

Abstract

Arabic is an attractive and popular language in the world. However, natural language processing for Arabic has not yet received enough attention from researchers. One of the important aspects of this process is the diacritic recognition or detection of words and letters in Arabic. In Arabic, changing the diacritics of a letter can change the meaning of a word in general. This is especially a big problem for non-Arabic speakers. In this research, a network based on Bi-LSTM deep learning, and CRF with two-level inputs is presented. The detection results of this algorithm are then improved by making some post-processing and corrections. The Implementation results show that the algorithm proposed in this research can perform up to twice as much as the best methods in this field and detect the diacritics in the Arabic phrase.

Keywords: Natural language processing, Deep learning, Arabic, diacritic, Bi-LSTM, CRF