ISCAS in Opinion Analysis Pilot Task: Experiments with sentimental dictionary based classifier and CRF model

Ruihong Huang，Longxi Pan，Le Sun
Institute of Software, Chinese Academy of Sciences
P.O.Box 8718, Beijing, 100080, P.R.China
ruihong05@iscas.cn

Abstract

The paper presents our work in the opinion pilot task in NTCIR6 in Chinese. In extracting opinion holders, we applied Conditional Random Field (CRF) model to find the opinion holders as a sequential labeling task, while in determining the subjectivity and the polarity, we adopted a simple empirical algorithms based on the sentimental dictionary to discriminate the subjective sentences from the objective ones and suggest their polarities. Besides the features used in the CRF model and the detailed specification in the machine learning system, the evaluation results and the error analysis will also be presented.

Keyword: *opinion analysis, Conditional Random Field, sentimental dictionary based classifier.*

1 Introduction

The existences of a large amount of written subjective texts referring to the quality of commercial products or the political selections and of the desire to extract and analyze the opinions expressed by the consumers and the voters have made opinion analysis a quite attracting and active research domain recently. Besides, discriminating the subjective part of the texts from the objective parts might be employed to improve the performance of many other NLP applications, such as text summarization, question answering.

There are four subtasks defined at the sentence level in the opinion pilot task in NTCIR6, including recognizing whether a sentence is subjective, is yes, determining its polarity, the relevance to a certain topic and the holder of the opinion. We participated in all subtasks except determining the relevance of a sentence to a certain topic.

The remainder of the paper is organized as follows: firstly, we will present the related work in section 2. Then our methods to determine the subjectivity and polarity and to extract opinion holders will be described separately in section 3 and 4. Finally, conclusions and future work will be presented in section 5.

2 Related Work

Subjectivity detection is the task of identifying subjective words, expressions ([4]), and sentences ([6], [3]), or documents ([5], [7]). Sentiment detection is the task of determining positive or negative sentiment of words ([6], [7]), phrases and sentences ([2]), or documents ([5]). Building on this work, more sophisticated problems such as opinion holder identification have also been studied.

In extracting the source holders, two research efforts, [8], [9], are closed related with ours. Specifically, in [8], a hybrid approach that combines Conditional Random Fields and a variation of AutoSlog (used for pattern extraction) was adopted applied, while in [9], the Maximum Entropy Ranking (MER) model was applied. In feature selection, both used rich structural features derived from syntactic parsers. In contrast, we just used limited grammatical information inherent in the Part-Of-Speech (POS) tagging and Named Entity Recognizing (NER) processing.

3 Determining the Subjectivity and the Polarity subtask

3.1 some considerations

Our aim is to find a simple but effective way that can perform well in the binary decision of sentence. Although many researchers had adopted supervised machine-learning methods to discriminate subjective sentences from objective ones in English employing a variety of lexical and contextual features, considering that machine-learning methods seem too complicated for our Chinese task, we used a simple empirical algorithms based on the sentimental dictionary to suggest the subjectivity and the polarity in sentence level.

Since our method was based on the sentimental dictionary, the first obstacle toward deciding the subjectivity and polarity of a sentence was how to collect as many emotion words as possible. For this, we used the emotion dictionary offered by NTCIR downloaded from its website as a base, then we enlarge the vocabulary by consulting tong2yi2ci2lin of HIT.

3.2 Algorithms description

According to [1], the meaning of a Chinese sentiment word is a function of the composite Chinese characters. P_{C_i} and N_{C_i} denote the weights of C_i as positive and negative characters respectively which can be computed by the times it appears in positive and negative dictionary. And they postulated an empirical formula (see Formula 1) to calculate the sentimental tendency of characters $C_1, C_2 \ldots\ldots C_P$.

$$S_{C_1} = P_{C_I} - N_{C_I} \quad . \qquad (1)$$

Then a sentiment score of a Chinese word *w* (see Formula 2) is the average of the sentiment scores of the composing characters $C_1, C_2 \ldots\ldots C_P$.

$$S_w = \tfrac{1}{P} \times \sum_{J=1}^{P} S_{C_J} \quad . \qquad (2)$$

So we can get the score of a sentence as long as we can get the scores of its comprised characters. The algorithms to attain the score of a sentence is presented below: [Sentence Level]

1. For every sentence
2. For every sentiment word in this sentence
3. If a negation operator appears before, then reverse the sentiment tendency.
4. Decide the opinionated tendency of this sentence by the function of sentiment words and the opinion holder as follows:

$$S_p = S_{opinion-holder} \times \sum_{j=1}^{n} S_{W_j}$$

. Where *Sp*, *Sopinion-holder*, and *Swj* are sentiment score of sentence *p*, weight of *opinion holder*, and sentiment score of word *wj*, respectively, and *n* is the total number of sentiment words in *p*. we do the experiment as above that we can discriminate sentences by S_P.

3.3 Evaluation results and result analysis

The evaluation results are presented in Table 1.

