Detecting Opinionated Sentences by Extracting Context Information[1]

Meng Xinfan, Wang Houfeng
Institute of Computational Linguistic
School of Information Science and Technology
Peking University
{mxf, wanghf}@pku.edu.cn

Abstract

In this paper, we briefly describe several experimental methods to solve MOAT at NTCIR-7. In the subtask of opinionated sentence detection, two methods aiming to extract the context information of each sentence are proposed. Maximum Entropy model is used to predict the polarity class. A rule-based pattern matching scheme is devised to find topic-relevant sentence. For the subtask of detecting holders and targets, the CRF model is adopted.

Keywords: *opinion analysis, maximum entropy, conditional random field*

1. Introduction

Opinion analysis or sentiment analysis refers to a broad area of natural language processing. It aims to determine the attitude or opinion of a speaker or writer with respect to a specific topic. This research area is relatively new in NLP but has drawn extensive attention in recent years. A basic task of opinion analysis is to determine if a document or a sentence expresses a subjective attitude of a speaker or writer. Other tasks in opinion analysis include: determining the polarity of an opinionated document or sentence; deciding the relevance between an opinionated sentence and a topic; extracting the opinion holder and opinion target from an opinionated expression，etc.

Many traditional methods have been adopted in opinion analysis ([1], [3]). However, the traditional machine learning algorithms used in NLP application do not perform well enough in opinion analysis ([1]).

Multilingual Opinion Analysis Task (MOAT) at NTCIR-7 includes all the tasks described above: opinion detection, polarity analysis, relevance determination, holder and target detection. We participated in all 5 evaluation subtasks in both Traditional Chinese side and Simplified Chinese side.

In the following sections, we first describe our approaches adopted in each subtask, then analyze the evaluation result and finally present our conclusion.

2. Opinionated Sentence Detection

Opinionated sentence detection is the only mandatory subtask defined in MOAT and is also the basis for further opinion analysis. Two groups of results are provided for this subtask. These results are predicted via two different models: Maximum Entropy model and Conditional Random Field model.

2.1. Maximum Entropy model

Maximum Entropy model is used in many NLP tasks such as opinion analysis. In previous research, each sentence is treated as an independent unit and features used typically are unigrams, bigrams or Part-Of-Speech tags ([1]).

The feature selection method mentioned above is easy to perform. However, the assumption that sentences are independent of each other is too rigid. In fact a sentence is usually related to its context.

The opinion of a sentence is directly related to the type of document it belongs to. For example, the sentences in a poem are mostly subjective while the sentences in a news story are largely objective. Thus besides the words and POS tags features, we introduce the global document subjectivity degree feature into Maximum Entropy model. This feature is defined as follows: let **p** be the percentage of opinionated sentence in the document and **D** be the global document subjectivity degree, then:

1. If **p** > 60%, **D** is high
2. If **p** < 30%, **D** is low
3. Otherwise **D** is medium

In addition, some other features are also considered. We assumed that the occurrence of a number suggests objectivity, as in the sentence “美國線上、時代華納換股合並-總值三千五百億美元歷來最大合並案將成首家全球性媒體與通信公司美股市開盤上揚” and sentence “合並後，美國線上的股東將持有新集團百

[1] Supported by National Natural Science Foundation of China (No. 60675035) and Beijing Natural Science Foundation (No. 4072012).

分之五十五的大多數股份，時代華納持有百分之四十五。”. We also assumed the occurrences of an opinion word and/or a proposal word indicate subjectivity. For example, the proposal word “稱” (means “say”) in sentence “但凱斯不為所動，他堅稱線上的業務應是重行銷而輕科技，到目前為止他這種遠見還是正確的，這使美國線上目前以兩千萬訂戶身居線上服務業盟主。” and the opinion word “懷疑” (means “suspicion”) in sentence “外界則更有懷疑” both indicate strong subjectivity bias. In our previous experiment with NTCIR-6 datasets, adding these clues into the classification models resulted in a 3 to 5 percent improvement in F-measure. Thus in our experiment, the number-exist indicator feature is used if the sentence contains a number. The opinion-word feature is set if the sentence contains a word in our pre-built opinion word dictionary. And the proposal-word feature is included if the sentence contains a word in proposal word dictionary. The proposal word dictionary is built from NTCIR-6 MOAT task sample release and the opinion word dictionary is combined from NTCIR-6 MOAT task samples and NTU sentiment dictionary.

