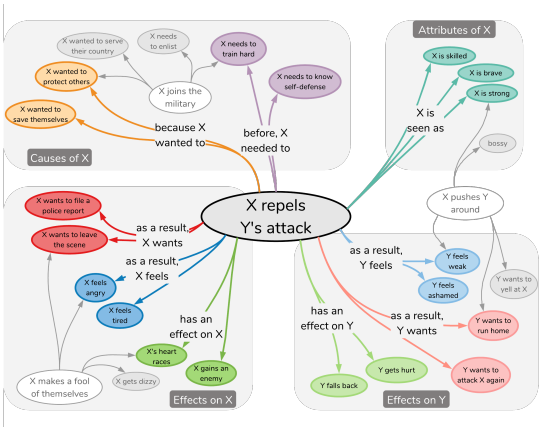
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Distributed Computing

Prof. R. Wattenhofer

Multimodal Pretraining with Commonsense Generation



Researchers have proposed models combining multimodal data, namely images and texts. These models have improved the performance of various downstream tasks. Despite the success, these models still suffer from drawbacks. On the one hand, many of the models are not suitable for generative tasks. On the other hand, these models are not aware of chronological and causal relations, which are important for reasoning tasks. Commonsense knowledge graphs are rich in chronological and causal relations, but none of the pre-trained models have leveraged commonsense knowledge graphs.

In this work, we will investigate a new pre-trained model to overcome the aforementioned drawbacks. Our proposed model will be the first generative neural architecture that incorporates and leverages structured commonsense knowledge. A recently proposed task, visual commonsense generation (VCG) will be deployed as the testing ground, which requires the generative model to generate textual information under the awareness of chronological and causal commonsense.

Specifically, our primary goal is as follows:

- Create a novel multimodal pre-trained model for generative tasks.
- Enhance our model expressiveness with pre-training tasks by leveraging commonsense knowledge graphs.
- Evaluate the proposed model on downstream multimodal tasks, for instance, commonsense reasoning tasks including VCR and VCG.

Requirements: Strong motivation, proficiency in Python & PyTorch, and prior knowledge in Deep Learning.

Interested? Please contact us for more details!

Contact

- Zhao Meng: zhmeng@ethz.ch, ETZ G61.3
- Yunpu Ma: cognitive.yunpu@gmail.com