背景

As we stand on the precipice of the age of artificial general intelligence (AGI), the potential for synergy between AI and chemistry is vast and promising.

The idea of creating AI powered chemistry assistants offers unprecedented opportunities to revolutionize the landscape of chemistry research by applying knowledge across various disciplines, efficiently processing labor-intensive and time-consuming tasks such as iterature searches, compound screening, and data analysis。

The comprehensive summary of chemical information from literature reports, such as publications and patents, and their subsequent storage in an organized database format is the next logical and necessary step toward the discovery of materials.

However, tradition approaches can be labor-intensive and necessitate expertise in coding, computer science, and data science. Furthermore, they are less generalizable, requiring the program to be rewritten when the target changes.

This paper demonstrates that LLMs, including ChatGPT based on the GPT-3.5 and GPT-4 models, can act as chemistry assistants to collaborate with human researchers, facilitating text mining and data analysis to accelerate the research process.

Now Let‘s get into next part .part two methods.

The authors did the following things to let enable ChatGPT to process each paper.

方法

**Design Considerations for ChatGPT-Based Text Mining**

1. Locating potential sections containing synthesis conditions with in the document.
2. Confirming the presence of synthesis conditions in the identified sections
3. Extracting synthesis parameters one by one

**Prompt Engineering**

1. Minimizing hallucination
2. Implementing detailed instructions
3. Requesting structured output

**ChatGPT-Assisted Python Code Generation and Data** **Processing**

Using ChatGPT create Python scripts for parsing academic papers, generating prompts, executing text , and collating the responses into cleaned, tabulated data .

Here is the Schematics of the ChatGPT Chemistry Assistant workflow

It has three different processes employing ChatGPT and ChemPrompt for efficient text mining and summarization of MOF synthesis conditions from a diverse set of published research articles. Each process is distinctively labeled with red, blue, and green dots.

For example ,process 1 is labeled by red dots. It is initiated with “Published Research Articles”, proceeds to “Human Preselection”, moves to the “Synthesis Paragraph”, integrates “ChatGPT with Chem-Prompt”, and culminates in “Tabulated Data”.

Here is the Illustration of a carefully designed ChemPrompt encapsulating all three fundamental principles of ChemPrompt Engineering

The prompt guides ChatGPT to systematically extract and summarize synthesis conditions from a specified section in a research article, organizing the data into a well-structured table.

总结

This research has successfully demonstrated the potential ofLLMs, particularly GPT models, in the domain of chemistry research. It presented a ChatGPT Chemistry Assistant that includes three different but connected approaches to text mining with ChemPrompt Engineering:

We elucidated two crucial insights from the data set ofsynthesis conditions.

First, the data can be employed toconstruct predictive models for reaction outcomes, which shed light on the key experimental factors that influence the MOF crystallization process.

Second, it is possible to create an MOF chatbot that can provide accurate answers based on text mining,

On a fundamental level, this study provides guidance on interacting with LLMs to serve as AI assistants for chemists, accelerating research with minimal prerequisite coding expertise and thus bridging the gap between chemistry and the realms of computational and data science more effectively.