

CityU-DAC: Disambiguating Sentiment-Ambiguous Adjectives within Context

Bin LU and Benjamin K. TSOU

Department of Chinese, Translation and Linguistics &
Language Information Sciences Research Centre
City University of Hong Kong
{lubin2010, rlbt sou}@gmail.com

Abstract

This paper describes our system participating in task 18 of SemEval-2010, i.e. disambiguating Sentiment-Ambiguous Adjectives (SAAs). To disambiguating SAAs, we compare the machine learning-based and lexicon-based methods in our submissions: 1) Maximum entropy is used to train classifiers based on the annotated Chinese data from the NTCIR opinion analysis tasks, and the clause-level and sentence-level classifiers are compared; 2) For the lexicon-based method, we first classify the adjectives into two classes: intensifiers (i.e. adjectives intensifying the intensity of context) and suppressors (i.e. adjectives decreasing the intensity of context), and then use the polarity of context to get the SAAs' contextual polarity based on a sentiment lexicon. The results show that the performance of maximum entropy is not quite high due to little training data; on the other hand, the lexicon-based method could improve the precision by considering the polarity of context.

1 Introduction

In recent years, *sentiment analysis*, which mines opinions from information sources such as news, blogs, and product reviews, has drawn much attention in the NLP field (Hatzivassiloglou and McKeown, 1997; Pang et al., 2002; Turney, 2002; Hu and Liu, 2004; Pang and Lee, 2008). It has many applications such as social media monitoring, market research, and public relations.

Some adjectives are neutral in sentiment polarity out of context, but they could show

positive, neutral or negative meaning within specific context. Such words can be called dynamic sentiment-ambiguous adjectives (SAAs). However, SAAs have not been intentionally tackled in the researches of sentiment analysis, and usually have been discarded or ignored by most previous work. Wu et al., (2008) presents an approach of combining collocation information and SVM to disambiguate SAAs, in which the collocation-based method was first used to disambiguate adjectives within the context of collocation (i.e. a sub-sentence marked by comma), and then the SVM algorithm was explored for those instances not covered by the collocation-based method. According to their experiments, their supervised algorithm achieves encouraging performance.

The task 18 at SemEval-2010 is intended to create a benchmark dataset for disambiguating SAAs. Given only 100 trial sentences, but not provided with any official training data, participants are required to tackle this problem data by unsupervised approaches or use their own training data. The task consists of 14 SAAs, 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. This task deals with Chinese SAAs, but the disambiguating techniques should be language-independent. Please refer to (Wu and Jin, 2010) for more descriptions of the task.

In our participating system, the annotated Chinese data from the NTCIR opinion analysis tasks is used as training data with the help of a combined sentiment lexicon. A machine learning-based method (namely maximum entropy) and the lexicon-based method are compared in our submissions. The results show that the performance of maximum entropy is not quite high due to little training data; on the other hand, the lexicon-based method could improve

the precision by considering the context of SAAs. In Section 2, we briefly describe data preparation of sentiment lexicon and training data. Our approaches for disambiguating SAAs are given in Section 3. The experiment and results are presented in Section 4, followed by a conclusion in Section 5.

2 Data Preparation

2.1 Sentiment Lexicon

Several traditional Chinese resources of polar words/phrases are collected, including NTU Sentiment Dictionary¹, *The Lexicon of Chinese Positive Words* (Shi and Zhu, 2006), *The Lexicon of Chinese Negative Words* (Yang and Zhu, 2006), and CityU's sentiment-bearing word/phrase list (Lu et al., 2008), which were manually marked in the political news data by trained annotators (Benjamin and Lu, 2008). Sentiment-bearing items marked with the *SENTIMENT_KW* tag (SKPI), including only positive and negative items but not neutral ones, were also automatically extracted from the Chinese sample data of NTCIR-6 OAPT (Seki et al., 2007). All these polar item lexicons were combined, and the combined polar item lexicon consists of 13,437 positive items and 18,365 negative items, a total of 31,802 items.

2.2 Training Data

The training data is extracted from the Chinese sample and test data from the NTCIR opinion analysis task, including NTCIR-6 (Seki et al., 2007), NTCIR-7 (Seki et al., 2008) and NTCIR-8 (Seki et al., 2010). The NTCIR opinion analysis tasks provide an opportunity to evaluate the techniques used by different participants based on a common evaluation framework in Chinese (simplified and traditional), Japanese and English.

For data from NTCIR-6 and NTCIR-7, three annotators manually marked the polarity of each opinionated sentence, and the lenient polarity is used here as the gold standard (please refer to Seki et al., 2008 for explanation of lenient and strict standard). For each opinionated sentence from NTCIR-8, only two annotators marked and the strict polarity is used as the gold standard. The traditional Chinese sentences are transferred into simplified Chinese. In total, there are about 12K opinionated sentences annotated with polarity, out of which about 9K are marked as

positive or negative, and others neutral. All the 9K sentences plus the 100 sentences from the trial data are used as the sentence-level training data.

