 of them have different key elements and same comment polarity; 12 have different key elements and different comment polarity). 72 comments have approximate-matched comment segments (the character string matching percentage is higher than 85%) (in which 31 have same key elements and same comment polarity, 35 have different key elements and same comment polarity, 6 have different key elements and different comment polarity) and 27 comments can not be matched. It means that the annotation agreement achieves 64.2% under strict metrics (complete match) and it achieves 94.6% under lenient metrics (approximate matched).

5. Observation on Opinion Expressions in Chinese Product Reviews

5.1 Observation on Opinion Words

Corresponding to the 8,483 annotated positive comments, 1,229 categories of positive opinion keywords are found. The average frequency of each kind of positive opinion keyword is 6.90. As for the 4,271 negative comments, 983 categories of negative opinion keyword are found. The average frequency of each kind of negative keyword is 4.34, which is obviously lower than the one of positive opinion keywords.

The frequencies of top-100 categories of positive keywords and negative keywords are further observed, which is shown in Figure 2.

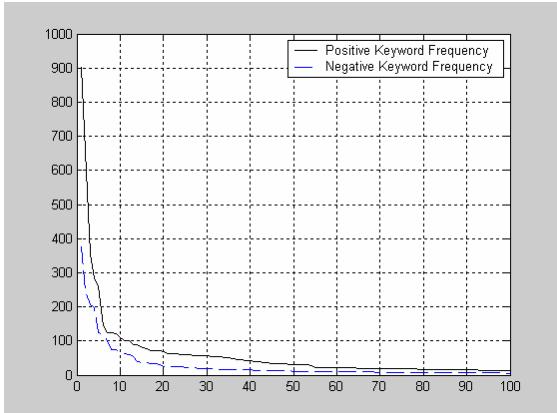


Figure 2. Frequencies of Top-100 Positive Keywords and Negative keywords

It is observed that the occurrences of top-frequency negative keywords are much lower than the occurrences of top-frequency positive keywords. It means that the negative expression in Chinese is more flexible than the positive expression.

By comparing with a static sentimental lexicon (consisting of 5,054 positive words and 3,493 negative words) adopted in (Xu et al. 2007), 317 categories of new positive words and 183 categories of new negative words are obtained. It is shown that the opinion corpus proves some opinion expression knowledge beyond the lexicon.

5.2 Observation on Context-Dependent Opinion Keywords

The status of context-dependent opinion keywords are observed. In the opinion corpus, there are 626 cases that the positive and negative comments having the same opinion keywords. It is $626/8,483=7.4\%$ of the total occurrences of positive keywords and $626/4271=14.7\%$ of the total occurrences of negative keywords.

This result indicates that context-dependent opinion keywords occurred in Chinese frequently, and thus the static sentimental lexicon, which assigns one fixed polarity to each opinion keyword, is not good enough for the determination of sentimental orientation.

Based on this observation, (Xia et al. 2007) proposed a unified collocation model, which determines the sentimental orientation based on the collocated observing attributes and sentimental words. This model is shown effective in Chinese opinion mining systems.

5.3 Observation on Negations and Modifiers

The annotated modifiers and negations are observed. In this opinion corpus, 434 categories of modifiers (with the total occurrence of 4,954) and 111 categories of negations (with the total occurrences of 1,222) are annotated. The frequencies of the top-100 modifiers and negations are shown in Figure 3, respectively.

