Arabic Light Word Analyser: Roles with Deep Learning Approach

Authors : Mohammed Abu Shquier

Abstract : This paper introduces a word segmentation method using the novel BP-LSTM-CRF architecture for processing semantic output training. The objective of web morphological analysis tools is to link a formal morpho-syntactic description to a lemma, along with morpho-syntactic information, a vocalized form, a vocalized analysis with morpho-syntactic information, and a list of paradigms. A key objective is to continuously enhance the proposed system through an inductive learning approach that considers semantic influences. The system is currently under construction and development based on data-driven learning. To evaluate the tool, an experiment on homograph analysis was conducted. The tool also encompasses the assumption of deep binary segmentation hypotheses, the arbitrary choice of trigram or n-gram continuation probabilities, language limitations, and morphology for both Modern Standard Arabic (MSA) and Dialectal Arabic (DA), which provide justification for updating this system. Most Arabic word analysis systems are based on the phonotactic morpho-syntactic analysis of a word transmitted using lexical rules, which are mainly used in MENA language technology tools, without taking into account contextual or semantic morphological implications. Therefore, it is necessary to have an automatic analysis tool taking into account the word sense and not only the morpho-syntactic category. Moreover, they are also based on statistical/stochastic models. These stochastic models, such as HMMs, have shown their effectiveness in different NLP applications: part-of-speech tagging, machine translation, speech recognition, etc. As an extension, we focus on language modeling using Recurrent Neural Network (RNN); given that morphological analysis coverage was very low in dialectal Arabic, it is significantly important to investigate deeply how the dialect data influence the accuracy of these approaches by developing dialectal morphological processing tools to show that dialectal variability can support to improve analysis.

Keywords : NLP, DL, ML, analyser, MSA, RNN, CNN