UBIQUITOUS
KNOWLEDGE
PROCESSING

Master Thesis

TECHNISCHE
UNIVERSITÄT
DARMSTADT

Multi-Modal Commonsense Reasoning

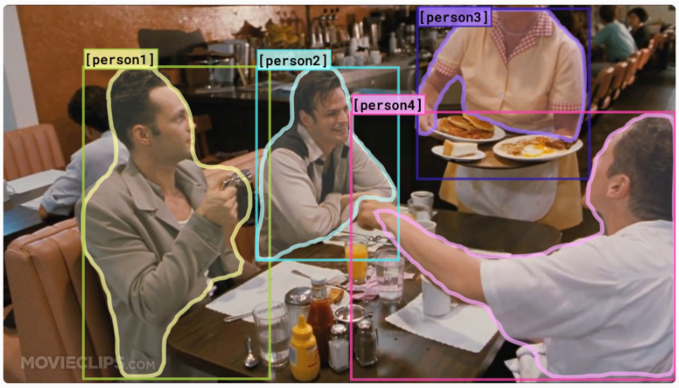
Motivation

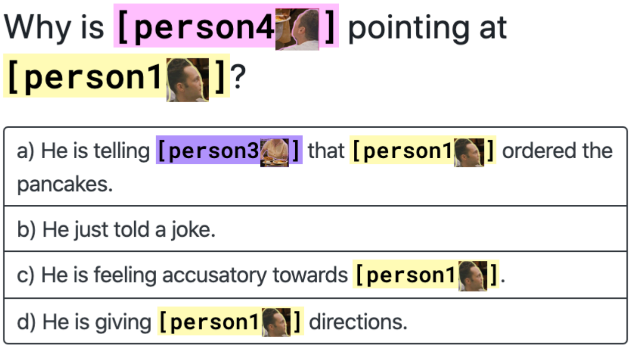
When we as humans reason about our world, we use information from multiple modalities (vision, sound, text, smell, etc.) to reach conclusions. These conclusions are often based on implicit experiences of our past and can be regarded as common sense reasoning. This thesis has the ambition of understanding what commonsense entails and how we can train an AI that is able to perform true reasoning.

Task Description

The focus of this Thesis lies on two interdependent and open research questions:

1) multi-modal representation learning
2) applying these representations to common sense reasoning tasks.





Why is [person4] pointing at [person1]?

a) He is telling [person3] that [person1] ordered the pancakes.

b) He just told a joke.

c) He is feeling accusatory towards [person1].

d) He is giving [person1] directions.

References

- “LXMERT: Learning Cross-Modality Encoder Representations from Transformers”. Hao Tan, Mohit Bansal: EMNLP 2019
- “From Recognition to Cognition: Visual Commonsense Reasoning”. Rowan Zellers, Yonatan Bisk, Ali Farhadi, Yejin Choi: CVPR 2019

Analysis	■■■■□
Programming	■■■■■
Literature	■■■□□

Contact

Prof. Dr. Iryna Gurevych

Jonas Pfeiffer

thesis@ukp.informatik.tu-darmstadt.de