Supplementary Materials

Anonymous Authors

1 LARGE LANGUAGE MODEL EXPERIMENTS

For large language models' results in the main experiment, the MNER and MRE results are from Chen and Feng [1], and the results on the MJERE dataset are implemented by us via in-context learning with 10 in-context examples. The sampling temperature is set to 1.0. These large language models are provided with the textual context and the text as input. As shown in Table 1, our prompt consists of an instruction, followed by 10 in-context examples, and finally a question. Each in-context example consists of the textual context of the image, the text and the corresponding entity-relation answer. The question has the same format as in-context examples except there is no answer.

2 CODE IMPLEMENTATIONS

The attachment contains some code implementations of the Shap-CA, including the parsing and processing of part-of-speech and positional embeddings, and the extracted image features in validation split. The complete code will be made public upon acceptance.

REFERENCES

[1] Feng Chen and Yujian Feng. 2023. Chain-of-Thought Prompt Distillation for Multimodal Named Entity and Multimodal Relation Extraction. *arXiv preprint arXiv:2306.14122* (2023).

Table 1: Prompts employed for ChatGPT/GPT4 on the MJERE dataset.

Please perform Named Entity Recognition and Relation Extraction in social media content. Given a tweet and accompanying textual descriptions of its related image, extract entities from the text and identify their associated relations. Entity types include [Entity category], while relation types encompass [Relation category].

Image description: The tweet is about a basketball game. There is a basketball player, holding up a trophy and yelling in celebration.
Text: Stephen Curry and Michael Jordan are both players who have/had 3 NBA Championships by age 30.
Answer:
Stephen Curry (per), Michael Jordan (per), peer.
Stephen Curry (per), NBA Championships (misc), awarded.
Michael Jordan (per), NBA Championships (misc), awarded.

...
<More In-Context Examples>
...

Image description: A black and white photo of a man and a woman. They are standing close to each other, with the man having a tie on. The man is looking at the camera while holding the woman, which gives an air of intimacy and connection between the two.
Text: Elizabeth Taylor and Montgomery Clift in A Place In the Sun, 1951 (dir. George Stevens).
Answer: