Are Large Language Models better than the Symbolic Reasoners for Question Answering over Theory of Mind Tasks?

Keywords: Large Language Models, Cognitive Sciences, Theory of Mind, Reasoning over Beliefs.

Summary

Theory of Mind (ToM) is the ability to attribute mental states – such as beliefs, intentions, and desires – to oneself and others. It is a crucial part of the development of cognitive processes as well as language understanding, leading to improved reasoning for question-answering tasks. One of the main cognitive processes in ToM is to reason about beliefs.

Example 0.1 (Belief)**.** Alex, a student, is preparing for an important final exam. He believes that if he studies all weekend, he'll perform well (belief). He intends to study diligently (intention) because he desires to achieve a top grade (desire).

During this internship, we are particularly interested in exploiting higher-order beliefs, i.e., beliefs about others' beliefs, or even beliefs about others' beliefs about others' beliefs, and so on, as well as false beliefs. Higher-order beliefs are crucial in social interactions, strategic games, ToM contexts, as they help in predicting and interpreting the intentions and actions of others based on their beliefs and perceptions. Qiana currently only provides the formalism for representation along with the reasoning capabilities without providing concrete grounds for its comparability to LLMs. Additionally, Qiana has not yet been proven to cope with higher-order beliefs.

Example 0.2 (Higher-Order Belief)**.** Alex is working with his friend Jamie, who also has an important exam. Alex believes that Jamie might not study hard, as he seems unmotivated. However, Jamie believes Alex thinks he's not serious about his studies, which affects Jamie's own motivation. Jamie then decides to study to prove Alex wrong, even though Alex might not actually care.

Recent studies suggest that Large Language Models (LLMs) exhibit some ToM abilities [1], such as reasoning about others' beliefs and intentions through tasks like the *false belief task* and *persuasion tasks*. These tasks explore various ToM components, including emotion, desire, intention, knowledge, belief, and non-literal communication. However, ToM Bench mainly introduces second-order beliefs in the benchmark dataset and exploits the capabilities of LLMs for answering multiple-choice questions by reasoning over those beliefs. On the other hand, Qiana [2] is a recently introduced symbolic framework called Qiana [2], which introduces a logical framework for representing and reasoning over beliefs. Qiana builds on top of First-Order Logic (FOL) and adds another level of expressivity.

The results of this study will set a milestone for future work on injecting higher-order beliefs into LLMs using Qiana representations and the other way around.

Contact Details

Mehwish Alam, Telecom Paris, Institut Polytechnique de Paris (mehwish.alam@telecom-paris.fr).

Pierre-Henri Paris, Universite Paris Saclay (pierre-henri.paris@telecom-paris.fr).

Fabian M. Suchanek, Telecom Paris, Institut Polytechnique de Paris (fabian.suchanek@telecom-paris.fr).

References

[1] Zhuang Chen, Jincenzi Wu, Jinfeng Zhou, Bosi Wen, Guanqun Bi, Gongyao Jiang, Yaru Cao, Mengting Hu, Yunghwei Lai, Zexuan Xiong, and Minlie Huang. Tombench: Benchmarking theory of mind in large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics, ACL*. Association for Computational Linguistics, 2024.

[2] Simon Coumes, Pierre-Henri Paris, François Schwarzentruber, and Fabian M. Suchanek. Qiana: A First-Order Formalism to Quantify over Contexts and Formulas. In *Proceedings of the Twenty-First International Conference on Principles of Knowledge Representation and Reasoning*, pages 295–305, Hanoi, Vietnam, November 2024. International Joint Conferences on Artificial Intelligence Organization.