

Context-Aware Meme Suggestion Using Large Language Models

Group:A515

Member:陳宥宇

Advising Professor:孫民

Abstract

This study proposes a large language model system that integrates visual and linguistic understanding to automatically recommend the most appropriate meme replies based on user input. The system utilizes the LLaVA-13B model to analyze meme content and generate three representative tags for each image—describing its emotion, theme, and situation. The LLaMA3-8B model then evaluates the semantic relevance between the user's message and the meme information. A staged filtering mechanism is applied to progressively narrow down selections, significantly improving recommendation accuracy. Experimental results show that when provided with well-structured information, the system demonstrates high usability and precision, with strong potential for integration into real-time messaging platforms.

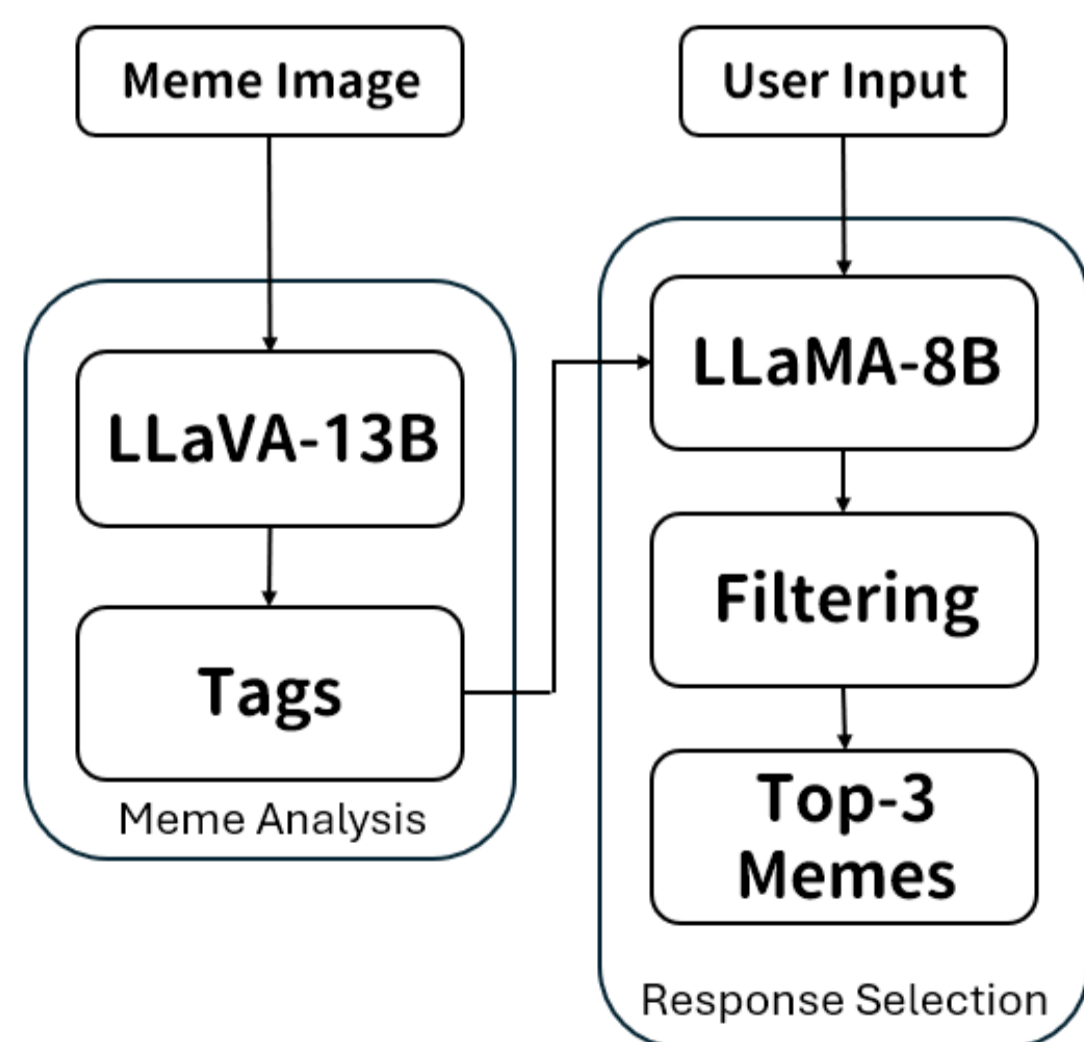
Research Methodology

The system is divided into two main modules:

Meme Analysis Module: The LLaVA-13B model processes image inputs and automatically generates three descriptive tags for each meme. These tags, along with the image filename and caption, are stored in a structured JSON format.

Recommendation Module: The LLaMA3-8B model analyzes the user's input and compares it with the meme data through multiple rounds of selection. A group-based elimination strategy is used to iteratively narrow down the candidates and identify the three most suitable memes for response.

Considering the model's performance constraints and input limitations, the recommendation process is designed as a multi-stage filtering mechanism to reduce semantic noise and improve inference efficiency.



Experimental Results

"I love you!" → recommend : ['011.jpg', '032.jpg', '027.jpg']



我還以為你喜歡的是我



好棒喔



我好愛你

"You are so ugly." → recommend : ['030.jpg', '040.jpg', '007.jpg']



靠近看變得更醜了



那樣，也許是因為你太醜了



你真的很臭

To compare the impact of different meme information formats on recommendation performance, four data versions were designed.

We perform user study to ablate our method. We ask the user whether the meme recommended by different methods are reasonable or not.

Number	Type	Score	Score(%)
Type 1	LLaVA-13B tags	183	91.5%
Type 2	PT-4o tags	181	90.5%
Type 3	LlaVA-13B 敘述	105	50.25%
Type 4	原始檔名與台詞	147	73.5%

The experimental results show that data formats containing tags performed better, and the high scores given by users confirm the feasibility and potential of this research.

Reference

- [1] Liu H, Li C, Wu Q et al. Visual instruction tuning. In: Proceedings of the 37th International Conference on Neural Information Processing Systems . Red Hook, NY: Curran Associates, 2023, 34892–916.
- [2] Introducing Meta Llama 3: The most capable openly available LLM to date. <https://ai.meta.com/blog/meta-llama-3/>
- [3] ChatGPT. <https://chatgpt.com/>
- [4] Ollama. <https://ollama.com/>
- [5] 【情報】有點高畫質的"海綿寶寶梗圖". <https://forum.gamer.com.tw/C.php?bsn=60076&snA=5491441>