

YSC-DSAA: An Approach to Disambiguate Sentiment Ambiguous Adjectives Based On SAAOL

Shi-Cai Yang

Ningbo University of Technology
Ningbo, Zhejiang, China
nbcysc@126.com

Mei-Juan Liu

Zhejiang Ocean University
Zhoushan, Zhejiang, China
azalea1212@126.com

Abstract

In this paper, we describe the system we developed for the SemEval-2010 task of disambiguating sentiment ambiguous adjectives (hereinafter referred to SAA). Our system created a new word library named SAA-Oriented Library consisting of positive words, negative words, negative words related to SAA, positive words related to SAA, and inverse words, etc. Based on the syntactic parsing, we analyzed the relationship between SAA and the keywords and handled other special processes by extracting such words in the relevant sentences to disambiguate sentiment ambiguous adjectives. Our micro average accuracy is 0.942, which puts our system in the first place.

1 Introduction

We participated in disambiguating sentiment ambiguous adjectives task of SemEval-2010(Wu and Jin, 2010).

Together 14 sentiment ambiguous adjectives are chosen by the task organizers, which are all high-frequency words in Mandarin Chinese. They are: 大|big, 小|small, 多|many, 少|few, 高|high, 低|low, 厚|thick, 薄|thin, 深|deep, 浅|shallow, 重|heavy, 轻|light, 巨大|huge, 重大|grave. These adjectives are neutral out of context, but when they co-occur with some target nouns, positive or negative emotion will be evoked. The task is designed to automatically determine the semantic orientation of these sentiment ambiguous adjectives within context: positive or negative (Wu and Jin, 2010). For instance, “价格高|the price is high” indicates negative meaning, while “质量高|the quality is high” has positive connotation.

Considering the grammar system of contemporary Chinese, a word is one of the most basic linguistic granularities consisting of a sentence. Therefore, as for the sentiment classification of a sentence, the sentiment tendency of a sentence can be identified on the basis of that of a word. Wiebe et al. (2004) proposed that whether a sentence is subjective or objective should be discriminated according to the adjectives in it. On the basis of *General Inquirer Dictionary, A Learner's Dictionary of Positive and Negative Words, HowNet, A Dictionary of Positive Words and A Dictionary of Negative Words* etc., Wang et al.(2009) built a word library for Chinese sentiment words to discriminate the sentiment category of a sentence using the weighted linear combination method.

Unlike the previous researches which have not taken SAA into consideration specially in discriminating the sentiment tendency of a sentence, in the SemEval-2010 task of disambiguating sentiment ambiguous adjectives, systems have to predict the sentiment tendency of these fourteen adjectives within specific context.

From the view of linguistics, first we developed a SAA-oriented keyword library, then analyzed the relationship between the keywords in the clauses and SAA, and classified its positive or negative meaning of SAA by extracting the clauses related to SAA in the sentence.

2 SAAOL

We create a SAA-oriented library marked as SAAOL which is made up of positive and negative words irrelevant to context, negative words related to SAA (NSAA), positive words related to SAA (PSAA), and inverse words. The above five categories of words are called keywords for short in the paper.

Positive and negative words irrelevant to context refer to the traditional positive or negative words which are gathered from *The Dictionary*

of Positive Words(Shi, 2005), *The Dictionary of Negative Words*(Yang, 2005), HowNet¹ and other network resources, such as Terms of Adverse Drug Reaction, Codes of Diseases and Symptoms, etc.

Distinguishing from the traditional positive and negative words, NSAA and PSAA in our SAAOL refer to those positive and negative words which are related to SAA, yet not classified into the positive and negative words irrelevant to context mentioned above.

We divide SAA into two categories: A category and B category listed in Table 1.

A category	B category
大 big	小 small
多 many	少 few
高 high	低 low
厚 thick	薄 thin
深 deep	浅 shallow
重 heavy	轻 light
巨大 huge	
重大 grave	

Table 1: SAA Classification Table

We identify whether a word belongs to NSAA or not on the following principle: any words when used with A category are negative; conversely, when used with B category, they are positive.

