AUTOMATIC PARSING OF THE METAPHOR POLARITY FOR OPINION MINING

Xiaoying Xu¹, Ya Li², Jianhua Tao³, Xuefei Liu⁴

¹, ⁴ PSC Center, Chinese Language & Literature Dept., Beijing Normal University, 100875, Beijing, China
¹, ², ³, ⁴ NLPR, Institute of Automation, Chinese Academy of Sciences, 100190, Beijing, China
¹xuxiaoying2000@bnu.edu.cn, ²yli@nlpr.ia.ac.cn, ³jhtao@nlpr.ia.ac.cn, ⁴liuxuefei1234@163.com

ABSTRACT

In recent years, both metaphor interpretation and opinion mining have drawn much attention in the natural language processing (NLP) field. This paper aims to make a connection between these two fields. In this paper, we propose to extend the glossary orientation annotation to the vehicle (the target concept part of the metaphor) by using an automatic annotation method, and based on the vehicle’s orientation corpus, we parse the metaphor’s polarity after extracting the metaphor (especially the simile) from the large-scale corpus. Two experiments are conducted to investigate the reliability of our proposal. The result of the first experiment shows the proposed method obtains better results than the system we proposed in 2009 in both precision and recall, while the result of second experiment shows that more than 65% metaphors have a very stable sentiment orientation. Generally, the results demonstrate the effectiveness of our approach and verifying our approach’s high reliability.

Index Terms— vehicle orientation, metaphor polarity parsing, opinion mining

1. INTRODUCTION

Metaphor is broadly used in human language, especially in expressing opinion. Conveying affect is one of important roles for metaphor, and metaphor is the important way of conveying affect. Emotional states and behavior often themselves described metaphorically (Kővecses, 2000; Fussell & Moss, 1998). In recent years, both metaphor interpretation and opinion mining have drawn much attention in the natural language processing (NLP) field.

Within opinion mining field, three subtasks can be identified, determining document subjectivity, (Pang and Lee, 2004), determining document polarity (Pang and Lee, 2004; Turney, 2002) and determining the strength of document orientation (Wilson et al., 2004). For realizing the tasks, Machine-learning Methods and Language-modeling Methods are two major approaches. As concern as the Language-modeling Methods, determining the opinion orientation of the glossary is a very important component in an opinion mining system. Large amounts of recent works have tackled the issue of term orientation and most of them focus on the orientation of the isolated words (Takamura et al., 2005; Wiebe and Mihalcea, 2006).

Within the metaphor interpretation field, the task of identifying metaphor from a large-scale corpus has received an increasing amount of attention due that metaphors have interesting applications in many NLP problems like machine translation, text summarization, information retrieval and question answering (Martin, 1990; Mason, 2004). Some of the works achieved good results in recognizing different types of metaphors (Krishnakumaran & Zhu, 2007), but few of them concern the metaphor interpretation in the opinion mining field.

In this paper, we try to make a connection between these two fields. We extend the glossary orientation annotation to the vehicle (the target concept part of the metaphor) by using an automatic annotation method, and based on the vehicle’s orientation corpus, we parse the metaphor’s polarity after extracting the metaphor (especially the simile) from the large-scale corpus.

For the following three reasons, the metaphor extraction and the metaphor polarity parsing can play quite an important role in the opinion mining system. Firstly, metaphor is frequently used when expressing opinion in human language, metaphor can’t be ignored in opinion mining system(Kővecses, 2000; Fussell & Moss, 1998); Secondly, without special processing, it’s difficult for the machine to recognize the metaphor and metaphor’s polarity automatically (Krishnakumaran & Zhu, 2007), which will directly decrease the accuracy of the polarity recognizing; Finally, we can easily observed that, comparing with that of the usual sentence, the syntactic structure of the sentences including metaphor is relatively simpler and is easier for the machine to label the word orientation and parse the polarity, which can avoid the difficulties caused by the sophistic syntax structure in the previous language-modeling Methods opinion mining system. This observation can be strongly proved by the results gained in this paper. Furthermore, the vehicle’s orientation corpus established in our paper is not
only helpful for opinion mining system, but also helpful for kinds of NLP fields such as machine translation, text summarization, information retrieval and question answering fields.

