When LLM Meets Material Science: An Investigation on MOF Synthesis Labeling

Xintong Zhao[1], Kyle Langlois[2], Jacob Furst[2], Scott McClellan[1], Rob Fleur[1], Yuan An[1], Xiaohua Hu[1], Fernando Uribe-Romo[2], Diego Gualdron[3] and Jane Greenberg[1]
[1]*Metadata Research Center*, *Drexel University*, Philadelphia, USA
[2]*Department of Chemistry*, *University of Central Florida*, Orlando, USA
[3]*Chemical and Biological Engineering*, *Colorado School of Mines*, Golden, USA

I. Introduction

Recent developments in Large Language Models (LLMs) have advanced the natural language processing (NLP) studies to a new era [1], [2], [4]–[6]. In generic domains, LLMs have become a key component in wide variety of state-of-the-art NLP tasks. In addition, prompt learning enables LLMs-based models to reach robust performance with much smaller training data.

Even with these advances, the performance of LLMs has yet to be fully explored for granular, domain-specific tasks. We consider this limitation, this paper reports on a case study focused on LLM and metal-organic frameworks (MOFs), which are part of part of the larger domain of materials science. The research investigates overall performance of LLMs for named entity recognition (NER) related to materials synthesis.

II. Method

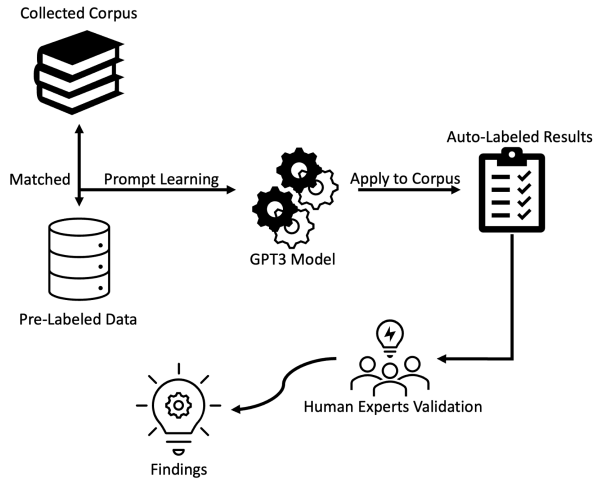

Fig. 1: Overall Workflow to Measure the Performance of LLM in MOF NER

The overall research workflow is shown in figure 2. We collected 172,000 research articles in xml format from ACS, then we matched these article texts with synthesis data extracted from previous study [3] by DOIs (digital object identifiers). We used prompt-based method and 400 matched records as examples for fine-tuning the GPT3 Curie model to perform named entity recognition. Next, we applied the fine-tuned GPT3 model to the rest of MOF articles to extract important entities from synthesis paragraphs. The fine-tuning and extracting process were done via OpenAI API. Finally, our collaborators, who were MOF researchers, validated the synthesis paragraphs annotated by GPT3 and finally we conducted quantitative analysis to measure the performance of GPT3 models from multiple aspects. The types of entities invovled in this study are shown in table I.

This work is supported by NSF OAC # 2118201.

Entity Types	Definition
Metal Precursor	Metal salt or coordination complex used to form the metal nodes/clusters in the metal-organic framework.
Organic Precursor	Organic ligands used in the synthesis of metal-organic frameworks. Form the connections between metal nodes/clusters.
Solvent Precursor	Solvent used in the reaction for the synthesis of the metal-organic framework. Metal and organic precursors usually dissolved in the solvents for the solvothermal synthesis.
Acid	Acid is added to the metal-organic framework mixture to help modulate the reaction to affect the quality of metal-organic frameworks produced.
Vessel	Container in which the reactions take place. Can also be external containers for reaction vessels such as insulated ovens.
Descriptor	General term to represent a classification or grouping. Such as physical terms like yields or temperature.
Resulted MOF Chemical Composition	Chemical formula representing the composition of the metal-organic framework.

TABLE I: Entity Types and Definition

III. Experiments and Pivot Findings

Compared GPT labeled entities and human experts labeled entities, there is a significant difference in the total number: we found that GPT3 recognized 439 entities vs. 833 entities recognized by domain scientists.

However, among 439 GPT-labeled entities, over 87% of their index span are also confirmed by human annotators, 81% of them are correct in terms of both index span and entity types.

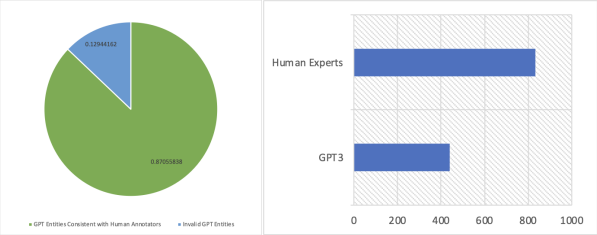

Fig. 2: Difference in GPT labeled entities and human labeled entities

IV. Conclusions and Next Steps

Overall, we found that the tested GPT3 model is reliable to extract important phrases, although it has limitations recognizing important entity types. Our next steps, we plan to introduce domain knowledge to the LLM model to improve its ability to identify domain-specific entity types.

References

[1] T. B. Brown et al., "Language Models are Few-Shot Learners." arXiv, 2020. doi: 10.48550/ARXIV.2005.14165.

[2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." arXiv, 2018. doi: 10.48550/ARXIV.1810.04805.

[3] H. Park, Y. Kang, W. Choe, and J. Kim, "Mining Insights on Metal–Organic Framework Synthesis from Scientific Literature Texts," Journal of Chemical Information and Modeling, vol. 62, no. 5. American Chemical Society (ACS), pp. 1190–1198, Feb. 23, 2022. doi: 10.1021/acs.jcim.1c01297.

[4] H. Touvron et al., "Llama 2: Open Foundation and Fine-Tuned Chat Models." arXiv, 2023. doi: 10.48550/ARXIV.2307.09288.

[5] H. Touvron et al., "LLaMA: Open and Efficient Foundation Language Models." arXiv, 2023. doi: 10.48550/ARXIV.2302.13971.

[6] Y. Liu et al., "RoBERTa: A Robustly Optimized BERT Pretraining Approach." arXiv, 2019. doi: 10.48550/ARXIV.1907.11692.