我在项目中负责的工作是通过大模型 API 调用来批量处理数据集的问题，并通过提示词工程进一步优化模型在数据集上的表现。

模型：deepseek-v3

使用Python语言编写代码，在代码中使用API 方式调用deepseek-v3模型，

Zero Shot

首先进行无user prompt，也就是进行zero-shot时，使用共同设置相同的系统提示词system prompt：

SYSTEM\_PROMPT = """You are an accurate table Q&A assistant. Please carefully analyze the following table and answer the questions accurately. Note: Only output the answers, do not explain the process.

"table\_text":table content;

"statement": question.

Please give the answer directly.

我（李凯文）、郑宇榕、何锦诚基于测试集的同一个随机子集[test\_100.jsonl](https://github.com/JacksonHe04/smart-table-llm/blob/main/datasets/test_100.jsonl)分别展开了测试。

将该提示词应用到 deepseek-v3 模型上，ACC 是 61%，并将错题保存到incorrect\_questions.jsonl文件中

Simple prompt

在对模型判断错误的案例进行分析后，对提示词部分进行了简单改动：添加了简单user prompt：

Strictly follow these rules:

1. Give the answer directly, without any explanation.

2. Keep the answer concise and accurate.

3. Do not show reasoning process.

再次对[test\_100.jsonl](https://github.com/JacksonHe04/smart-table-llm/blob/main/datasets/test_100.jsonl)进行测试，ACC为64%，提升了3%，说明简单提示词对于模型性能有提升

尝试使用不**同的简单提**示词，如小组其他成员prompt：yurong-prompt

USER\_PROMPT\_TEMPLATE = """

You are a table QA assistant. Your task is to answer questions based on the given table data. Follow these rules:

1. Analyze the table structure (header + rows) and the question.

2. Search for the exact match or logical condition in the table.

3. Return only the answer value, without additional explanations!!!

再次对[test\_100.jsonl](https://github.com/JacksonHe04/smart-table-llm/blob/main/datasets/test_100.jsonl)进行测试，ACC为59%，降低了，分析原因可能为该提示词中对于表格读取的指令影响模型原本理解能力，反而降低了模型性能

Complex prompt

尝试简单的提示词后，使用小组成员的复杂提示词进行测试 jincheng prompt：

USER\_PROMPT\_TEMPLATE = """

You are an accurate table Q&A assistant.

Only output the final answer, without any explanation

The answer must be in exactly the same format as required by the question

If it is a country name, use the full country name instead of the abbreviation

If the quantity is asked, the count must be accurate

If there are multiple possible answers, only output the one that best meets the requirements of the question

The answer must be based on the data in the table and do not use external knowledge

If calculations are involved, they must be accurately calculated without estimation

Keep the case of the answer consistent with what is required by the question

For dates and times, keep the original format

If there are null or missing values in the table, exclude these values during calculations

For the content within quotation marks, the original format must be maintained, including the quotation marks themselves

Carefully check all cases that meet the criteria when counting

For sorting and comparison, consider the data in all relevant columns

If the answer involves specific text content, it must exactly match the original text, including case and punctuation marks

Please carefully analyze the following table and answer the questions accurately. Note: Only output the answers, do not explain the process.

Table Content: ${tableText}

Question：${statement}

Please give the answer directly.

再次对[test\_100.jsonl](https://github.com/JacksonHe04/smart-table-llm/blob/main/datasets/test_100.jsonl)文件进行测试，ACC为60%，相比simple降低，分析原因：可能为过于复杂的提示词对于模型判断造成较大影响，不适用于deepseek-v3模型的小批量处理，后续将再次对比simple prompt进行确认。

随机100\*5测试

已经在100固定数据集（[test\_100.jsonl](https://github.com/JacksonHe04/smart-table-llm/blob/main/datasets/test_100.jsonl)）测试过，将目标范围放大到整体，在总数中随机取100进行测试，共5次取平均，减少误差。

根据kw prompt和jincheng prompt的结果，将两者对比分析，

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| deepseek-v3 | 60.20% | API + 提示词工程 | 5次随机100平均 | zero-shot | 李凯文 |  |
| deepseek-v3 | 65.80% | API + 提示词工程 | 5次随机100平均 | kw-prompt | 李凯文 | 提升5.6% |
| deepseek-v3 | 62.40% | API + 提示词工程 | 5次随机100平均 | jincheng-prompt | 李凯文 | 提升2.2% |

