

## Exploiting Discourse Structure for Sentiment Analysis of Text

As consumers nowadays generate increasingly more Web content describing their experiences with and opinions on, e.g., products and brands, information systems monitoring people's sentiment with respect to such entities are crucial for today's businesses – not in the least as one fifth of all tweets and one third of all blog posts discuss products or brands. The exploration of the potential of such automated sentiment analysis techniques has only just begun.

A typical automated sentiment analysis approach is to use frequencies of positive and negative words in order to determine whether a text is predominantly positive or negative. Such an approach ignores structural aspects of a text, whereas such aspects may contain valuable information. We hypothesize that it may not be so much the sentiment-carrying words per se that convey a text's overall sentiment, but rather the way in which these words are used. Sentiment-carrying words in a conclusion may for instance contribute more to the overall sentiment of a text than sentiment-carrying words in, e.g., background information.

In this light, we propose to guide automated sentiment analysis by a text's discourse structure, as identified by applying Rhetorical Structure Theory on sentence level. We use the identified rhetorical roles to distinguish important text segments from less important ones in terms of their contribution to a text's overall sentiment. We subsequently weight the sentiment conveyed by the identified text segments in accordance with their respective importance when determining a text's overall sentiment. Weights optimized by a genetic algorithm yield significant improvements in sentiment classification performance in comparison to a baseline not taking into account text's discourse structure.