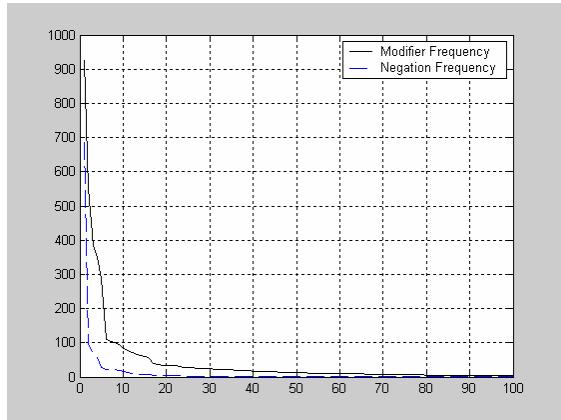


Figure 3. Frequencies of Top-100 Modifiers and Negations

It is observed that the number of frequently used modifiers (frequency >5) is 79 and the number of frequently used negations (frequency>5) is 16. Especially, the frequencies of negations after top-16 ranks are very low. It indicates that the use of negations is relatively fixed comparing with the use of modifiers.

6. Conclusions

In this study, a new scheme for annotating opinions in Chinese product reviews is proposed. Following this scheme, a Chinese opinion corpus on online product reviews is constructed. This corpus is a useful resource to analyze the characteristics of opinion expressions in Chinese product reviews. Furthermore, this corpus is useful to evaluate the automatic opinion mining algorithms as a standard answer. Currently, more reviews are annotated to expand the scale of this opinion corpus. Furthermore, the revision and deep annotation of the opinion corpus are conducted such as annotate the equivalents of the informal words or abbreviations in the comments.

Acknowledgements

This research is partially supported by The Chinese University of Hong Kong under the Direct Grant Scheme project (2050417) and Strategic Grant Scheme project (4410001) and The Hong Kong Polytechnic University (Project Code A-P203) and a CERG Grant (Project code 5087/01E). The authors also acknowledge the great work of the annotators.

References

- Hu, M.Q. and Liu, B. (2004). Mining and Summarizing Customer Reviews, In *Proceedings of ACM SIGKDD 2004*, pp.168--177.
- Hu, M.Q. and Liu, B. (2006). Opinion Extraction and Summarization on the Web, In *Proceedings of the 21th National Conference on Artificial Intelligence and the*

18th Innovative Applications of Artificial Intelligence Conference, Boston.

- Yu, H. and Hatzivassiloglou, V. (2003). Towards Answering Opinion Question: Separating Facts From Opinions and Identifying the Polarity of Opinion Sentences, In *Proceedings of the 8th Conference on Empirical Methods in Natural Language Processing*, Japan.
- Hatzivassiloglou, V. and McKeown, K.R. (1997). Predicting the Semantic Orientation of Adjectives. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, Madrid, Spain, pp.174--181.
- Gordon, A. Kazemzadeh, A., Nair, A., and Petrova, M. (2003). Recognizing Expressions of Commonsense Psychology in English Text. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, Sapporo, Japan, pp.208--215.
- Wiebe, J., Wilson, T., Bruce, R., Bell, M., and Martin, M. (2004). Learning Subjective Language. *Computational Linguistics*, 30(3), pp. 277--308.
- Riloff, E., Wiebe, J., and Wilson, T. (2003). Learning Subjective Nouns Using Extraction Pattern Bootstrapping. In *Proceedings of the 7th Conference on Natural Language Learning*. Edmonton, Canada, pp. 25--32.
- Seki, Y. (2007). Opinion Holder Extraction from Author and Authority Viewpoints. In *Proceedings of the 30th Annual International ACM SIGIR Conference*, Netherlands, pp. 841--842.
- Ku, L.W. Wu, T.H., Lee, L.Y. and Chen, H.H. (2005). Construction of an Evaluation Corpus for Opinion Extraction, In *Proceedings of NTCIR-5 Workshop Meeting*, pp.513--520.
- Pang, B., Lee, L. and Vaithyanathan S. (2002). Thumbs up? Sentiment Classification Using Machine Learning Techniques. In *Proceedings of EMNLP-2002*, pp. 79--86.
- Wiebe, J., Wilson, T., and Cardie, C. (2005). Annotating Expressions of Opinions and Emotions in Language, *Computer Intelligence*, 39(2), pp.165--210
- Xia, Y.Q., Xu, R.F., Wong, K.F. and Zheng F. (2007). The Unified Collocation Framework for Opinion Mining, In *Proceedings of IEEE ICMLC2007*, pp. 844--850
- Xu, R.F., Wong, K.F. and Xia, Y.Q. (2007) Opinmine – Opinion Analysis System by CUHK for NTCIR-6 Pilot Task, In *Proceedings of NTCIR-6*, pp.350--357