对比得出，添加提示词的比zero-shot准确率更高，simple相比于complex对于deepseek-v3模型ACC的提升更大。可以得知，deepseek-v3适和简单提示词

One-shot

通过阅读老师给的文章，了解可在prompt中添加example以提升ACC，添加ACC后进行测试

SYSTEM\_PROMPT = """You are an accurate table Q&A assistant. Please carefully analyze the following table and answer the questions accurately. Note: Only output the answers, do not explain the process.

"table\_text":table content;

"statement": question.

Please give the answer directly.

Example:

{"statement": "which team won previous to crettyard?", "table\_text": [["team", "county", "wins", "years won"], ["greystones", "wicklow", "1", "2011"], ["ballymore eustace", "kildare", "1", "2010"], ["maynooth", "kildare", "1", "2009"], ["ballyroan abbey", "laois", "1", "2008"], ["fingal ravens", "dublin", "1", "2007"], ["confey", "kildare", "1", "2006"], ["crettyard", "laois", "1", "2005"], ["wolfe tones", "meath", "1", "2004"], ["dundalk gaels", "louth", "1", "2003"]], "answer": ["Wolfe Tones"], "ids": "nt-2"}

Question: which team won previous to crettyard

Answer: Wolfe Tones"""

添加后，随机概率变为64%，由此可见，example对于ACC来说有正面提升。

分析原因：

1.example为模型提供了一个正确读取数据给出回答的模版

2.example规范了模型处理表格的方法

Vision模型（doubao-1-5-vision-pro）

通过组员提示，尝试使用version模型进行测试

try:

        completion = client.chat.completions.create(

            #model="deepseek-r1-250120",

            #model="deepseek-v3-250324",

            model="doubao-1.5-vision-pro-250328",

            messages=[

                {"role": "system", "content": SYSTEM\_PROMPT},

                {"role": "user", "content": get\_table\_qa\_prompt(table\_str, question)}

            ],

            temperature=0.1

        )

        answer = completion.choices[0].message.content.strip()

        return answer.split('\n')[0].strip()

    except Exception as e:

        print(f"API调用失败: {str(e)}")

        return None

使用kw prompt测试，ACC为：69.8%，由此可见，视觉模型在处理表格问答方面更有优势，分析原因：

1. 表格问答的jsonl文件代码本质是体现图表，以代码形式呈现，因此，视觉模型可能在将代码处理成图表并获取、处理信息能力更具优势。
2. 视觉模型对图表类信息进行特化训练，在处理相关信息方面具有较大优势

下图为测试结果汇总

升降与zero-shot对比

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| deepseek-v3 | 61% | API + 提示词工程 | test\_100.jsonl | zero-shot | 李凯文 | 109.17s |  |
| deepseek-v3 | 59% | API + 提示词工程 | test\_100.jsonl | yurong-prompt | 李凯文 | 113.61s | 降6% |
| deepseek-v3 | 64% | API + 提示词工程 | test\_100.jsonl | kw-prompt | 李凯文 | 120.46s | 升3% |
| deepseek-v3 | 60% | API + 提示词工程 | test\_100.jsonl | jincheng-prompt | 李凯文 | 104.14s | 降1% |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| deepseek-v3 | 60.20% | API + 提示词工程 | 5次随机100平均 | zero-shot | 李凯文 |  |
| deepseek-v3 | 65.80% | API + 提示词工程 | 5次随机100平均 | kw-prompt | 李凯文 | 升5.6% |
| deepseek-v3 | 62.40% | API + 提示词工程 | 5次随机100平均 | jincheng-prompt | 李凯文 | 升2.2% |
| deepseek-v3 | 64.00% | API + 提示词工程 | 5次随机100平均 | kw-prompt+example | 李凯文 | 升3.8% |
| doubao-1-5-vision-pro | 69.80% | API + 提示词工程 | 5次随机100平均 | jincheng-prompt | 李凯文 | 升9.6% |