Shuhui Zhu
sz649

Critique 6: “Event Extraction by Answering (Almost) Natural Questions”

Xinya Du and Claire Cardie

“Event Extraction by Answering (Almost) Natural Questions” provides an alternative solution to approach event extraction tasks, incorporating BERT-based QA models. This solution could be generalized to transform other labeling tasks into QA tasks. For example, a similar framework could facilitate NER tagging. We could feed in BERT QA model by formulating questions such as “[CLS]what are the organizations, including companies, institutes, teams and agencies [SEP]+<input sentence>” and “[CLS]the event happened in/at where, the location, including mountains, the regions [SEP]+<input sentence>” to get outputs for ORG and LOC tags accordingly. These are one possible phrasing of template questions, which entails further examination since models’ performance may vary depending on the formulation of questions.

This paper provides a thorough analysis in question formulation. The improvement in result with template 3 and the addition of “in <trigger>”, incorporating more semantic information shows the significance of question formulation in QA tasks. Template 3 is developed from a set of existing annotation guidelines, which first categorizes events into different event types, such as Movement.Transport and provides different question templates according to the event types. We could also generalize this method to other tasks with a question formulation component. One example is to refine the above proposed question template in NER tasks. After a careful scan of the annotation guidelines provided in this paper’s appendix, many event types have output argument roles similar to NER tags such as “Person” (“PER”), “Org” (“ORG”) and “Place” (“LOC”). We could use the same questions to get output for NER tags. In semantic labeling tasks, questions are designed by researchers; however, in many QA tasks, questions are already given by users. In this scenario, to improve the answer quality an additional step to rewrite the query is desired.

As template 3 and adding “in <trigger>” achieve a better performance by encompassing more information than the baseline, further improvement could be made by incorporating the previous questions’ answers. For example, after getting an answer “French company” from “Who is the buying agent in sale?” we could formulate the next question as “What was bought in sale by French company?” This addition is meant to give more information when training BERT model.

The sophistication of BERT models and careful design in question formulation though achieves a better performance than other conventional models, but some problems are still left to be solved in event extraction. Spotted in the paper, one difficulty lies in complex sentence with multiple clauses. Given a sample mixing complex sentences and simple sentences, the ratio of complex sentence may be too little to have BERT models to learn the complex rules. To improve the performance on complex sentences, a different sample containing all complex sentences could better suit the purpose. We could either use the length of sentence as a proxy, or the appearance of initiation words in clauses like what and that. If a sentence is longer than 15 words or contains initiation words of clauses, we keep it in the new sample.

In general, this paper provides great insights on event extraction. The advances could be generalized to other labeling and QA tasks.