Automatic Identification of Rhetorical Questions

Shohini Bhattasali, Jeremy Cytryn, Elana Feldman, Joonsuk Park
Cornell University

ACL 2015

Overview

Identifying rhetorical questions is relevant for many NLP tasks, including information extraction and text summarization. We present a simple n-gram based language model to classify rhetorical questions in the Switchboard Corpus. We find that incorporatating contextual cues achieves the highest performance.

Rhetorical Questions

Rhetorical questions act as emphatic statements and are often lexically and syntactically indistinguishable from information-seeking questions. Contextual cues help disambiguate as seen below.

Question: *After all, who ever lifted a finger to help George?*
(Rhetorical reading: *No one helped George.*)

Question: *Who likes winter? It is always cold and windy and gray and everyone feels miserable all the time.*
(Rhetorical reading: *No one likes winter in Ithaca.*)

Features for Identifying Rhetorical Questions

- Unigrams and bigrams to capture speaker intent
- Linguistic cues namely, strong negative polarity items (NPIs), modal auxillaries, and certain expressions such as *yet* and *after all*
- POS n-grams to capture predictive common grammatical relations
- Contextual cues i.e., phrases uttered prior to the questions and the response to the questions which reflect discourse patterns.

Example conversation:
"... you give them an F on something that does not seem bad to me. ***What are you telling that student?*** *You're telling them that hey, you might as well forget it, you know."*

If we only consider the question in the above example, it could be a regular question. However, if we consider the precedent and subsequent utterances, we know that it is a rhetorical question.

Accordingly, we consider the following feature sets:

- Unigrams
- Bigrams
- POS bigrams
- POS trigrams

Question + Precedent + Subsequent

context

Experiment

We used a Support Vector Machine (SVM) as our principal classifier. All unigrams and bigrams in the training data are considered as potential candidates for features. For each feature set, we estimated the maximal predictivity over both rhetorical and non-rhetorical classes. POS features were computed similarly. In order to assess the relative value of question-based and context-based features, we designed seven feature sets: Question (baseline), Precedent (Pre), Subsequent (Sub), Q+Pre, Q+Sub, Pre+Sub, Q+Pre+Sub.

Result & Analysis

[Table 1] shows the performance of the feature sets cross-validated and trained on 5960 questions in the corpus and tested on 2555 remaining questions. We use F_1-score as our main metric due to the skewed nature of our dataset (~5% of test questions rhetorical).

Feature set	Acc	Pre	Rec	F1	Error 95%
Question	92.41	35.00	60.16	44.25	7.59 ±1.02
Precedent	85.64	12.30	30.47	17.53	14.36 ±1.36
Subsequent	78.98	13.68	60.16	22.29	21.02 ±1.58
Question + Precedent	93.82	41.94	60.94	49.68	6.18 ±0.93
Question + Subsequent	93.27	39.52	**64.84**	49.11	6.73 ±0.97
Precedent + Subsequent	84.93	19.62	**64.84**	30.14	15.07 ±1.38
Question + Precedent + Subsequent	**94.87**	**49.03**	59.38	**53.71**	**5.13± 0.86**

Table 1. Experimental Results (%) Accuracy, precision, recall, f1-score, and training error within 95% confidence interval

Our results largely reflect our intuition on the expected utility of our various feature sets. While features in the question group prove the most useful single source, the subsequent features outperform precedent features.

Our results suggest that incorporating contextual cues from both directly before and after the question itself outperforms classifiers trained on a naive question-only feature space.

Conclusion

Overall, adding in n-gram features from the context improves the performance. We achieve a 53.71% F1-score by adding features extracted from the preceding and the subsequent utterances, which is about a 10% improvement from a baseline classifier using only the features from the question itself.