For example, in the following clauses,

“油价很高|oil prices are high” ,

“责任重大|the responsibility is important” ,

“任务很重|the task is very heavy” ,

“工作量很大|the workload is very large” ,

“油价|oil prices”, “责任|responsibility”, “任务|task”, “工作量|workload” are NSAA.

Correspondingly, we identify whether a word belongs to PSAA or not on the following principle: any words when used with A category are positive; however, when used with B category, they are negative.

In the clauses,

“粮食很多| much food” ,

“效率极低| efficiency is extremely low” ,

“存款利率高|interest rate on deposit is high” ,

“粮食| food” , “效率| efficiency” , “存款利率|interest rate on deposit” are PSAA.

In general, when two negative words are used together, the sentiment tendency that they show is negative. For instances, “糖尿病发病率

|incidence of diabetes”, “病毒感染|virus infection”, “战争破坏|destruction of wars”. However, in certain cases, some words play a part in eliminating negative meaning when used with negative words, for example, “反|anti-” , “抑制|restrain”, “避免|avoid”, “抗|resist”, “降低|reduce”, “降幅|fall”, “减少|decrease”, “控制|control”, “成本|cost”, “反对|oppose”, “下调|decrease”, “非|non-”, “不|not”. These special words are called inverse words in our SAAOL.

In the following instances, “减轻伤害|reduce the injury”, “抑制通胀|curb inflation”, “反战|anti-war”, the words “伤害|injury”, “抑制|inflation”, and “战争|war” themselves are all negative. When used with the inverse words“减轻|reduce”, “抑制|curb”, “反|anti-”, they express positive meaning instead.

On the basis of the above collected word library, we discriminate manually the positive and negative meaning, PSAA, NSAA, and inverse words in 50,000 Chinese words according to *Richard Xiao's Top 50,000 Chinese Word Frequency List*, which collects the frequency of the top 50000 Chinese words covered in the just published frequency dictionary of Mandarin Chinese based on a balanced corpus of ca. 50 million words. The list is available at http://www.lancs.ac.uk/fass/projects/corpus/data/top50000_Chinese_words.zip.

Based on HowNet lexical semantic similarity computing(Liu, 2002), Yang and Wu(2009) selected the new positive and negative benchmark words to identify the sentiment tendency by adopting the improved benchmark words and the modified method of computing similarity between words and benchmark words. Their accuracy rate arrived at 98.94%.

In light of the errors of manual calibration, we extended the keywords in SAAOL by applying Yang and Wu's (2009) method and added synonymic and antonymous words in it. Eventually we proofread and revised manually the new extended keywords.

3 Our method

According to the structural characteristics of the sentence, the sentence can be divided into simple sentences and complex sentences. A simple sentence consists of a single clause which contains a subject and a predicate and stands alone as its own sentence. However, a complex sentence is the one which is linked by conjunctions or con-

¹ <http://www.keenage.com>.

sists of at least two or more clauses without any conjunctions in it.

A complicated sentence in structure is divided into several clauses in accordance with punctuations, such as a full stop, or an exclamatory mark, or a comma, or a semicolon, etc. We analyze the syntax of the clause by extracting the clause including SAA and the adjacent one. We extract SAAOL keywords in the selected clauses, and then analyze the grammatical relationship between the keywords and SAA.

Wang et al.'s research of extraction technology based on the dependency relation of Chinese sentimental elements indicated that the dependency analyzer designed by Stanford University had not showed a high rate of accuracy. And the wrong dependency relation will interfere with the subsequent parsing process seriously (Wang, et al., 2009).

Taking the above factors into consideration, we have not analyzed the dependency relation at present. Through studying abundant instances, we specialize in the structural relationship between the keywords and SAA to extract the relation patterns which have a higher occurrence frequency. In the meantime, inverse words are processed particularly. Eventually we supplemented modification of the inaccuracy of automatic segmented words and some special adverbs, such as 偏|prejudiced, 过|excessive, 太|too.

