Commonsense Reasoning For Question Answering

Contact: Julien Romero julien.romero@telecom-sudparis.eu

1   Context

Commonsense is a skill every human has but that is hard to get for computers. A simple observation can convince us: When we write a text, we rarely state the obvious, what is commonsense. For example, we will rarely say that at night, the sun is not visible!

We can divide the problem of commonsense into two parts. First, there is commonsense knowledge, i.e. statements that we intuitively know are true. For example, the fact that elephants have a trunk. It is opposed to encyclopedic knowledge that is acquired by studying. For example, we learn that Paris is the capital of France at school. The second part is commonsense reasoning, i.e. using reasoning over commonsense knowledge. This kind of reasoning is particularly useful when it comes to question-answering. For example, to the question "where would I not want a fox?", I could answer a hen house as foxes eat hens and hens are found in hen houses.

The goal of this project is to study the limitations of the current approaches. In particular, we will be interested in the CommonsenseQA dataset [1]. State-of-the-art algorithms rely on a knowledge base called ConceptNet [2]. This is a problem for several reasons:

1. CommonsenseQA is partially built from ConceptNet, which biased the results.

2. It is not clear if the approaches would generalize to other knowledge bases.

3. They rely on a clear path between the question and the answer in the knowledge graph.

In the first part of this internship, we will study the existing methods such as MHGRN [3] or QA-GNN [4]. We will compare them by changing the knowledge base they use to see how they generalize. Then, we will see how we can leverage the weaknesses to propose a new approach.

2   Involved Skills

- English (French can be useful for daily life).

- Good knowledge of Python.

- Experience with machine learning and deep learning, in particular with frameworks like Pytorch.

- Basic knowledge about knowledge bases/ontologies.

3   Practical Information

The SAMOVAR lab from Telecom SudParis potentially finances this 6-months internship. It will take place either on the Evry campus or the Palaiseau campus. The expected starting date is between February 2022 and July 2022.

References

[1] A. Talmor, J. Herzig, N. Lourie, and J. Berant, "Commonsenseqa: A question answering challenge targeting commonsense knowledge," *arXiv preprint arXiv:1811.00937*, 2018.

[2] R. Speer and C. Havasi, "ConceptNet 5: A large semantic network for relational knowledge," in *Theory and Applications of Natural Language Processing*, 2012.

[3] Y. Feng, X. Chen, B. Y. Lin, P. Wang, J. Yan, and X. Ren, "Scalable multi-hop relational reasoning for knowledge-aware question answering," *arXiv preprint arXiv:2005.00646*, 2020.

[4] M. Yasunaga, H. Ren, A. Bosselut, P. Liang, and J. Leskovec, "Qa-gnn: Reasoning with language models and knowledge graphs for question answering," *arXiv preprint arXiv:2104.06378*, 2021.