

From Text to Data: Sentiment Analysis of Presidential Election Political Forums

Authors : Sergio V Davalos, Alison L. Watkins

Abstract : User generated content (UGC) such as website post has data associated with it: time of the post, gender, location, type of device, and number of words. The text entered in user generated content (UGC) can provide a valuable dimension for analysis. In this research, each user post is treated as a collection of terms (words). In addition to the number of words per post, the frequency of each term is determined by post and by the sum of occurrences in all posts. This research focuses on one specific aspect of UGC: sentiment. Sentiment analysis (SA) was applied to the content (user posts) of two sets of political forums related to the US presidential elections for 2012 and 2016. Sentiment analysis results in deriving data from the text. This enables the subsequent application of data analytic methods. The SASA (SAIL/SAI Sentiment Analyzer) model was used for sentiment analysis. The application of SASA resulted with a sentiment score for each post. Based on the sentiment scores for the posts there are significant differences between the content and sentiment of the two sets for the 2012 and 2016 presidential election forums. In the 2012 forums, 38% of the forums started with positive sentiment and 16% with negative sentiment. In the 2016 forums, 29% started with positive sentiment and 15% with negative sentiment. There also were changes in sentiment over time. For both elections as the election got closer, the cumulative sentiment score became negative. The candidate who won each election was in the more posts than the losing candidates. In the case of Trump, there were more negative posts than Clinton's highest number of posts which were positive. KNIME topic modeling was used to derive topics from the posts. There were also changes in topics and keyword emphasis over time. Initially, the political parties were the most referenced and as the election got closer the emphasis changed to the candidates. The performance of the SASA method proved to predict sentiment better than four other methods in Sentibench. The research resulted in deriving sentiment data from text. In combination with other data, the sentiment data provided insight and discovery about user sentiment in the US presidential elections for 2012 and 2016.

Keywords : sentiment analysis, text mining, user generated content, US presidential elections

Conference Title : ICBDM 2018 : International Conference on Big Data Management

Conference Location : Rome, Italy

Conference Dates : May 03-04, 2018