Discourse structure and attitudinal valence of opinion words in sentiment extraction

We present a method for extracting the sentiment contained in texts based on the attitudinal valence of opinion words, combined with the hierarchy of discourse relations proposed by Rhetorical Structure Theory (Mann & Thompson 1988). Following the idea of Polanyi and Zaenen (2006), according to which contextual information influences basic valence of individual lexical items, Taboada et al. (2008) propose a word-based method for extracting sentiment from texts that relies on the most relevant parts of a text. The method predicts that opinion words found in the nuclei (more important parts) of a document are more significant for the overall sentiment, whereas opinion words found in the satellites (less important parts) only potentially interfere with the overall sentiment. However, as pointed out by Taboada et al. (2008) and Narayanan et al. (2009), for certain discourse relations the calculation of sentiment should involve both parts of the relation. Consider the following example from Taboada et al. (2008), where S and N mark the beginning of the satellite and nucleus, respectively.

(1) [S] If the plot had been more gripping, more intense, [N] this would have worked perfectly.

Disregarding the satellite in the above sentence means that we miss the condition imposed on worked perfectly. In this case, the calculation of sentiment for the Condition relation ought to take into account the satellite as well as the nucleus, as shown in Trnavac and Taboada (2012). Heerschop et al. (2011) hypothesize that there is a possible hierarchy in relations – the satellites of some relations may contribute more to the overall sentiment than others. Based on our analysis of the affective content expressed by automatically extracted discourse relations from the Simon Fraser University Corpus (Taboada 2008) and the Penn Discourse Treebank (Prasad et al. 2008), we propose to classify all the discourse relations into 4 categories: (1) relations that reverse polarity, (2) intensify polarity, (3) downtown polarity, (4) produce no change in polarity.

While adopting the idea of Polanyi and Zaenen (2006) that discourse relations are valence shifters which change the base value of opinion words, we propose to include discourse relations as textual modifiers that have a fixed percentage scale associated with them and are presented in a similar fashion as intensifiers in the sentiment analysis systems (e.g., Taboada et al. 2011). We make a list of explicit discourse markers that in combination with opinion words are required for creation of the polarity classifier. We compare the performance of a sentiment analysis system (SO-CAL, Taboada et al. 2011) when opinion words are detected only in the nuclei with its performance when both parts of the relation are analyzed in combination with the opinion words. The preliminary results of the analysis show that opinion words important for the sentiment analysis are part of the nucleus only when a discourse relation reverses polarity, while for all other types of relations, both the nucleus and the satellite are relevant for the extraction of the sentiment.

Words: 500
References


