Abstract

In Natural Language Processing (NLP), keywords play a vital role in information retrieval, content summarization, and search engine optimization. However, manual extraction becomes impractical with the increasing volume of information. Automatic keyword extraction becomes essential for scalability. This study addresses the limitations of methods relying solely on global semantic features or local statistical features. It proposes a hybrid approach, combining local statistical features with embedding models, enhancing keyword extraction performance. The method, evaluated on the SemEval2017 dataset using SciBERT and SVM, outperforms baseline methods with an F-score of 0.70. This research emphasizes the importance of leveraging both local statistical information and contextualized semantic features for effective keyword identification in scientific publications.