Meanwhile, we also try to get the clause-level training data since the context of collocation within sub-sentences are quite crucial for disambiguating SAAs according to Wu et al. (2008). From the 9K positive/ negative sentences above, we automatically extract the clause for each occurrence of SAAs.

Note the polarity for a whole sentence is not necessarily the same with that of the clause containing SAAs. Consider the sentence *在当前的世界大格局中，中俄两国相互支持* (*In the current large circumstance of the world, China and Russia support each other*). The polarity of the whole sentence is positive, while the clause *在当前的世界大格局中* (*In the current large circumstance of the world*) containing a SAA *大* (*large*) is neutral, and the polarity lies in the second part of the whole sentence, i.e. *相互支持* (*support each other*).

Thus, we manually checked the polarity of clauses containing SAAs. Due to time limitation, we only checked 465 clauses. Plus the clauses extracted from 100 trial sentences, the final clause-level training data consist of 565 positive/negative clauses containing SAAs.

3 Our Approach for Disambiguating SAAs

To disambiguating SAAs, we use the maximum entropy algorithm and the sentiment lexicon-based method, and also combine them together.

3.1 The Maximum Entropy-based Method

Maximum entropy classification (MaxEnt) is a technique which has proven effective in a number of natural language processing applications (Berger et al., 1996). Le Zhang's maximum entropy tool² is used for classification.

The Chinese sentences are segmented into words using a production segmentation system. Unigrams of words are used as basic features for classification. Bigrams are also tried, but does not show improvement, and thus are not described in details here.

3.2 The Lexicon-based Method

For the lexicon-based method, we first classify the 14 adjectives into two classes: intensifiers

¹ <http://nlg18.csie.ntu.edu.tw:8080/opinion/index.html>

² http://homepages.inf.ed.ac.uk/lzhang10/maxent_toolkit.html

and suppressors. Intensifiers refer to adjectives intensifying the intensity of context, including 大|big, 多|many, 高|high, 厚|thick, 深|deep, 重|heavy, 巨大|huge, 重大|grave, while suppressors refer to adjectives decreasing the intensity of context, including 小|small, 少|few, 低|low, 薄|thin, 浅|shallow, 轻|light.

Meanwhile, the collocation nouns are also classified into two classes: positive and negative. Positive nouns include 素质|quality, 标准|standard, 水平|level, 效益|benefit, 成就|achievement, etc. Negative nouns include 压力|pressure, 差距|disparity, 问题|problem, 风险|risk, 污染|pollution etc.

The hypothesis here is that intensifiers will receive the polarity of their collocations while suppressors will get the opposite polarity of their collocations. For example, 成就|achievement could be collocated with one of the following intensifiers: 大|big, 多|many or 高|high, and the adjectives just receive the polarity of 成就|achievement, which is positive. Meanwhile, 污染|pollution could be collocated with one of the following suppressors: 小|small, 少|few, 低|low, and the adjectives just receive the opposite polarity of 污染|pollution, which is also positive.

Based on this hypothesis, we could get the polarity of SAAs through their collocation nouns within the clauses containing SAAs. The context of SAAs is a sub-sentence marked by comma. The sentiment lexicon mentioned in Section 2.1 is used to find polarity of collocation nouns.

3.3 Combining Maximum Entropy and Lexicon

To combine the two methods above, the lexicon-based method is first used to disambiguate the sentiment of SAAs, and the context of collocation is a sub-sentence marked by comma. Then for those instances that are not covered by the lexicon-based method, the maximum entropy algorithm is explored.

4 Experiment and Results

The dataset contains two parts: some sentences were extracted from Chinese Gigaword (LDC corpus: LDC2005T14), and other sentences were gathered through the search engine like Google. Firstly, these sentences were automatically segmented and POS-tagged, and then the ambiguous adjectives were manually annotated

with the correct sentiment polarity within the sentence context. Two annotators annotated the sentences double blindly, and the third annotator checks the annotation. All the data of 2,917 sentences is provided as the test set, and evaluation is performed in terms of micro accuracy and macro accuracy.

We submitted 4 runs: run 1 is based on the sentence-level MaxEnt classifier; run 2 on the clause-level MaxEnt classifier; run 3 is got by combining the lexicon-based method and the sentence-level MaxEnt classifier; and run 4 by combining the lexicon-based method and the clause-level MaxEnt classifier. The official scores for the 4 runs are shown in Table 2.

Table 2. Results of 4 Runs

Run	Micro Acc. (%)	Macro Acc. (%)
1	61.98	67.89
2	62.63	60.85
3	71.55	75.54
4	72.47	69.80

From Table 2, we can observe that:

- 1) Compared the highest scores achieved by other teams, the performance of maximum entropy (run 1 and 2) is not quite high due to little training data;
- 2) By integrating the lexicon-based method and maximum entropy (run 3 and 4), we improve the accuracy by considering the context of SAAs;
- 3) The sentence-level maximum entropy classifier shows better macro accuracy, and clause-level one better micro accuracy.