		P	R	F
Opinion-ated	Lenient strict	0.590 0.221	0.664 0.662	0.625 0.331
OpAnd-Polarity	Lenient strict	0.231 0.059	0.261 0.314	0.264 0.099

Table 1: Chinese opinion analysis results

Seen from the above table, the results were somewhat depressing especially the results on polarity recognizing. As an explanation to the bad results, we supposed that the semantics contained in words comprising a sentence was not enough to express the sentence's subjectivity or polarity. For example,「暴力」(violence: negative) is negative in itself definitely, however the phrase 「暴力行为」(violence activity) in a sentence is usually used to state a objective fact. Additional, words are not enough to express syntax and semantic roles with emotional tendency.

4 Extracting Opinion Holders subtask

Extracting opinion holders matching with some certain opinion words is really a hard work. To locate the opinion holders exactly, we need to diagnose the connection between the two objects, especially when there are more than one opinion holders for an opinion word and there are more than one opinion words existing in the sentences.

4.1 CRF Model

Conditional Random Field [10], model has been used successfully in a variety of sequential segmenting and labeling tasks. For its immunity to bias to some local optimization, it has improved the labeling performance.

In our source finding task, we employed a C++ implementation of Conditional Random Field model, CRF++ [11], which base its training on LBFGS algorithms, making the training procedure quite fast.

Through applying the CRF model, we integrate a variety of features presented in the next part into a universal theoretical framework.

4.2 Features used

Intuitively, most opinion holders would be supposed to be named entities (PERSON or ORGANIZATION) and probably be of noun type in Part-Of-Speech, additionally, they can be preceded by an opinion word.

With these properties in mind, we employ a variety of features in our opinion holder extraction task. Specifically, the feature annotation unit is Chinese characters and the features are not all binary as in [8]. We will outline the features we used below.

Lexical features

In Chinese, the names of persons and organizations are usually comprised of particular characters frequently. So we define the lexical features for neighboring characters in a [-2, +2] window to cover most of Chinese word length.

Part-of-speech features

We used our own POS tagger to label the Part-of-speech features.

In the experiments, we didn't filter the noun type out of the other POS types to make our part-of-speech features binary, but to directly define the output from the POS tagger as our part-of-speech features for we observed that the part-of-speech types of the words neighboring with the source holder present some patterns. So we defined the part-of-pattern features for neighboring characters in a [-1, 6] window to capture the POS patterns existing in the forwarding two or three words.

In particular, we extended the window to the character just before the current character to capture the source holder appearing at the beginning of one sentence for the ending of the last sentence was denoted by an asterisk.

Semantic class features

We applied our own Named Entity Recognizer (NER) to find out the possible candidates for source holders.

The labels for this category of features are trinary indicating whether the labeled character is in a person's name or in a organization's name or in none of the above two.

Opinion trigger features

As annotated in the training data, nearly every OPINION_SRC is followed by an OPINION_OPR phrase not too far from the according OPINION_SRC. To make use of the important cue, we extracted all the opinion trigger words in the training data to form an opinion trigger lexicon which will be listed below in Table 3. We defined the opinion trigger features to be binary by marking the words appearing in the lexicon positive, negative otherwise.

Table 3: Opinion trigger lexicon

trigram	即表示
bigram	并称 报导 要求 回答 承认 声明 表明 希望 告诉 指出 答称 表态 坚称 分析 指控 建议 提出 深表阐述 宣布 相信 认为 警告 觉得 预料 指称 宣称 以为 发表 还说 透露 预期 同意 声称 明示 预测 盛赞 表示 询问 指责 谈话 强调
unigram	指 答 道 说 称

		CRT-w	CRT-wo	P-CRT	InCRT	Miss	F-A
Sentence-based	lenient	665	1724	175	354	447	257
	strict	293	544	84	157	188	89
		CRT	InCRT	Miss	F-A	P-H	CRT-NUM
Holder-based	lenient	871	422	0	396	1689	1958
	strict	391	189	0	162	742	857

Table 2: The evaluation results for opinion holder extraction subtask

As indicated in section 2, we didn't employ complex structural features separately except for the structural information inherent in the Part-Of-Speech tagging and the Named Entity Recognizing and the opinion trigger features presenting the dependency relationship between the source holders and the opinion expressions.

4.3 Evaluation results and result Analysis

Of 32 topics in *NTCIR-6 Opinion Analysis Pilot Task* test collection in Chinese, four topics were provided as a sample (training) data to participants in Chinese [12] which were used to extract the opinion trigger words we used to generate the opinion trigger features.

In the opinion analysis pilot evaluations, we achieved an F-measure of 0.436 in sentence-based scoring and 0.489 in holder-based scoring in the source holder extracting subtask. The detailed evaluation results is presented in Table 2, the detailed meanings of the items are listed in the table are specified in [12].

Seen from Table 2, the recall and the precision scores isn't quite high, for we have omitted a lot of source holders while recognized some of none-source-holder phrases as holders wrongly at the same time. Especially, after examining the labeled data, we noticed the results depend on the opinion trigger words severely compared with the other features. We supposed several reasons might explain.

Firstly, in Chinese as in English, the opinion holders are usually closely related with the some frequent opinion trigger words and the relationship isn't too hard to capture. So even not using much structural patterns as in [8], [9], we had found out a large ratio of opinion holders exactly by using the opinion trigger words features. But the drawback is obvious, when we came across the words not appearing in the training data, we would have failed to recognize the opinion holder probably.

