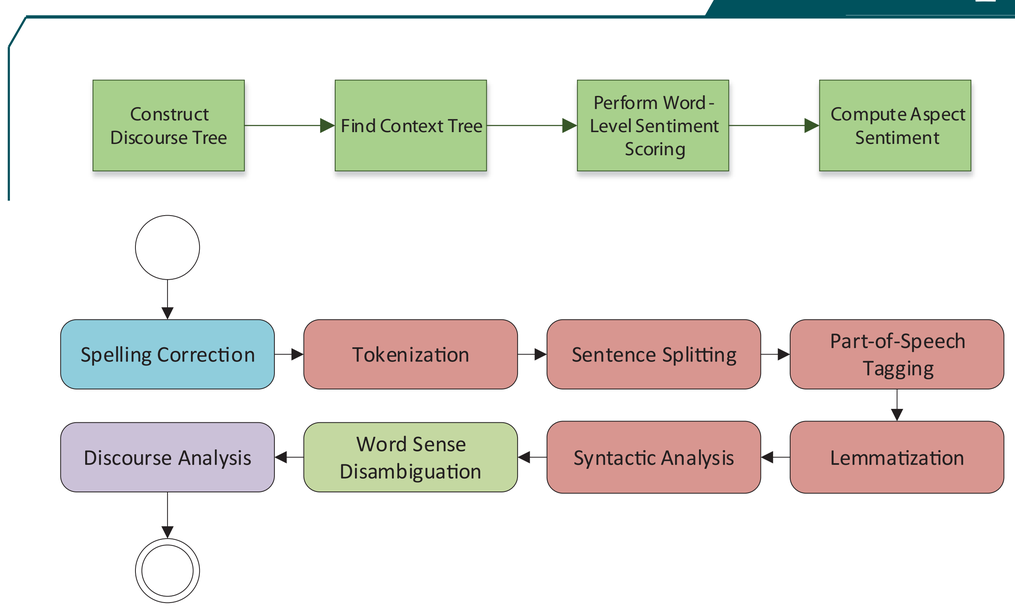
COMMIT at SemEval-2016 Task 5:
Sentiment Analysis using Rhetorical Structure Theory

