Edward Wawrzynek

*A Novel Approach to Authorship Attribution Using Word Vectors and Stylistic Features*

Writers have distinctly different styles of writing that stay consistent across all the texts they write. Authorship attribution attempts to utilize these differences to guess the author of a document by calculating features of text and comparing them across different authors. This particular approach to authorship attribution used features of word frequency, punctuation, word and sentence length, word diversity, as well as a novel approach involving the frequency of certain types of words, as determined by groups of semantic word vectors, and tried weighting each feature in a number of different ways to increase accuracy. This method of attribution achieved accuracy of 100% on fourteen novels from seven different authors, and accuracy between 64.3% and 100% on shorter texts. In addition, there was a relationship between the frequency of grouped word vectors and authors' style, and the frequency of grouped word vectors was second in accuracy only to the frequency of certain words, indicating that the use of word vectors could be a useful tool in authorship attribution.