Research Proposal of Visual Spatial Reasoning of Large Language Models

Jinlong Liu
Department of Computer Science, University of Liverpool
J.Liu157@liverpool.ac.uk

***Abstract*—Large Language Models has become a hot topic in the field of Natural Language Processing. The Visual Spatial Reasoning is a key ability of human intelligence. In this paper, we will review the recent research on the Visual Spatial Reasoning of Large Language Models.**

I. Reasearch Topic

In recent years, Large Language Models (LLMs) have achieved great success in the field of Natural Language Processing (NLP). The LLMs are trained on large-scale text corpus, and can be used to solve many NLP tasks, such as machine translation, question answering, and text generation. The LLMs are usually trained on text corpus, and can be used to solve many NLP tasks, such as machine translation, question answering, and text generation.

However, the LLMs are not good at solving the Spatial Reasoning tasks. In generally, Spatial knowledge have these aspects: including (i) mereotopology, (ii) direction and orientation, (iii) size, (iv) distance and (v) shape [1]. And Spatial Reasoning is reasoning more than one of these aspects of Spatial knowledge at same time. Cohn *et al.* [2] build a Spatial Reasoning dataset, which contains 1000 questions. The questions are about the Spatial knowledge of the objects in the images. The LLMs can not solve the Spatial Reasoning tasks well. Borji *et al.* [3] proposed a comphensive analysis of ChatGPT's failures and Spatial Reasoning is one of the failures. Furthermore, Lin *et al.* [4] tested the Spatial Reasoning ability of GPT-4 and found that GPT-4 can not solve the Spatial Reasoning tasks well, but it can consider as a knowledge acquisition tool.

Following the Research of testing Spatial Reasoning of LLMs, I noticed that there a few researches focus on the Visual Spatial Reasoning of Visual Language Models, such as ViLBERT [**?**], LXMERT [**?**], and UNITER [**?**]. The Visual Spatial Reasoning is a key ability of human intelligence. In this paper, we will review the recent research on the Visual Spatial Reasoning of Large Language Models.

References

[1] A. G. Cohn and J. Renz, "Qualitative spatial representation and reasoning," *Foundations of Artificial Intelligence*, vol. 3, pp. 551–596, 2008.

[2] A. G. Cohn and J. Hernandez-Orallo, "Dialectical language model evaluation: An initial appraisal of the commonsense spatial reasoning abilities of llms," *arXiv preprint arXiv:2304.11164*, 2023.

[3] A. Borji, "A categorical archive of chatgpt failures," *arXiv preprint arXiv:2302.03494*, 2023.

[4] F. Lin, Z. Shou, and C. Chen, "Using language models for knowledge acquisition in natural language reasoning problems," *arXiv preprint arXiv:2304.01771*, 2023.