The whole paper is organized as follows. Section 1 is the introduction of the paper. Section 2 is a brief review of previous related work. Section 3 describes our annotation scheme for vehicle polarity and parsing scheme for metaphor polarity in detail. Section 4 introduces the corpus and method used in this paper, Section 5 presents the experiments and evaluation. Section 6 discusses the conclusion and feature research.

2. RELATED WORK

Large amounts of recent works have tackled this issue of term orientation. Most of the previous works (Esuli and Sebastiani, 2005, 2006; Hatzivassiloglou and McKeown, 1997; Kamps et al., 2004; Takamura et al., 2005; Turney and Littman, 2003; Xiaoying Xu et al, 2009) dealing with term orientation established their lexicons automatically or semi-automatically based on some available resource. Few of the previous works (Wiebe and Mihalcea, 2006; Karo Moilanen and Stephen Pulman, 2008; Xiaoying X u et al, 2009) established their lexicons manually. All of the work mentioned above focuses on the orientation of the isolated words.

Our work pays attention to vehicle orientation of the term for the first time, which can not only significantly increased the entry of the term orientation lexicon, but also add the indispensable contextual-based orientation to the lexicon. Together with the isolated term orientation lexicon established in 2009(Xiaoying Xu et al, 2009), we can build a more complete term orientation lexicon than any other previous work.

Within opinion mining, lots of works focus on determining the subjectivity and polarity of the document as well as the strength of document orientation (Turney, 2002) by using machine-learning Methods or Language-modeling Methods, but no work tackled the topic of parsing the metaphor polarity. Much research has gone into the processing of metaphors and provides some metaphor recognizing and understanding systems (Gentner, 1983 : Martin, 1990). Some systems had got quite good result in metaphor extraction (Krishnakumaran & Zhu, 2007; Gedigian et al, 2006). But most works on the metaphor processing field set their target on machine translation, text summarization, information retrieval and question answering, few of them tackle the issue of opinion mining.

This paper focus on parsing the metaphor polarity with the language-based method and the recognizing result had been significantly improved.

3. THE PAR Sing OF METAPHOR POLARITY IN OPINION MINING SYSTEM

3.1. The annotation of tenor orientation in metaphor

Metaphor expresses an analogy between two seemingly unrelated concepts. Metaphoric usages enhance the attributes of the source concept (which is named as “vehicle” in the rhetoric) by comparing it with the attributes of the target concept (which is named as “tenor” in the rhetoric).

For example, as concern as a certain word, when it is used in an ordinary sentence, the lexical orientation is stable and can be reflect in the isolated words. Most previous researchers established their polarity lexicon by annotated this kind of orientation of the isolated words (Wiebe and Mihalcea, 2006; Karo Moilanen and Stephen Pulman, 2008; Xiaoying Xu et al, 2009). When it used as a tenor in a metaphoric sentence, some neutral orientation words will change their orientation to positive or negative polarity. For example, “wolf” is a neutral orientation noun in the ordinary Chinese sentence “there is a wolf”, and so is the “pine tree” in the sentence “this pine tree is very small”, but when they are used as vehicles in the sentence as “he is a wolf”, “he is just like a pine tree”, “wolf” has a negative orientation while “pine tree” has a positive orientation. We call this kind of orientation as the vehicle orientation of the lexical.

The words’ vehicle orientation is as stable as the isolate orientation and can be well known by the native speaker of the language. It’s also an important semantic resource for recognizing the polarity in opinion mining.

Metaphor can be divided into two types: the simile and the metaphor. In simile, three part of the analogy, e.g. the tenor, the figurative verb and the vehicle appear together in one sentence, while in metaphor, only the vehicle must appear in every metaphorical sentence.

In this paper, according to mapping relation between the vehicle and the words with semantic orientation which are co-appeared in simile sentences collected from large-scale corpus, we automatically label the vehicle orientation. Then, we parse the metaphor polarity based on the vehicle orientation corpus and extend the metaphor parsing to opinion mining.

