Abstract

Authorship attribution is one of the important problem, with many applications of practical use in the real-world. Authorship identification determines the likelihood of a piece of writing produced by a particular author by examining the other writings of that author. Every author has a unique style of writing pattern. This paper identifies the unique style of an author(s) using lexical stylometric features including function words using balanced training corpus. The present paper calculates the frequencies of the lexical based stylometric features by balancing training and test corpus on English editorial documents. The present paper compares various machine learning algorithms for the authorship attribution and achieved highest average accuracy 95.58 using Random Forest classifier and 92.59 using Multilayer Perceptron algorithms.

References


Index Terms
Computer Science
Artificial Intelligence
Keywords

Authorship Clustering; Stylometry; Supervised Classification