
Abstract

A spell checker is a tool for detecting and correcting various spelling errors. While it might be trivial for humans, spell detecting and correcting can be very useful for machines, because machines could not detect spelling errors and correct them automatically. In Natural Language Processing (NLP), detecting and correcting spelling errors is a task that has been widely performed to normalize data, since most raw texts are noisy and have many spelling errors. In recent years, Long Short-term Memory (LSTM) has shown to give an extraordinary result in solving sequential problems, including spelling correction. In this paper, we propose an LSTM model that encodes input word at character level, that also uses word and POS tag contexts as features. We performed the experiment on an artificial dataset based on Indonesian Wikipedia articles that we made by simulating some artificial spelling errors at character level and tested it on real dataset, mostly are Indonesian online news articles. The evaluation on test dataset gives 83.76% accuracy.

Keywords: Indonesian spell checker, LSTM, deep learning