3.2. Extending metaphor polarity parsing to Opinion Mining

Figure 1. (a) Metaphor polarity stability and (b) polarity distribution of 400 metaphor nouns
Figure 1 shows flowchart of extending the vehicle orientation to the metaphor polarity parsing in opinion mining. Firstly, the word segmented and POS tagged figurative sentences which are collected from large-scale corpus, and then the vehicle and the words which express the semantic orientation are detected, afterwards, the vehicle orientation are automatically parsed. Though the method, a large vehicle orientation lexicon could be obtained, which will be a great helpful language resource for opinion mining as well as other NLP fields. In opinion mining system, word segment and POS tagging (not necessary in all the opinion mining system) are performed on the raw input text, and then metaphors are detected and recognized by the sentiment orientation tagged in the lexicon as well as the context information. Finally, the sentiment orientation could be applied to opinion mining.

4. CORPUS AND METHOD

4.1. Corpus Construction

The typical figurative verbs (TFV) in Chinese are “xiang4, hao3xiang4, ru2, wan3ru2, fang3fu2, hao3bi3, hao3si4, you2ru2, huang3ruo4 (All of them are synonyms and mean as, seem, like)”, and the typical sentence templates is “……Tenor (T)……TFV ……vehicle (V)…… “. To simplify our work, this paper only focuses on the words of plants and animals that serve as a vehicle in a figurative sentence. A lexicon with more than 3000 words of the name of plants and animals is used in our experiment.

In order to build a large-scale figurative sentence corpus, we utilize search engines as well as several pre-constructed corporsa to search for the possible sentences which match the typical sentence templates. The detailed construction procedure is as followings.

Step 1: We construct queries, such as “as*” (* is the wildcard) and submit them to search engines (SE). In order to get the desired corpus as much as possible, we use seven famous SEs including Baidu, Google, Sogou, Soso, yahoo, Microsoft Bing and Youdao. For each query, the search depth is 3, and in each search level, we select the first 100 web pages, and that adds up to 300 web pages for each query. All the web pages constitute the raw corpus C1.

We also introduce raw Corpus C2 and Sogou News corpus (SogouCS1 and SogouCA2) into this work. C2 is a small corpus which is about 600M Bytes crawled from http://www.163.com/ and its sub channels, e.g., Education, News, Books and Sports. SogouCS is 6 months corpus from Olympic, Sports, IT etc. 18 channels of http://www.sogou.com/. SogouCA is 2 months corpus from all the main portal websites. The size of SogouCS and SogouCA are 3.3G Bytes and 2.1G Bytes.

Step 2: The corpus C1, C2, SogouCS and SogouCA finally constitute the raw corpus. Then we separated the paragraphs into sentences by period, comma, exclamatory mark and question mark etc. Afterwards, we use regular expressions to match all the figurative sentence templates to build a potential figurative sentence corpus PC.

Step 3: Word segmentation and Part-of-Speech tagging were carried out on the potential figurative sentence corpus PC. Maximum Entropy (ME) model is adopted in this task. Then we can get the more specific figurative sentence templates “as N……A……”, in which, N is a noun, and represents the vehicle V, and A is a adjective and may probably imply the emotion of the speaker, which is helpful for us to detect the orientation of the vehicle. Table 3 shows some Noun-Adjective pairs in simile sentence instances.

Step 4: Because some sentences which match the templates do not have figurative sense, a research assistant is asked to seek out a small amount of the true figurative sentences and label the vehicle polarity manually. The corpus was finally used to test the feasibility of the proposed method.

4.1. Automatic annotation of the metaphor’s polarity

We have already built a large-scale Chinese sentiment lexicon in (Xiaoying Xu et al. 2009), which contains more than 11 000 000 words. In the automatic annotation of vehicle’s orientation in a sentence, we firstly search the adjective which is included in the sentiment lexicon, and then the adjective’s polarity is directly mapped to the vehicle. For instance,

(1) xiang4 hua1 yi1yang4 piao4liang4
    as    flower    AUX    beautiful (as beautiful as flowers)

We automatically segment the sentence into words, and tag their Part-of-Speech, and then match the Noun-Adjective pair.

Xiang4/v hua1/n yi1yang4/u piao4liang4/a

In the above example N = hua1 (flowers), A = piao4liang4 (beautiful). Afterwards we search the adjective in sentiment lexicon and find that the adjective A has a positive polarity, and then the noun “hua1” is considered as a positive metaphor. If the adjective was not collected in the sentiment

---

1 http://www.sogou.com/labs/dl/cs.html
2 http://www.sogou.com/labs/dl/ca.html
lexicon, we currently consider that metaphor is neutral in sentiment.

