Note on [Learning to Edit: Aligning LLMs with Knowledge Editing]

NoteAuthor: XIONG ZHIPENG

June 11, 2024

1 Which problem is addressed by the paper?

Efficiently modify LLMs' outputs towards targeted queries while preserving overall performance across other unrelated ones

   For example: updating the knowledge of "*The current British Prime Minister is Rishi Sunak*" modifies the response to "*Who is married to the PM of the UK?*"

2 Why is it important?

The dynamic nature of the world necessitates frequent updates to LLMs to rectify outdated information or integrate new knowledge, thereby safeguarding their sustained pertinence.

3 How is it solved?

4 DataAnalysis, Applications, Conclusion and Future Work