

Abstract

Named entity recognition (NER) is one of the major issues in natural language processing. Named entities consist of specific nouns like names of person, organization, location etc. which refer to essential perceptions in a text. NER is contributive in better performance of applications like machine translation, text summarization and text classification. In the recent decade, with respect to development of deep learning (DL) methods, considerable advances are recorded in this field. The objective here is to propose an efficient method for NER in Farsi (Persian) text through DL methods. Because the deep neural networks require more data for better learning and due to lack of tagged Farsi data, attempt is made to apply transfer learning and active learning approaches. BERT pre-trained model is applied here, which is based on transfer learning to take advantage of transferring knowledge from source task to destination task. BERT is capable of supporting more than 100 languages including Farsi. The architecture of our proposed method is based on BERT and conditional random field (CRF). The results of applying supervised learning method on Arman corpus is 84.23% and 80.80% word-level and phrase-level F1-score, respectively. Our proposed method on PEYMA corpus has 86.14% and 82.05% word-level and phrase-level F1-score, respectively. By applying active learning methods with 30% of the Arman corpus and 40% of the PEYMA corpus separately, a 90% efficiency is obtained. As to the objective here, it is revealed that by applying less labeled data, the main challenge here, a close efficiency to the supervised learning method is obtained.

Keywords: *Natural Language Processing, Named Entity Recognition, Transfer Learning, BERT, Active Learning*