Abstract

This paper presents a Convolutional Neural Network CNN Models to classify Arabic sentences into three topics. These sentences are derived from Essex Arabic Summaries Corpus (EASC) corpus, tokenized to words and transformed to sequences of word indices. All sequences are padded to be in the same length. The models of Convolution Neural Network are built on top of word embedding layer. The word embedding layer is either pre-trained or jointed into the model. Dropout and l2 weight regularization are used to overcome the overfitting during training. The CNN models achieve high performance in accuracy for Arabic sentences classification.

References


7. Francois chollet, “Deep learning with python”, Manning publication Co, 2018


Index Terms

Computer Science

Artificial Intelligence

Keywords

Classification, Convolutional neural network, Word Embedding