To classify a sentence with this method, we conduct a two phase classification. In the first phase, we find the global subjectivity degree of the document to which it belongs. Support Vector Machine is an effective model to do document classification. We first extract the features from test dataset, which include words, POS tags, then train the SVM model with the documents in NTCIR-6 opinion analysis pilot task, and finally predict the global subjectivity degree of each document to be tested. In the second phase, we introduce global subjectivity degree into Maximum Entropy model to perform the opinionated/objective binary classification.

2.2. Conditional Random Field Model

Conditional Random Field (CRF) is a sequential labeling model, which is a more aggressive way to extract the context information. We employed this model as an alternative way to detect opinionated sentence.

There are two ways to use this model. One is to classify based on sentences and the other is based on words. We choose the second way since that will make it relatively easy to select features. Here we simply use words, POS tags and opinion-word indicators as features. Other features like number-exist feature is not included to simplify our features templates used in the CRF model.

To do the classification, we segment each document into a sequence of words, generate the features (words, POS tags and opinion-word indicator), and label each word as either opinionated or objective with CRF. After the labeling procedure finished, we count the number of each type of words in every sentence, and then tag a sentence as opinionated if it contains more opinionated words than objective words.

3. Polarity Analysis

The polarity classification task is done independently with the opinionated sentence detection subtask. We treat this subtask as a 4-class classification procedure. The 4 classes are positive, negative, neutral and non-opinionated.

Maximum Entropy model is used to predict the polarity class. Features here used contain words, part-of-speech tags, number-exist indicator, proposal-words indicator and opinion-words indicator.

Since we conduct the opinionated sentence detection and polarity classification independently, inconsistency will arise between those two results. We resolve this inconsistency with some simple rules:

1. If a sentence is predicted to be opinionated in opinionated sentence detection but received a non-opinionated class in polarity classification task, its polarity class would be changed to neutral.
2. Conversely, if a sentence is predicted to be non-opinionated in opinionated sentence detection task but receive one of the {positive, negative, neutral} classes in polarity classification task, we would assign a non-opinionated class to this sentence.

4. Relevance Analysis

The topic relevance detection task is done via a rule-based approach.

The rules are:

1. Multiple patterns are defined for every topic. Each pattern is a modified version of regular expression.
2. A pattern is a set of words concatenated by operators which include +, *, ~ and #. + is “or” operator, * is “and” operator, \ is “exclude” operator and # is “omit” operator.
3. If a sentence satisfies one of the patterns, we judge that the sentence is relevant to this topic.

For example, we defined a pattern {核试,核爆, 爆,试爆,核子爆炸}+{核子,核武}*{试验,实验,测试} for topic N09 in traditional Chinese evaluation side. According to the pattern mentioned above, the sentence “可能为次临界温度的核子试爆，亦即用来模拟核爆的小规模爆炸” and the sentence “中共上次进行核子相关测试是在一九九九年” satisfy this pattern, while the sentence “印度自从一九七四年首度进行核子试爆後，该国政客、科学家及战略专家无不为这个模棱两可，进退维谷的核武政策感到头痛” does not satisfy the pattern.

Patterns are generated via the following steps:

1. First, we extract the keywords from topic description;
2. Then we expand this keywords to a larger set by searching them in the search engines or by finding similar words in documents;
3. Finally we inspect the documents to revise the keywords set and devise the pattern.

5. Holder and Target Detection

Holder and target detection is very similar to the Name Entity Recognition task. Both aim to extract some entities from a sentence. Thus we treated this task as a sequence labeling task and adopted CRF Model. NTCIR-6 opinion analysis pilot task release data and NTCIR-7 MOAT sample release data are used to train the CRF model.

The features we used include three categories:

1. Neighbor words in a [-4, 4] window;
2. Neighbor POS tags in a [-4, 4] window;
3. Neighbor proposal tags in a [-4, 4] window. A proposal tag of a word is set to YES if this word belongs to the proposal word dictionary defined previously.

In all those features used above, we set all the windows to the maximum value CRF++ can accept, which can improve the model's ability to detect very long patterns but might also introduce some noises.

The label set {O, H, T} is used. If a word is labeled as H/T/O, it means the word belongs to a holder/target/other entity.