Secondly, for the Chinese characters appearing in person names and organization names can be large in number, there may be sparseness in the lexical features, similarly, the variation of Chinese word length may result in the sparseness of the part-of-speech and the lexical features defined on the unchangeable width of window. So the lexical and part-of-speech features employed in our experiment might be not quite efficient to extract the opinion holders. Besides, the original part-of-speech features comprised of all kinds of POS labels may be more efficient when used after some preprocessing to distinguish some labels such as noun or pronoun.

Thirdly, the structural information included in our features might be not powerful enough to capture the various complex structural relationships between the source holder and opinion expression on one hand or to put enough constraints to the sentence structures to limit the model to only extract the correct opinion holder on the other hand, so it's quite possible that we have omitted some opinion holders while got some incorrect opinion holders at the same time.

The last but not the least important reason is that the errors generated in the POS tagging and the NER may propagate and affect our final opinion holder extracting result.

Below, we will present some typical false instances. Doubly underlined phrases indicate incorrectly extracted sources (either false positives or false negatives). Opinion words are singly underlined.

（1）柯恩在转往澳洲进行为期两天的访问时表示：「我认为， …

（2）柯恩坦承， …

（3）美国国防部长柯恩今天在澳洲堪培拉表示， …

In (1), (2), we all failed to extract the opinion holder for the distance from the opinion word larger than the window width of ten set for the opinion trigger features in (1) and the failure of the opinion word lexicon to include “坦承”. In contrast, in (3), aside from the correct opinion holder “美国国防部长柯恩”, we extract the location phrase “澳洲堪培拉” incorrectly, we supposed it’s for the lack of enough structural information.

5 Conclusions and Future Work

In the subtask of determining the sentence subjectivity and the polarity, we haven’t got good evaluation scores using our rule-based classifier. After speculation, we realized that it was necessary to incorporate other lexical features that can reflect syntax and semantic roles in sentences including not only single words, multi-word N-grams, but also phrases and lexicon-syntactic patterns to improve the classifying performance.

When we introduce these more complicated lexical features, machine learning methods are more suitable for sentence classify compared with rule-based way. To implement the machine-learning algorithms in Chinese, we can employ classical bag-of-features classifiers such as Naïve Bayes, ME and SVM, besides, we might need to do some efforts to collect some primitive resources.

In our experiments for opinion holder extraction, we used the Conditional Random Field model to integrate the lexical, grammatical and semantic features together efficiently and achieved No.2 in the evaluation results.

Through the analysis to the results of the opinion holder extraction subtask, we believe exploring more rich structural features to present the close and complex relationships between the opinion words and the opinion holders will improve both the recall and the precision. We also doubted the reasonability to the design of the part-of-speech and the lexical features defined on an unchangeable width of window, we may do some refinements to it in the future.

References

[1] Ku, L-W., Liang, Y-T. and Chen, H-H. Opinion extraction, summarization and tracking in news and blog corpora. *AAAICAAW'06*, 2006.

[2] S-M. Kim and E. Hovy. 2004. Determining the sentiment of opinions. In *Coling 2004*.

[3] J. Wiebe and E. Riloff. 2005. Creating subjective and objective sentence classifiers from unannotated texts. In *CICLing-2005*.

[4] Riloff, E. and Wiebe, J. Learning extraction patterns for subjective expressions. *Proceedings of the 2003 Conference on EMNLP*, pages 105-112. 2003.

[5] Pang, B., Lee, L., and Vaithyanathan, S. Thumbs up? Sentiment classification using machine learning techniques. *Proceedings of the 2002 Conference on EMNLP*, pages 79-86. 2002.

[6] Hatzivassiloglou, V., & McKeown, K.R. 1997. Predicting the semantic orientation of adjectives. *Proceedings of the 35th Annual Meeting of the ACL and the 8th Conference of the European Chapter of the ACL* (pp. 174-181). New Brunswick, NJ: ACL.

[7] Peter Turney. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In Proc. of the ACL.

[8] Y. Choi, C. Cardie, E. Riloff, and S. Patwardhan. Identifying sources of opinions with conditional random fields and extraction patterns. In *Proc. of the 2005 Human Language Technology Conf. and Conf. on Empirical Methods in Natural Language Processing (HLT/EMNLP 2005)*, Vancouver, B. C., 2005.

[9] S. Kim and E. Hovy. 2005. Identifying opinion holders for question answering in opinion texts. In *Proceedings of AAAI Workshop on Question Answering in Restricted Domains*.

[10] J. Lafferty, A. K. McCallum & F. Pereira. 2001. Conditiona Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *Proceedings of 18th International Conference on Machine Learning*.

[11] http://chasen.org/~taku/software/CRF++/

[12] Yohei Seki, David Kirk Evans, Lun-Wei Ku, Hsin-Hsi Chen, Noriko Kando, and Chin-Yew Lin. Overview of Opinion Analysis Pilot Task at NTCIR-6. In Proceedings of the Sixth NTCIR Workshop.