Abstract

Authorship attribution (AA) can be defined as the task of inferring characteristics of a document's author from the textual characteristics of the document itself. In this paper we evaluated the compression model for AA on Telugu text. We considered six different compressors namely Zip, BZip, GZip, LZW, PPM and PPMd in combination with three different compression distance measures such as Normalized Compressor Distance (NCD), Compression Dissimilarity Measure (CDM) and Conditional Complexity of Compression (CCC). The result shows that the compression models are good alternatives for Authorship attribution instead of classification model with various features.

References

Authorship Attribution based on Data Compression for Telugu Text

206–215.
- Bratko, A., Filipic, B.: Spam Filtering Using Compression Models. Department of
Authorship Attribution based on Data Compression for Telugu Text

Intelligent Systems, Jozef Stefan Institute, Ljubljana, Slovenia, IJS-DP-9227 (2005)

Index Terms

Computer Science Security

Keywords

Authorship attribution Compressors Compression distance measures

Macro-average

Micro-average

Accuracy
Telugu data set.