We conduct the opinion sentence detection and holder/target recognition task independently, therefore the results of these two tasks may not be consistent. The following methods are presented to resolve the possible contradiction:

1. Many sentences directly express the opinion of the author, whose name is usually not present in the text. Thus if a sentence is labeled as opinionated but has no recognized holder, we set the holder field of this sentence to POST_AUTHOR. However, we are unable to find a default value or pattern for target field. So we simply set the target field to blanks when contradiction happens.
2. If a sentence is labeled as non-opinionated but has recognized holder/target, we just remove the holder and target field.

6. Evaluation and Analysis

We participated in all subtasks in both Traditional Chinese and Simplified Chinese side.

As stated in previous sections, we provide two groups of results for opinionated sentence detection subtask. One is predicted by Maximum Entropy model (iclpku-1) and the other is by Conditional Random Field model (iclpku-2). In Traditional Chinese side, iclpku-1 achieves a relatively high precision 0.7, while the recall is not very high, just 0.63. For iclpku-2, on the contrary, the recall is higher than precision. But in Simplified Chinese side, the performances of iclpku-1 and iclpku-2 are very similar: precisions are both very low and recalls are both high. The topics in Traditional Chinese and Simplified Chinese corpus are largely overlapped, thus this distinction might reflect some tagging-scheme-inconsistency between the Traditional Chinese and Simplified Chinese corpus.

Table 1 Opinionated Sentence Detection Traditional Chinese

	Lenient			Strict		
	P	R	F	P	R	F
Iclpku1	0.70	0.63	0.66	0.86	0.70	0.77
Iclpku2	0.58	0.74	0.65	0.74	0.79	0.76

Table 2 Opinionated Sentence Detection Simplified Chinese

	Lenient			Strict		
	P	R	F	P	R	F
Iclpku1	0.48	0.80	0.60	0.45	0.82	0.58
Iclpku2	0.44	0.80	0.57	0.40	0.83	0.54

The polarity analysis subtask has very strong connection with opinionated sentence detection. As expected, the polarity subtask results reveal patterns that are very similar to those ones in opinionated sentence detection.

The relevance detection subtask is very challenging. But with a simple rule-based pattern extraction method, we achieve the 5th highest result on F-measure (strict) in Traditional Chinese side. This result suggests that the keywords we pick can effectively distinguish the relevant sentences from irrelevant ones on Traditional Chinese dataset. However, since most sentences in Simplified Chinese dataset are topic-relevant, it is difficult to evaluate performance of our method on this dataset.

In holder detection subtask, with a simple method, we achieve a relatively high result. But the same method does not generalize well on target detection subtask. Two reasons can explain this performance difference:

1. There exist some strong clues to suggest the occurrence of holders, like "专家认为" or "他表示",but the patterns for detecting target are much less obvious.
2. The dataset for training CRF model contains more instances of holder than of target.

7. Conclusion and Future Works

All the approaches we adopted here are relatively simple, for the reasons that we believe simple method could be effective if it is in the right direction and we don't have much experience in this area.

Basically, the results are close to our expectation but not to our satisfaction. In the opinionated sentence detection subtask, the two simple methods we provided achieve an average result. To improve the performance of models, some methods could be used in future work:

1. For Maximum Entropy model, it is necessary to revise the features selection scheme, and more opinion indicator features should be introduced into models.
2. For Conditional Random Field model, predicting based on words is too artificial. It is more natural to predict based on sentences or clauses.

In the subtask of detecting holders and target, the same methods get completely different results. CRF model is suitable for holder detection subtask, to further improve the performance, it is necessary to change the template patterns and the features. On the other hand, we should try other methods and models to detect target.

Reference

[1] 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.
[2] Ku, L.-W., Liang, Y.-T. and Chen, H.-H. Opinion extraction, summarization and tracking in news and blog Corpora. *Proceedings of AAAI-2006 Spring Symposium on Computational Approaches to* Analyzing Weblogs, AAAI Technical Report. Pages 100-107. 2006.
[3] 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/EMNLP2005)*, Vancouver, B. C., 2005
[4] E. Riloff and J. Wiebe. Learning extraction patterns for subjective expressions. *Proceedings of the 2003Conference on EMNLP*, pages 105-112. 2003.
[5] 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.
[6] H. Yu and V. Hatzivassiloglou, "Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences, in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2003.
[7] http://chasen.org/~taku/software/CRF++/
[8] http://homepages.inf.ed.ac.uk/s0450736/maxent_toolkit.html