PRGC: Potential Relation and Global Correspondence Based Joint Relational Triple Extraction

https://aclanthology.org/2021.acl-long.486.pdf

Abhinav Kumar Thakur (Id: 6180633316) and Hassan Hamad (Id: 9491306979)

Citation to the original paper

Hengyi Zheng, Rui Wen, Xi Chen, Yifan Yang, Yunyan Zhang Ziheng Zhang, Ningyu Zhang, Bin Qin, Ming Xu, Yefeng Zheng. PRGC: Potential Relation and Global Correspondence Based Joint Relational Triple Extraction. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing.

What is the general problem this work is trying to do?

The problem is called: Relation Extraction (RE). We are given some unstructured text, e.g. "*Jeff Bezos* founded *Amazon* from his garage in Bellevue, Washington, on July 5, 1994". The goal is to extract the triplet relation (Jeff Bezos, FOUNDER, Amazon). This usually requires recognizing that *"Jeff Bezos"* is an entity of a PERSON and *"Amazon"* is an entity of an ORGANIZATION. Then we should extract the FOUNDER relation between these two entities. Note that the problem does not have to be solved in such a pipeline manner.

What is the new specific approach being taken in this work, and what is interesting or innovative about it,in your opinion?

In this paper the author proposes a new approach called Potential Relation and Global Correspondence (PRGC). The model first encodes the sentence using the pre-trained model BERT. What's new in this paper is the following 3 stage process: Relation Judgement, Entity Extraction and Subject-object Alignment. This means that when we want to predict the relation in the sentence "*Jeff Bezos* founded *Amazon* from his garage in Bellevue, Washington, on July 5, 1994", the model first tries to identify the potential relations, e.g.: *"Founder"*. Then it extracts the subject-object pairs, e.g.: *"Jeff Bezos"* and *"Amazon"*. After which the Subject-object Alignment stage aligns the subject-object pair into a relation triplet. The proposed model has better performance than the baseline models on sentences with different overlapping entity-relationship patterns which appeared quite interesting to us.

What are the specific hypotheses from the paper that you plan to verify in your reproduction study?

- On the task of relation extraction over the datasets NYT and WebNLG, the PRGC model outperforms the 8 baseline models mentioned in the paper.
- The PRGC model is less complex in terms of FLOPS and number of parameters compared to the given previous models.

What are the additional ablations you plan to do, and why are they interesting?

If time permits and we are not limited by hardware constraints, we plan to do an ablation study on the input word embeddings. Specifically, the authors currently use the BERT pretrained model to encode the input sentence. We plan to test different variants e.g. RoBERTa, static word/character embedding like Glove.

Briefly state how you are assured that you have access to the appropriate data including, where relevant, data splits

The paper considers two datasets that are publicly available: WebNLG and NYT. Also, the data splits are given.

Briefly state whether you will re-use existing code or whether you will implement yourself

The authors of the paper included a github repository for their code but we will try to replicate the results using our own implementation and using standard machine learning libraries in python.

Discuss the computational feasibility of your proposed work– make an argument that the reproduction will be feasible.

The authors used NVIDIA Tesla V100 GPU and trained the model for 100 epochs. The hardware available to us is: NVIDIA GeForce RTX 2080Ti GPU and Intel Xeon Gold 5220 CPU. While our GPU is less powerful, we have access to multiple such GPUs and so we expect to be able to train the model but with more time overhead.