Appendix. Annotation Examples

```
<ReviewID>11</ReviewID>
<ProductCategory>DC</ProductCategory>
<ProductName>Canon PowerShot S30</ProductName>
<ReviewSource>www.xitek.com</ReviewSource>
<User>eudemon7</User>
<ReviewDate>2003-02-02 19:35</ReviewDate>
<Annotator>angela</Annotator>

<Pros>
<paragraph>图像质量优秀，功能齐全，设计合理，操作方便</paragraph>
<Sentence no="1">图像质量优秀，功能齐全，设计合理，操作方便
    <Comment no="1" target_attribute="画面" visibility_of_target_attribute="true" attribute_segment="图像质量" visibility_of_opinion_expression="true" opinion_expression="优秀" opinion_keyword="优秀" negation="" modifier="" polarity_degree="1" degree="1">图像质量优秀</Comment>
    <Comment no="2" target_attribute="功能设计" visibility_of_target_attribute="true" attribute_segment="功能" visibility_of_opinion_expression="true" opinion_expression="齐全" opinion_keyword="齐全" negation="" modifier="" polarity_degree="1" degree="1">功能齐全</Comment>
    <Comment no="3" target_attribute="设计合理" visibility_of_target_attribute="true" attribute_segment="设计" visibility_of_opinion_expression="true" opinion_expression="合理" opinion_keyword="合理" negation="" modifier="" polarity_degree="1" degree="1">设计合理</Comment>
    <Comment no="4" target_attribute="操作" visibility_of_target_attribute="true" attribute_segment="操作" visibility_of_opinion_expression="true" opinion_expression="方便" opinion_keyword="方便" negation="" modifier="" polarity_degree="1" degree="1">操作方便</Comment>
<SentencePolarity>0:4:0:0:0<SentencePolarity>
</Sentence>
<ParagraphPolarity>0:4:0:0:0<ParagraphPolarity>
</Pros>

<Cons>
<paragraph>电池很不耐用，附件太贵</paragraph>
<Sentence no="1">电池很不耐用，附件太贵
    <Comment no="1" target_attribute="电池相关" visibility_of_target_attribute="true" attribute_segment="电池" visibility_of_opinion_expression="true" opinion_expression="很不耐用" opinion_keyword="耐用" negation="不" modifier="很" polarity_degree="-1" degree="2">电池很不耐用</Comment>
    <Comment no="2" target_attribute="配件价格" visibility_of_target_attribute="true" attribute_segment="附件" visibility_of_opinion_expression="true" opinion_expression="太贵" opinion_keyword="贵" negation="" modifier="太" polarity_degree="-1" degree="2">附件太贵</Comment>
<SentencePolarity>0:0:0:0:2<SentencePolarity>
</Sentence>
<ParagraphPolarity>0:0:0:0:2<ParagraphPolarity>
</Cons>

<General>
<paragraph>就同档价格而言，没有比他更出色的相机了</paragraph>
<Sentence no="1">就同档价格而言，没有比他更出色的相机了
    <Comment no="1" target_attribute="总体" visibility_of_target_attribute="true" attribute_segment="相机" visibility_of_opinion_expression="true" opinion_expression="没有比他更出色" opinion_keyword="出色" negation="没有" modifier="更" polarity_degree="1" degree="2">没有比他更出色的相机了</Comment>
<SentencePolarity>1:0:0:0:0<SentencePolarity>
</Sentence>
<ParagraphPolarity>1:0:0:0:0<ParagraphPolarity>
</General>

</Review>
```