

Natural Language Processing Term Project Report

Sisi Li & Yuqian Zhang

In this project, we performed sentiment analysis on the reviews of college professors by using the Naïve Bayes classifier and the Maximum entropy model.

All review data were obtained from the Rate My Professor(<http://www.ratemyprofessors.com>), which is a platform for college students across the country to evaluate their professors. Our data is basically focused on the reviews of Computer Science professors at NYU.

We classified our data based on the score each review provides, and the range is from 1 to 5. Professors scored at 3 and above were categorized as “Positive”, and below 3 were categorized as “Negative”. There were about 400 reviews in total, and we used about 250 to train the model, and the rest as test corpus.

Naïve Bayes classifier used “smoothing”, and the word to word tagging accuracy is 76%, 5058 out of 6655 tags correct. We also went back and score our review unit tagging, and the accuracy is 86%, 125 out of 145 tags right, which means of all the 145 reviews, our classifier tagged 125 right.

In order to refine the classifier, we examined the review with the wrong tags. We found out that a couple of reviews scored 3 or above consists more negative sentence than positive sentences. For example , the review scored 3 like” Very smart man, but it is often difficult to understand the point he is trying to get across. Homework is really difficult and takes a LONG time. He gave up 3 programming assignments due the day before the midterm. The exams are difficult but not impossible, and he curves generously. If you do all the homework, you will do fine in the class.” However, we did not find a lot review like this, so we decided to stick to recording word frequency method.

Besides Naïve Bayes classifier, we also tried to tag the corpus by using the Maximum entropy model. With different features, combinations of preToken, currentToken, postToken and preTag, we have gained following result.

1. cur, pre, post, preTag
2812 out of 6655 tags correct
Accuracy:42.25
2. cur, pre, preTag
3272 out of 6655 tags correct
Accuracy:49.17
3. cur, post, preTag
2682 out of 6655 tags correct
Accuracy:40.30
4. cur, preCur, curPost, preTag

2292 out of 6655 tags correct

Accuracy:34.44

5. cur, preCurpost, preTag

2752 out of 6655 tags correct

Accuracy:41.35

In our Maximum entropy model, the chunk tagger is different from the one introduced in the class. We don't have the three-part chunk as a reference. Thus, F-measure, together with recall and precision value are not available in our score result. Instead, by taking a look at the accuracy, we found that the highest accuracy is 49.17, produced by having preToken as our main feature. Compared to the Naïve Bayes classifier, the Maximum entropy model is not a good way to analyze the scored reviews.

Naïve Bayes classifier yielded a better result in our sentiment analysis.

Attachment:

Training.txt (Our training corpus with words tagged with Negative and Position)

Words.txt (Our testing corpus with words to be tagged)

Key.txt (The right tagged correct answer to words.txt)

Code for Naïve Bayes Model:

Naïve_Bayes.java (Naïve Bayes classifier)

Score.java (used to score the accuracy of Naïve Bayes)

Code for Maximum entropy model:

ChunkRange.java (MEMM Model)

FeatureBuilder.java (MEMM Model)

FeatureBuilder2.java (MEMM Model)

MEMMmain.java (MEMM Model)

MEtag.java (MEMM Model)

MEtrain.java (MEMM Model)

Score_MEMM.java (used to score the accuracy of MEMM Model)

Response1.txt (MEMM Model tagging result)

Response2.txt (MEMM Model tagging result)

Response3.txt (MEMM Model tagging result)

Response4.txt (MEMM Model tagging result)

Response5.txt (MEMM Model tagging result)