Project Proposal

Team Members: C. Hidey (ch3085)

March 4, 2015

I propose a project to analyze interview descriptions using sentiment and discourse features. I plan to build on existing work into using linguistic connectives that may identify similar or contrasting sentiment. Being able to identify the perceptions of a company by a potential employee may be useful for customer relations and to improve interview experiences.

The website glassdoor.com provides reviews of companies and people can also describe their interview experience. Each interview description also contains 3 types of labels: whether they were offered the job and accepted, whether their experience was positive or negative, and how difficult the interview was (easy/average/hard). I hypothesize that sentiment and discourse will be useful in predicting interview difficulty. I also plan to implement current research in sentiment and discourse to predict the polarity of the interview. Simultaneously, I hope to determine the correlation between the difficulty of an interview and the polarity.

In the timeframe of this project, I plan to create a classifier to predict the labels for interview descriptions for difficulty and polarity. I plan to create a corpus by harvesting data from the website and using a supervised learning framework. During this project, there may arise some corpus-specific obstacles. Crawling the website and creating the corpus is possible, but it is difficult to say what complications may arise. Also, while the data is mostly formatted, it is closer to Twitter than newswire, which may require some normalization. Finally, since these are interview descriptions, there will be a significant amount of terminology that is not used in standard lexicons, and the variety across many different fields is potentially vast.

At the discourse level, there is evidence that sentiment is contrasted with discourse relations. Hatzivassiloglou and McKeown use common discourse markers to predict the orientation of adjectives (1). Zhou focuses on predicting sentiment analysis at the discourse level (4). Trivedi and Eisenstein predict fine-grained sentiment analysis at the level of discourse argument spans (3). Lazaridou et al. have also created a model for unsupervised joint inference for discourse and sentiments using Bayesian networks (2).

I plan to incorporate the results of these projects into this current research. Although this analysis would be at the document level (2-3 paragraph documents), I theorize that discourse features should contribute to the identification of the overall polarity and difficulty expressed in the interview descriptions.

References

[1] Vasileios Hatzivassiloglou and Kathleen R. McKeown. 1997. *Predicting the semantic orientation of adjectives*. In Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics, 174-181.

[2] Angeliki Lazaridou, Ivan Titov, and Caroline Sporleder. 2013. *A Bayesian Model for Joint Unsupervised Induction*. Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pages 1630-1639.

[3] Rakshit Trivedi and Jacob Eisenstein. 2013. *Discourse Connectors for Latent Subjectivity in Sentiment Analysis*. Proceedings of NAACL-HLT 2013, pages 808-813.

[4] Yudong Zhou. 2013. *Fine-grained Sentiment Analysis with Discourse Structure*. Master Thesis, Saarland University.