In addition to the official scores, we also evaluate the performance of the lexicon-based method alone. The micro and macro accuracy are respectively 0.847 and 0.835665, showing that the lexicon-based method is more accurate than the maximum entropy algorithm (run 1 and 2). But it only covers 1,436 (49%) of 2,917 test instances.

Because the data from the NTCIR opinion analysis task is not specifically annotated for this task, and the manually checked clauses are less than 600, the performance of our system is not quite high compared to the highest performance achieved by other teams.

5 Conclusion

To disambiguating SAAs, we compare machine learning-based and lexicon-based methods in our submissions: 1) Maximum entropy is used to train classifiers based on the annotated Chinese data from the NTCIR opinion analysis tasks, and the clause-level and sentence-level classifiers are

compared; 2) For the lexicon-based method, we first classify the adjectives into two classes: intensifiers (i.e. adjectives intensifying the intensity of context) and suppressors (i.e. adjectives decreasing the intensity of context), and then use the polarity of context to get the SAAs' contextual polarity. The results show that the performance of maximum entropy is not quite high due to little training data; on the other hand, the lexicon-based method could improve the precision by considering the context of SAAs.

References

- Adam L. Berger, Stephen A. Della Pietra, and Vincent J. Della Pietra. 1996. A maximum entropy approach to natural language processing. *Computational Linguistics*, 22(1):39-71.
- Vasileios Hatzivassiloglou and Kathleen McKeown. 1997. Predicting the Semantic Orientation of Adjectives. *Proceedings of ACL-97*. 174-181.
- Minqing Hu and Bing Liu. 2004. Mining Opinion Features in Customer Reviews. In *Proceedings of the 19th National Conference on Artificial Intelligence*, pp. 755-760.
- Bin Lu, Benjamin K. Tsou and Oi Yee Kwong. 2008. Supervised Approaches and Ensemble Techniques for Chinese Opinion Analysis at NTCIR-7. In *Proceedings of the Seventh NTCIR Workshop (NTCIR-7)*. pp. 218-225. Tokyo, Japan.
- Bo Pang and Lillian Lee. 2008. *Opinion mining and sentiment analysis, Foundations and Trends in Information Retrieval*, Now Publishers.
- Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of EMNLP 2002*, pp.79-86.
- Yohei Seki, David Kirk Evans, Lun-Wei Ku, Le Sun, Hsin-His Chen, Noriko Kando. 2007. Overview of Opinion Analysis Pilot Task at NTCIR-6. *Proc. of the Seventh NTCIR Workshop*. Japan. 2007.6.
- Yohei Seki, David Kirk Evans, Lun-Wei Ku, Le Sun, Hsin-His Chen, Noriko Kando and Chin-Yew Lin. 2008. Overview of Multilingual Opinion Analysis Task at NTCIR-7. *Proc. of the Seventh NTCIR Workshop*. Japan. Dec. 2008.
- Yohei Seki, Lun-Wei Ku, Le Sun, Hsin-His Chen, Noriko Kando. 2010. Overview of Multilingual Opinion Analysis Task at NTCIR-8. *Proc. of the Seventh NTCIR Workshop*. Japan. June, 2010.
- Jilin Shi and Yinggui Zhu. 2006. The Lexicon of Chinese Positive Words (褒義詞詞典). Sichuan Lexicon Press.
- Benjamin K. Tsou and Bin Lu. 2008. A Political News Corpus in Chinese for Opinion Analysis. In *Proceedings of the Second International Workshop on Evaluating Information Access (EVAL2008)*. pp. 6-7. Tokyo, Japan.
- Peter D. Turney. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews, In *Proceedings of ACL-02*, Philadelphia, Pennsylvania, 417-424.
- Yunfang Wu, Miao Wang, Peng Jin and Shiwen Yu. 2008. Disambiguate sentiment ambiguous adjectives. In *Proceedings of IEEE International Conference on Natural Language Processing and Knowledge Engineering (NLP-KE'08)*.
- Yunfang Wu, and Peng Jin. 2010. SemEval-2010 Task 18: Disambiguate sentiment ambiguous adjectives. In *Proceedings of SemEval-2010*.
- Ruifeng Xu, Kam-Fai Wong and Yunqing Xia. 2008. Coarse-Fine Opinion Mining - WIA in NTCIR-7 MOAT Task. In *Proceedings of the Seventh NTCIR Workshop (NTCIR-7)*. Tokyo, Japan, Dec. 16-19.
- Ling Yang and Yinggui Zhu. 2006. The Lexicon of Chinese Negative Words (貶義詞詞典). Sichuan Lexicon Press.