Abstract

In this thesis the author seeks to establish the most appropriate mechanism for conducting sentiment analysis with respect to political debates; firstly so as to predict their outcome and secondly to support a mechanism to provide for the visualisation of such debates in the context of further analysis. To this end two alternative approaches are considered, a classification based approach and a lexicon based approach. In the context of the second approach both generic and domain specific sentiment lexicons are considered. Two techniques to generating domain-specific sentiment lexicons are also proposed: (i) direct generation and (ii) adaptation. The first was founded on the idea of generating a dedicated lexicon directly from labelled source data. The second approach was founded on the idea of using an existing general purpose lexicon and adapting this so that it becomes a specialised lexicon with respect to some domain. The operation of both the generic and domain specific sentiment lexicons are compared with the classification based approach. The comparison between the potential sentiment mining approaches was conducted by predicting the attitude of individual debaters (speakers) in political debates (using a corpus of labelled political speeches extracted from political debate transcripts taken from the proceedings of the UK House of Commons). The reported comparison indicates that the attitude of speakers can be effectively predicted using sentiment mining.

The author then goes on to propose a framework, the Debate Graph Extraction (DGE) framework, for extracting debate graphs from transcripts of political debates. The idea is to represent the structure of a debate as a graph with speakers as nodes and "exchanges" as links. Links between nodes were established according to the exchanges between the speeches. Nodes were labelled according to the "attitude" (sentiment) of the speakers, "positive" or "negative", using one of the three proposed sentiment mining approaches. The attitude of the speakers was then used to label the graph links as being either "supporting" or "opposing". If both speakers had the same attitude (both "positive" or both "negative") the link was labelled as being "supporting"; otherwise
the link was labelled as being "opposing". The resulting graphs capture the abstract representation of a debate where two opposing factions exchanging arguments on related content.

Finally, the author moves to discuss mechanisms whereby debate graphs can be structurally analysed using network mathematics and community detection techniques. To this end the debate graphs were conceptualised as networks in order to conduct appropriate network analysis. The significance was that the network mathematics and community detection processes can draw conclusions about the general properties of debates in parliamentary practice through the exploration of the embedded patterns of connectivity and reactivity between the exchanging nodes (speakers).

Keywords: Sentiment Analysis, Machine Learning, Debate Visualisation, Debate Analysis & Information Retrieval.