

[< BACK](#)[Home](#) **Big Data** > **VOL. 11, NO. 1** | Original Articles normal

Siamese-Based Architecture for Cross-Lingual Plagiarism Detection in English–Hindi Language Pairs

Basant Agarwal, Mukesh Kumar Gupta, Harish Sharma, and Ramesh Chandra Poonia

Published Online: 13 Feb 2023 | Doi: <https://doi.org/10.1089/big.2020.0243>

Abstract

The cross-lingual plagiarism detection (CLPD) is a challenging problem in natural language processing. Cross-lingual plagiarism is when a text is translated from any other language and used as it is without proper acknowledgment. Most of the existing methods provide good results for monolingual plagiarism detection, whereas the performances of existing methods for the CLPD are very limited. The reason for this is that it is difficult to represent the text from two different languages in a common semantic space. In this article, a novel Siamese architecture-based model is proposed to detect the cross-lingual plagiarism in English–Hindi language pairs. The proposed model combines the convolutional neural network (CNN) and bidirectional long short-term memory (Bi-LSTM) network to learn the semantic similarity among the cross-lingual sentences for the English–Hindi language pairs. In the proposed model, the CNN model learns the local context of words, whereas the Bi-LSTM model learns the global context of sentences in forward and backward directions. The performances of the proposed models are evaluated on the benchmark data set, that is, Microsoft paraphrase corpus, which is converted in the English–Hindi language pairs. The proposed model outperforms other models giving 67%, 72%, and 67% weighted average precision, recall, and F1-measure scores. The experimental results show the effectiveness of the proposed models over the baseline models because the proposed model is very efficient in representing the cross-lingual text very efficiently.

We use cookies to give you a better experience on liebertpub.com. By continuing to use our site, you are agreeing to the use of cookies as set in our Cookie Policy.

OK