论文第六章中的提示工程-AZM

1. 辅助生成脚本的提示工程

1.1 生成使用Elsevier数据挖掘API下载文献的脚本

You are an expert in coding. Generate a Python script to fulfill the following functions:

1. Search papers with a given query string using Elsevier Text mining API. I have already acquired an API key from Elsevier and installed the needed Python module.

2. Download the full text of 1000 papers in JSON format. If there are any papers that failed to download the full text. An error information should be logged and print on the screen.

Should you have any questions, ask me in your response.

1.2 生成使用GPT-4进行信息挖掘的脚本

You are an expert at coding. Help me generate a Python script. The script needs to contain the following parts:

1. Text extraction. Text should be extract from multiple JSON files. The JSON file has a key named 'full-text-retrieval-response'. This key contains several sub-keys, one of which is 'originalText'. Text in the key ‘originalText’ should be extracted. The result should be a data frame with two columns: ‘file name’ and ‘content’. The ‘file name’ column is the name of the JSON file from which the text is extracted. The ‘content’ column is the text extracted from the JSON file. Note that there are 130 JSON files in the folder, and the text in every JSON file should be extracted. If the text in any JSON file failed to be extracted, output an error message contains the name of the file.

2. Text Segments. The text extracted from Json file should be parsed into several sections based on some keywords. The keywords include: ‘Introduction’, “‘Experimental section’ OR ‘Experimental’ OR ‘Materials and methods’ OR ‘Methods’”, ‘Results and discussion’, ‘Conclusion’. Note that the keywords are sorted by their location that may appear in the text. The output should be a data frame that includes: ‘file name’, ‘section name’ and ‘content’. The ‘file name’ column is the name of the JSON file from which the text is extracted. The ‘section name’ column is the keyword that used to split the text. The ‘content’ column is the text extracted from the JSON file. For example, a paragraph of text is split using the keywords ‘Results and discussion’ and ‘Conclusion’. The ‘section name’ of this paragraph should be ‘Results and discussion’. Note that all tools from OpenAI is useable. I have already obtained an API key from OpenAI and installed the Python module OpenAI.

3. Call GPT-4 API with the same prompt. Every ‘content’ in the data frame should be input to a prompt that to be used in GPT-4 API. The model used in API is ‘GPT-4.1-2025-04-14’. I will provide you the prompt later.

Should you have any questions or find any aspects of this prompt unclear, please include your inquiries in your response.

1.3 从JSON文件中读取Markdown表格的脚本

You are an expert at coding. Your job is to write a Python script aiming to extract information of lithium metal oxide that is used as lithium-ion battery cathode from the given JSON file. The script should realize the following functions:

1. Extract the markdown tables in all 'response' keys. Note that I need only the content of the markdown tables.

2. Add one column named "PII of the source paper" to the table extracted, fill this column with the file name of the file from which the table is extracted. Note that ".json" is not needed in the table. For example, the "PII of the source paper" of the table you extracted is "S2352152X23023447", and the "PII of the source paper" of the table I provide to you is "S0167577X22021231". Note that the JSON file has a key named "file name". The content of this key is like "S0167577X22021231.json". You can extract the content to fill the "PII of the source paper" from this key.

3. Merge all markdown tables to one table

I have provided four JSON files for you to analyze.

1.4 生成辅助清洗无关材料的脚本

You are an expert at coding. Help me write a Python script. The script should realize the functions below:

1. Read data from the Excel file I provided to you, extract all the unique value in “Zeolite name” column.

2. Print one unique value and the corresponding "PII of the source paper" column on the screen each time. If one unique value corresponds multiple rows in "PII of the source paper" column, print one of them is enough. Ask me whether it needs to be moved. IF my answer is “y”, execute Step 3. IF my answer is “n”, move the corresponding row into another Excel file named “DeleteTable.xlsx”.

3. First, ask me whether any information should be added. IF my answer is “y”, ask me what information should be added. After I tell the script the information that I want to add, the script should add the information to “Condition” column. Then, move the corresponding row into an Excel file named “ResultTable.xlsx”. IF my answer is “n”, move the corresponding row into an Excel file named “ResultTable.xlsx” directly.

4. The script should document all the error when executing the script. And the script should show me the progress of the execution.

Should you have any questions, ask me in your response.

