

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).

59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116

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:

117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232