A Framework for Chinese Domain-Specific Distant Supervised Named Entity Recognition

Authors : Qin Long, Li Xiaoge

Abstract : The Knowledge Graphs have now become a new form of knowledge representation. However, there is no consensus in regard to a plausible and definition of entities and relationships in the domain-specific knowledge graph. Further, in conjunction with several limitations and deficiencies, various domain-specific entities and relationships recognition approaches are far from perfect. Specifically, named entity recognition in Chinese domain is a critical task for the natural language process applications. However, a bottleneck problem with Chinese named entity recognition in new domains is the lack of annotated data. To address this challenge, a domain distant supervised named entity recognition framework is proposed. The framework is divided into two stages: first, the distant supervised corpus is generated based on the entity linking model of graph attention neural network; secondly, the generated corpus is trained as the input of the distant supervised named entity recognition model to train to obtain named entities. The link model is verified in the ccks2019 entity link corpus, and the F1 value is 2% higher than that of the benchmark method. The re-pre-trained BERT language model is added to the benchmark method, and the results show that it is more suitable for distant supervised named entity recognition tasks. Finally, it is applied in the computer field, and the results show that this framework can obtain domain named entities.