Kim Schouten & Flavius Frasincar

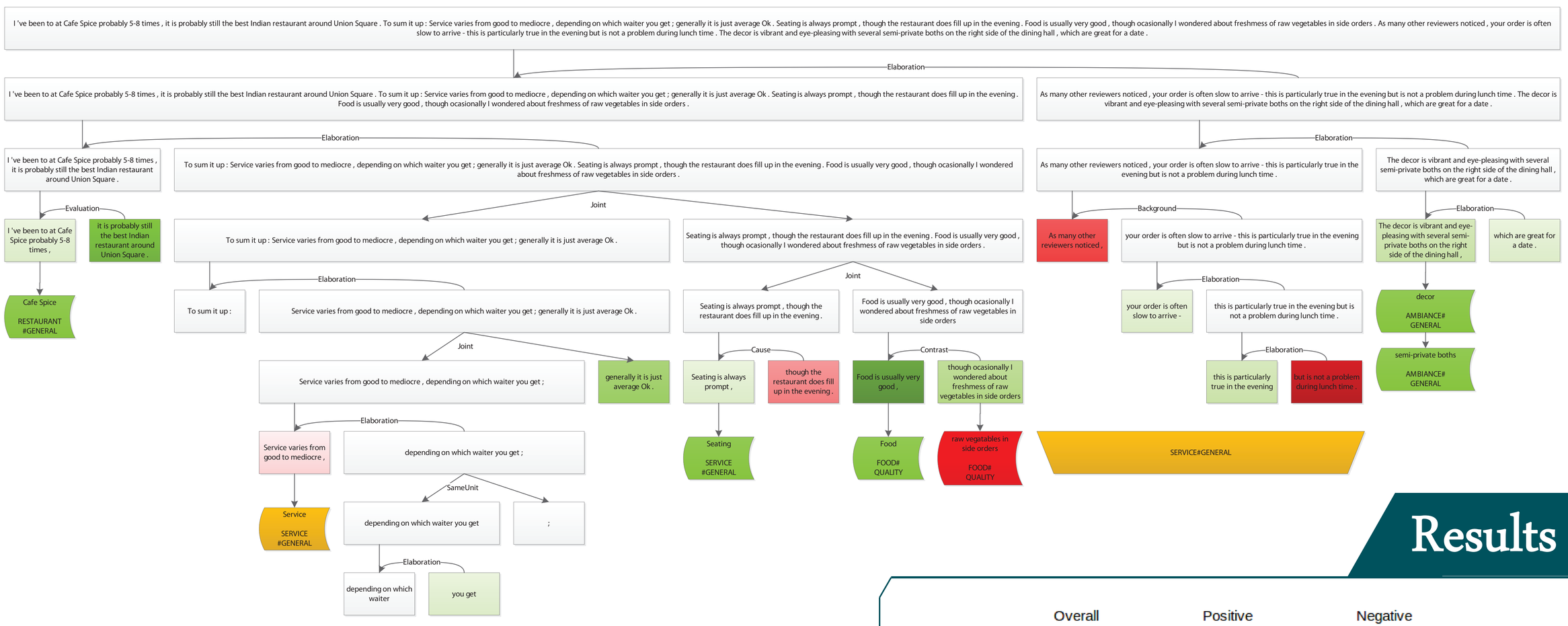
Abstract

This paper reports our submission to the Aspect-Based Sentiment Analysis task of SemEval 2016. It covers the prediction of sentiment for a given set of aspects (i.e., subtask 1, slot 2) for the English language using discourse analysis. To that end, a discourse parser implementing the Rhetorical Structure Theory is employed and the resulting information is used to determine the context of each aspect, as well as to compute the expressed sentiment in that context by weighing the discourse relations between words.

Setup



Running Example



Computing sentiment

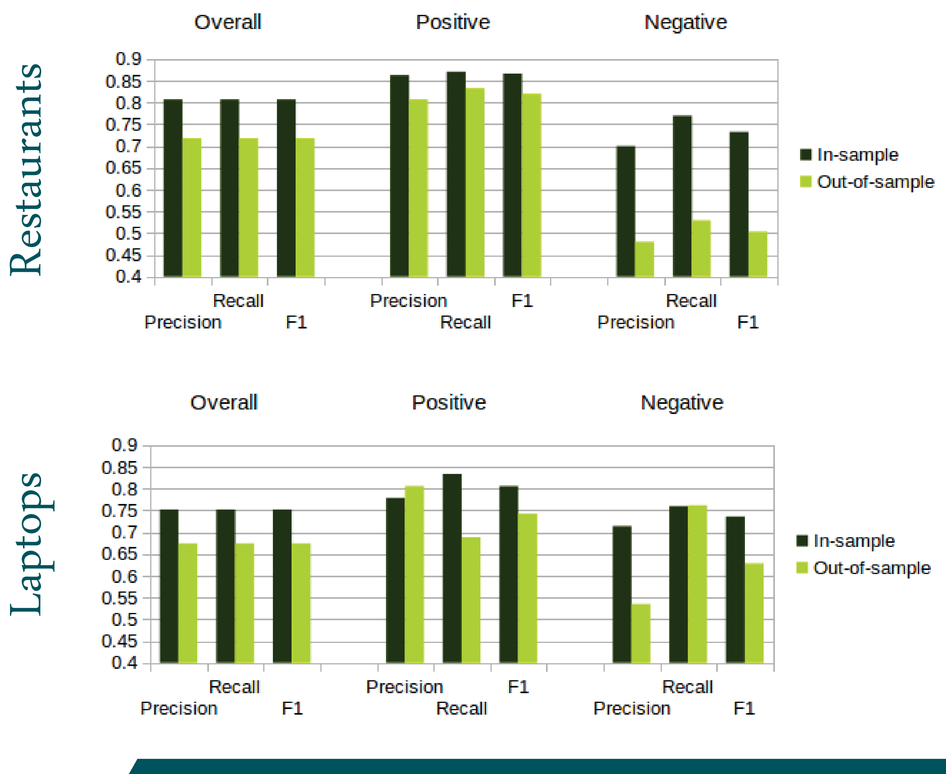
For each leaf in the context tree

$$sent(s_i) = \sum_{t_j \in s_i} sent(t_j) \times \prod_{r_n \in P_{s_i}} w_{r_n}, \forall s_i \in S$$

For each aspect

$$sent(s_a) = \sum_{s_i \in S_a} sent(s_i)$$

Results



Conclusions

While discourse analysis yields high level linguistic information that can be used to better predict sentiment, the proposed algorithm does not yet stack up to the high-performing machine learning approaches that are commonly exploited for this task.

Future Work

1. Include the neutral class
2. Incorporate negations and amplifiers/diminishers
3. Use a better sentiment lexicon or multiple ones
4. Improve context tree determination algorithm
5. Improve sentiment aggregation over context tree
6. Address overfitting issues
7. Combine RST method with a machine learning classifier

ERASMUS UNIVERSITEIT ROTTERDAM

COMMIT/ SIKS