1.5 生成训练机器学习模型的脚本

You are an expert at coding and data science. Generate some MATLAB scripts for me to establish machine learning models. The functions of the scripts are shown below, each function should be written to a single MATLAB code (\*.m) file:

1. Load data function code file. The function should load the original data from Excel file, then save the original data table in a MATLAB data file (OriginalData.mat).

2. Divide the training set and testing set function code file. The function should first load OriginalData.mat, then randomly divide training set and testing set in a ratio of 8:2. This function should save the training set and testing set in Dataset.mat, and save the testing set location in TestLocationInformation.mat.

3. Train machine learning models function code file. The function should first load Dataset.mat. Then the script should train five regression models: support vector machine, Gaussian process regression, decision tree regression, regression train ensemble and neural network. Each model should be saved in a single .mat file. For example, the script saves support vector machine model to Model\_SVM.mat, and saves Gaussian process regression model to Model\_GPR.mat.

4. Two model evaluation function code files. The first function code file receives actual and predicted values of training set and testing set, and outputs the Pearson correlation coefficient, mean absolute error and root mean square error between the actual values and predicted values. The second function code file should load all the model data files and Dataset.mat. Then the script should predict the output variables based on the input variables of training set and testing set and the models. Then the script calls the first function file to calculate the evaluation parameters of all five models, and saves the evaluation results of each model to a single .mat file. For example, the script saves evaluation results of support vector machine model to Evaluation\_SVM.mat, and saves evaluation results of Gaussian process regression model to Evaluation \_GPR.mat.

5. Shapley explanation function file. The script should load all the model data files and Dataset.mat. Then the script should use every data point as query point to calculate Shapley values of all models. The script should save the Shapley calculation results in ShapleyExplanation.mat.

6. Main code file. The file should call all the other functions and output a message on the command window when each step is completed.

Should you have any questions, ask me in your response.

2. 信息提取脚本

You are a researcher in the field of anomalous diffusion in zeolite who is good at extracting information from text. Your job is to extract information of zeolite from the given text. This is the text which may or may not include the information wanted:

'''

[text content]

'''

Your mission is to extract the following information from the given text:

1. The name and composition of the zeolite;

2. The framework density and the space group of the zeolite;

3. The species that diffuse in the zeolite, including their name, composition and carbon number;

4. The concentration of the species that diffuse in the zeolite;

5. The anomalous diffusion coefficient and the anomalous diffusion exponent of the species that diffuse in the zeolite;

5. The temperature when calculating the anomalous diffusion coefficient;

6. The pore structure of the zeolite, including its accessible surface area, not accessible surface area, accessible volume, not accessible volume, volume fraction, largest free sphere, largest included sphere, largest included free sphere, channel dimensionality, and combined structure number;

7. Other parameters of the zeolite.

Please follow these steps to extract information from the given text:

1. Analyze the text, answer this question according to the content of the text: “Whether the text about the information mentioned above or not? Your answer should only be yes or no.” Keep in mind that do not output your answer, but execute Step 2.

2. If your answer in Step 1 is “Yes”, continue the process to step 3. If your answer in Step 1 is “no”, end the process and output nothing.

3. Identify the information mentioned previously. IF you find no relevant variables, output nothing.

4. Extract name of the information, their values and their units.

5. Organize these into a table with following columns: Zeolite Name, Variable Name, Numeric Value and Unit.

6. END.

Below I will provide you a sample table:

'''

| Zeolite Name | Variable Name | Numeric Value | Unit |

| -|anomalous diffusion coefficient | 5.8E-5 | Å2/psα|

'''

My other requirements are:

1.You have unlimited time to analyze the text.

2. If one variable has multiple values, you should show me all values rather than only one value. For example, you should show me:

'''

| Zeolite Name | Variable Name | Numeric Value | Unit |

|- |anomalous diffusion coefficient |5.8E-5|Å2/psα|

|- |anomalous diffusion coefficient |1.7E-5|Å2/psα|

'''

You should not show me:

'''

| Zeolite Name | Variable Name | Numeric Value | Unit |

|- |anomalous diffusion coefficient |5.8E-5 to 1.7E-5|Å2/psα|

'''

3. If you failed to extract the information wanted or there is no wanted information in the text, your output should only be:

'''

Nothing extracted.

'''

If you succeed in extracting the information wanted, your output should only be:

'''

Succeed in extracting information:

[The markdown table]

'''

Remember, NO other sentences or other reply is needed.