ABSTRACT

Template extraction is a form of Information Extraction (IE) which aims to obtain patterns from data. A few studies considered that template extraction is beneficial in some tasks such as question template extraction. However, utilization of sequence labeling approach in questions template extraction as a task was less explored in prior studies. Several studies suggested that sequence labeling is beneficial in IE and could be implemented in template extraction. However, in those studies, detailed explanation about utilization of sequence labeling approach in questions template extraction task is not justified yet and further exploration and analysis about question template extraction has not been carried out yet. Based on those limitations, this study proposed a novel question template extraction model using sequence labeling approach by utilizing BiLSTM algorithm. According to the experimental results and followed by comprehensive analysis that carried out in this study, it can be concluded that the proposed model is superior to the state-of-the-art model from prior related study in extracting question template using sequence labeling approach. The model with best performance in this study is a BiLSTM which achieved 0.857 in ROUGE-L score. This model employed 200 LSTM units and used Adam as the optimizer with learning rate of 0.001 and was trained using 100 epochs.

Keywords: NLP, information extraction, BiLSTM, CRF, sequence labeling