ISCAS at Multilingual Opinion Analysis Task

Yunping Huang, Yulin Wang, Le Sun
Institute of Software, Chinese Academy of Sciences
P.O.Box 8718, Beijing, 100190, P.R.China

Abstract

The paper presents our work at the multilingual opinion analysis task in Simplified Chinese in NTCIR7. In detecting opinionated sentences, an EM algorithm was proposed to extract the sentiment words based on the sentimental dictionary, and then an iterative algorithm was used to estimate the score of the sentiment words and the sentences. In detecting relevant sentences, we solve this problem by analogizing the task to the traditional information retrieval task. The difficulty lies in that some sentence is relevant to the topic even if there are no key words hit in it. In this situation, we use a pseudo feedback and query extension method to refine the result. The evaluation results and the result analysis will also be presented.

Keywords: NTCIR, Opinion Analysis.

1. Introduction

The processing of opinion information has been widely discussed these days. There are a large amount of written subjective texts. People are concerned about opinion existed in these data, which makes opinion analysis a quite attracting and active research domain recently.

There are five subtasks defined at sentence level in MOAT task. We participated in two subtasks: detecting the opinion sentences, and detecting the relevant sentences. In the subtask of detecting opinion sentences, we use a lexicon based algorithm, which considers the sentiment words in the sentiment dictionary as the seed sentiment words, and uses the statistical method to expand the sentiment words. At last, we compute the opinion score of a sentence through the sentiment words. In the subtask of detecting relevant sentences, We borrow the pseudo feedback and query extension method to refine the topics to improve the recall rate.

2. Detecting Opinionated Sentences

Discriminating subjective sentences from objective ones can be considered as a binary classification problem. Many researchers had adopted supervised machine-learning method for it. However, it needs a variety of lexical and contextual features. These machine-learning methods seem too complicated for our Chinese task, so we used a simple empirical algorithms based on the sentiment lexicons to detect opinionated sentences.

The first problem is how to collect as many sentiment words as possible since our method was based on the sentimental lexicons. For this, we use several existed emotion dictionaries to extract seed sentiment words. After these operations, we can get a relatively big emotion dictionary. However, this can not cover all sentiment words. Many sentiment words will not be included by these dictionaries, and there are also many topic-related sentiment words, thus, we need to use statistical method to extract sentiment word. Turney used the statistical method to extract the sentiment words.

We assume that the word usually occurred around sentiment word will have a big possibility to be sentiment word. Based on this assumption, we proposed an EM algorithm to extract the sentiment words. After the EM algorithm converges, we select M words as sentiment words. The most frequent words and the word which just contains one character are removed.

After extracting the sentiment words, the other key problem is how to estimate the score of the sentiment word. Since the dictionary does not define the strength of the words, we need to compute the sentiment score of the words. Ku et, al. use the sentiment scores of the composing characters to compute the score of a sentiment word, and they use the dictionary to estimate the score of the sentiment score of the character. In our system, we use $OpinSF(w_j)/SF(w_j)$ as the score of the sentiment word W_j. $OpinSF(w_j)$ is number of opinion sentences which contain word Wj; $SF(w_j)$ is the sentence frequency of word Wj.

We propose an iterative algorithm to estimate the opinion score of the words and the sentences. First, a rule-based algorithm is used to obtain an initial opinionated sentence set, and then the score of the sentiment word can be computed through this opinioned sentence set. The new opinion scores of the word are used to compute the opinion scores of the sentence. This process can be continued until convergence.

3. Detecting Relevant Sentences

In Relevance subtask of MOAT, we judge if one sentence in a document is relevant to a given topic. We find that this task is very similar to the traditional information retrieval task. But there is another problem. Some sentences that don't contain any key words from the topics, but they indeed are relevant to the topics. We use the pseudo feedback and query expansion method to refine the topics to improve the recall rate.

We use the topic title as a query to start the first stage retrieval, and then we can get a relatively small relevant result set. Although the result is of high precision, the recall rate is low. So we use pseudo feedback to expand the query, and restart the process of retrieval.

4. Experimental Results

(1) Table 1 shows Opinion Analysis Results of Simplified Chinese. Seen from the table, our system received average scores, and the precision and recall value are not quite high. The possible reason is our method may bring noise when using the rule-based algorithm. The other possible reason is that there are many parameters to be tuned, but we failed to tune it to be optimal.

(2) Table 2 shows the evaluation results for relevance subtask. Seen from the table, our system perform very well. By using pseudo feedback and query expansion, we got a relatively higher recall rate.

Table 1. Opinion Analysis Results of Simplified Chinese

L/S	P	R	F
Strict	0.4271	0.8118	0.5597
Lenient	0.4649	0.7442	0.5723

Table 2. The evaluation results for relevance subtask

L/S	P	R	F
Strict	0.9828	0.9396	0.9593
Lenient	0.9703	0.9288	0.9491

5. Conclusions and Future Work

This paper describes our system in detail for NTCIR 7 MOAT track. In the subtask of detecting opinionated sentences, our system just receive average scores. The precision and recall scores are not quite high. After speculation, we find that our method has many limitations: 1). When using the EM algorithm to extract sentiment words, the effect is influenced by the initial seed sentiment words, so maybe some preprocesses are needed. 2). In the iterative algorithm, there are also many problems. The initial rule may bring noise. 3). When estimating the score of sentiment word, it is sensitive to the scale of the collection, and just using the collection to estimate the score may be not very suitable, so it is necessary to combined it with other method. For example, if we have a dictionary with strength score, we can incorporate it as a prior value.

In future work, we will reinforce our method through overcoming the above limitation, and we will test sensitivity of the performance to the introduced parameters, and then develop some methods to automatically tune the parameters.