To sum up, based on the word library SAAOL and structural analysis, SAA classification procedures are as follows:

- Step 1 Extract unidentified clauses including SAA;
- Step 2 Extract the keywords in SAAOL from the clause;
- Step 3 Label the sentiment tendency of each sentiment word by using SAAOL;
- Step 4 Discriminate the positive or negative meaning of a sentence in accordance with the different relationships. If there are no keywords in the sentence, perform step 5; otherwise, discrimination is over.
- Step 5 Extract the clauses next to SAA, and identify them according to Steps 2-4. If there are no extractable clauses, mark them as SAA which will be recognized. A is for the positives, and B for the negatives.

4 Evaluation

In disambiguating sentiment ambiguous adjectives task of SemEval-2010, there are 2917 instances in test data for 14 Chinese sentiment ambiguous adjectives. According to the official result of the task, our micro average accuracy is 0.942, which puts our system in the first position among the participants.

Depending upon the answers from organizers of the task, we notice that errors occur mainly in the following cases.

Firstly, there is a key word related to SAA, but it has no such key word in our SAAOL.

For instance,

为什么我的电脑上 pf 使用率很<head>高</head>啊 | Why is the usage rate of pf so high in my computer?

“pf 使用率|The usage rate of pf” should be NSAA, but it does not exist in our SAAOL.

Secondly, the sentence itself is too complicated to be analyzed effectively in our system so far.

Thirdly, as the imperfection of SAAOL itself, there are some inevitable mistakes in it.

For instance,

这位跳水运动员的动作难度很<head>大</head> | The diver's feat is extremely difficult.

It is generally known that if the bigger the difficulty of the dive is, the better the diver's performance will be, both of which are of proportional relation. However, generally speaking, the degree of difficulty is negative. For this reason, we made a mistake in such instance.

5 Conclusions

In this paper, we describe the approach taken by our systems which participated in the disambiguating sentiment ambiguous adjectives task of SemEval-2010.

We created a new word library named SAAOL. Through gathering words from relative dictionaries, HowNet, and other network resources, we discriminated manually the positive and negative meaning, PSAA, NSAA, and inverse words in 50,000 Chinese words according to *Richard Xiao's Top 50,000 Chinese Word Frequency List*. And then we extended the keywords in SAAOL by applying Yang's (2009) method and added synonymic and antonymous words in it. Eventually the new extended keywords were proofread and revised manually.

Based on SAAOL and structural analysis, we describe a procedure to disambiguate sentiment

ambiguous adjectives. Evaluation results show that this approach achieves good performance in the task.

References

- Qun Liu, JianSu Li. 2002. Calculation of semantic similarity of words based on the HowNet. *The Third Chinese Lexical Semantics Workshop*. Tai Bei.
- Jilin Shi, Yinggui Zhu. 2005. *A Dictionary of Positive Words*. Lexicographical Publishing House, Chengdu, Sichuan.
- Su Ge Wang, An Na Yang, De Yu Li. 2009. Research on sentence sentiment classification based on Chinese sentiment word table. *Computer Engineering and Applications*, 45(24):153-155.
- Qian Wang, TingTing He, et al. 2009. Research on dependency Tree-Based Chinese sentimental elements extraction, *Advances of Computational Linguistics in China*, 624-629.
- Janyce Wiebe, Theresa Wilson, et al. 2004. Learning subjective language. *Computational Linguistics*, 30(3): 277-308.
- Yunfang Wu, Peng Jin. SemEval-2010 task 18: Disambiguating sentiment ambiguous adjectives.
- Yu bing Yang, Xian wei Wu. Improved lexical semantic tendentiousness recognition computing. 2009. *Computer Engineering and Applications*, 45(21): 91-93.
- Ling Yang, Yinggui Zhu. 2005. *A Dictionary of Negative Words*. Lexicographical Publishing House, Chengdu, Sichuan.