5. EXPERIMENTS AND EVALUATION

This section introduces the automatic metaphor polarity annotation results. To clearly demonstrate the effectiveness of the proposed method, we compared its performance with the sentence sentiment recognition system introduced in (Xiaoying Xu et al. 2009). The main idea of the previous sentiment recognition system proposed 2009 is that the sentiment propagates word by word and its influence fades away after several steps. The test data set contains 550 figurative sentences which are manually labeled, among which, 469 sentences are successfully mapped, and the others are not. There are two reasons, the first one is the adjective in the sentence is not included in the sentiment lexicon and the other is the adjective is neutral in sentiment. Table 4 and Table 5 show the detailed performance of two metaphor polarity experiments.

<table>
<thead>
<tr>
<th>Precision</th>
<th>Recall</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>0.86</td>
<td>0.74</td>
</tr>
<tr>
<td>Negative</td>
<td>0.57</td>
<td>0.74</td>
</tr>
</tbody>
</table>

The results show that the proposed method obtains better results than the system we proposed in 2009 in both precision and recall. Considering that it is a simple direct mapping-based recognition method the performance of our approach is very promising.

We also examined the sentences which are not successfully recognized. The main reason is some words are incorrectly segmented thus they can not be found in the sentiment lexicon and a few of the errors are caused by POS tagging. The word segmentation error could be overcome by integrating the sentiment lexicon into the Maximum Entropy based word segmentation model. Through this method, the words could be identified as long as the word is in the large-scale sentiment lexicon.

The second experiment was conduct to test the stability of metaphor polarity. The metaphor polarity stability in this paper is defined as the absolute difference of metaphor’s positive and negative polarity occurrence percentage, which is shown as the following equation.

\[ \text{Metaphor polarity stability} = \text{abs} \left( \frac{\text{Positive} - \text{Negative}}{\text{frequency of occurrence}} \right) \]  

High metaphor polarity stability would benefit the understanding of the sentence’ figurative sense. Fig. 1(a) shows the metaphor polarity stability distribution. From Fig.1(a), we can see that more than 65% metaphors have a very stable sentiment orientation. The difference of polarity percentage is above 0.8, in other words, more that 65% metaphors have a specific polarity (positive/negative) with the probability of more than 0.9. Fig. 1(b) shows the positive and negative polarities distribution. The overwhelming percentage of positive polarity indicates that people tend to use figurative sense to show their positive sentiment.

The above experiments shows the feasibility of the proposed method, which using the adjective’s polarity to map the metaphor’s sentiment polarity. This ensures that we can build a reliable metaphor polarity lexicon according to the direct-mapping mechanism. With the lexicon, people could know that “ta1 xiang4 hua1 yi2yang4 (She looks like a flower)” expresses a positive sentiment because the Vehicle “hua1” has a very strong positive possibility even if there is no obvious adjective in this sentence. This work is especially helpful for non-native speakers read between the lines. The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

6. CONCLUSION

We propose a metaphor parsing scheme based on the vehicle annotation corpus. According to mapping relation between the vehicle and the words with semantic orientation which are co-appeared in simile sentences collected from large-scale corpus, we automatically label the vehicle orientation. Then, we parse the metaphor polarity based on the vehicle orientation corpus and extend the metaphor parsing to opinion mining. The results demonstrate the effectiveness of our approach and verifying our approach’s high reliability.
In our experiment, we only labeled the vehicle orientation of animals and plants name, and the amount of simile sentence templates is relatively small. In future, we will explore to acquire the orientation knowledge of human language from different figures of speech, such as metonymy, exaggeration, the irony.

ACKNOWLEDGMENT

The work was supported by the National Science Foundation of China (No. 60873160, 90820303), 863 Programs (No. 2009AA01Z320), China-Singapore Institute of Digital Media (CSIDM) and the Beijing Normal University (11-30, 10-12-02), the Fundamental Research Funds for the Central University (2010105565004GK) and the State Language Commission 12th Five-Year language research programme (YB125-41)

REFERENCES


