Analyzing Text Corpora to Determine the Emotional State of Humans

Wodarz, Paulina (School: University High School)

People’s feelings are often reflected in the way they write. For this project, texts were used to characterize the emotional state of people both throughout the decades and in different parts of society in the present time. The emotional characteristics of words can be quantified by valence (the degree of positivity/negativity) and energy (the degree of passion). For temporal analysis, words (with known valence/energy values) were searched in an online Corpus of Historical American English to determine how their frequency of usage changed over time. For the present-day analysis, a collection of bloggers’ posts from Kaggle (grouped by gender, age, and occupation) was put through a sentiment analysis tool. It was found that over the course of 200 years, energetic words decreased in frequency, and less energetic words increased. Negative and positive words decreased, and neutral words increased, indicating a rise of apathy in society. Further, it was found that in present day’s common usage, females, younger people, and those with a background in the arts exhibit more negative emotions than the other groups. All these results were statistically significant (T-test). These findings indicate that linguistic analysis can be used to detect not only long-term societal trends, but also to study the emotional characteristics of different groups of people. In particular, this method could be a valid tool to identify vulnerable populations that can be targeted for depression evaluation.

Awards Won: