Aniket Pant

Abstract and Methods

**Lexicon-based Sentiment Analysis of Dante’s Divine Comedy**

In this work, we investigate the use of modern sentiment analysis algorithms in the analysis of Dante’s Divine Comedy. We utilize lexicon-trained sentiment analysis per canto through the three sections of Dante’s Inferno, Purgatorio, and Paradiso. For the investigation, we utilize the Python 3.6 scripting language, the Pandas module for data processing, and the TextBlob module for text processing and sentiment polarity calculations. The TextBlob module utilizes a lexicon-based sentiment mapping, meaning it is trained on a specific, internal dataset that has been previously labeled with sentiment and generalized for cross-domain use to account for dependencies in context and sense of usage. In this case, the TextBlob sentiment model is trained on the WordNet3 lexical database (CITATION). For analysis, we utilize the Project Gutenberg rendition of Dante’s Divine Comedy as translated by Rev. H. F. Cary (CITATION). Given the construction of the sentiment analysis algorithm, the clear dependence of the training lexicon should be noted, giving rise to a limitation in this study: the translation of meaning from WordNet3 lexicon to our own text-specific, Dante based lexicon. Since we produce our own domain-specific lexicon, there is some loss in meaning, and hence, some inaccuracy in calculating sentiment. This limitation is additionally compounded by issues and stylistic choices in translation from the text’s original language, Italian. In this work, we present key findings in a study of average sentiment across all Canto’s and all sections of Dante’s Divine Comedy, as presented in the Project Gutenberg edition (CITATION). We discuss average sentiment along with average rates of changes and points of inflections in the story’s chronology.