Research on Named Entity Recognition from Patent Texts with Local Large Language Model*

Chi Yu[1,†], Liang Chen[1,*,†] and Haiyun Xu[2,†]

[1] Institute of Scientific and Technical Information of China, Fuxing Road 15, 100038 Beijing, P.R.China
[2] Business school, Shandong University of Technology, Xinchunxi 266, 255000 Zibo, P.R.China

Abstract

Named entity recognition (NER) from patent texts is one of the fundamental tasks in technical intelligence analysis. However, state-of-the-art performance of NER is achieved at the cost of massive labeled data and intensive labor, which is cumbersome and time consuming. To address the issue, this paper proposed a new framework which employs large language model (LLM) to fulfill the task of NER. Specifically, 3 different prompt templates are designed for NER and efficient fine-tuning algorithm is also utilized to improve their performances. To demonstrate the characteristics of our method, extensive experiments are conducted based on a patent dataset pertained to magnetic head in hard disk drive, namely TFH-2020. Experimental results show that, even though from the perspective of supervised learning, there is a considerable gap between our method and SOTA methods, from the perspective of few-shot learning, our method outperformances similar methods by a large margin.

1. Introduction

Named entity recognition (NER) seeks to locate and classify named entity mentions in unstructured text into pre-defined categories, thus to resolve the issue of ambiguity within free texts. But when it comes to patent text, challenges arise not only from the scarcity of labeled patent dataset, but also from the characteristics of patent texts, such as long sentence with domain-specific and novel terms that are difficult to understand, complex structure of sentences in patent claims that difficult for syntactic parsing. These issues severely hinder the application of NER technologies in patent texts.

The emergence of large language model (LLM) provides a new way to address the issues. That is, by leveraging its capabilities of knowledge storage, semantic understanding and text generation, NER task can be conducted with minimal labeled data and applicable to all domains. To demonstrate the validity and feasibility of the thought, this study proposes a framework for NER based on locally deployed LLM, as ChatGPT or GPT4 is inaccessible in China. Furthermore, efficient fine-tuning algorithm is also utilized in our study to explore the capability of LLM in NER task.

The organization of the rest of this paper is as follows. In Section 2, a LLM-based method is put forward with 3 types of prompt template for NER in patent texts. Then, extensive experiments are conducted on the corpus of TFH-2020[1] to illustrate its performance in Section 4 . The last section concludes this contribution with future study directions.

2. Methodology

Since LLM primarily serves as a chatbot, NER is treated as a question-answering task in this paper. Specifically speaking, through prompt templates, the target text is transformed into a question and submitted to LLM to

identify named entities within it. In this process, the quality of prompt template determines the performance of LLM in NER task. According to the principles of prompt template design proposed by I.Fulford and N.Andrew[2], three prompt templates are proposed include baseline prompt, two-step prompt and multi-entity-type prompt which are shown in Fig.1. In the meanwhile, efficient fine-tuning algorithm is also utilized to improve the performance of LLM.
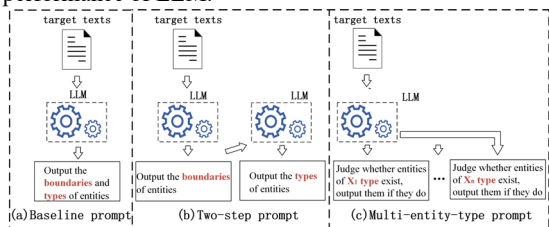

Fig.1 The three prompt templates for NER task

3. Experiment

3.1. Data Preparation and LLM Selection

TFH-2020 corpus [2] is taken as the experimental dataset in this study. It contains 1,010 patent abstracts pertaining to thin film head technology in hard-disk drive collected from the USPTO (United States Patent and Trademark Office) database. To describe the structure of the invention, L.Chen et al [1] defined 17 types of entities such as *system*, *component*, *function*, *effect*, *consequence* etc., we refer the readers to L.Chen et al [1] for more details on this corpus. We construct prompts, instructions and test data from the sentence level. In detail, the patent abstracts in TFH-2020 are divided into 3,384 sentences, which are then randomly split into training and testing sets at a 9:1 ratio. The sentences in the training set are used as examples to fill in the prompts and build instructions as well, while those in the testing set serve to evaluate the performance of LLM in NER task. LLAMA-7B-chat released by META Inc is employed as the LLM in this study.

3.2. Experimental result and analysis

To evaluate the performance of the proposed method, the three prompt templates mentioned above are utilized to for NER task, and the performance of LLM are analyzed from two perspectives: first, the impact of varying number of examples in prompt, and second, the performance of efficient fine-tuning algorithms for different prompt templates. The weighted-average precision, recall, F_1-value and the error rate of output format are shown in Fig. 10. It can be observed that:

(1) For contextual learning, as the number of examples increases, the performance of baseline prompt undergoes a trend of first increasing and then decreasing. When the number of examples is 3, its F1-value reaches maximum of 31%, with corresponding precision rate of 32% and recall rate of 31%, respectively. The performance of two-step prompt exhibits a similar trend. As for the counterpart of multi-entity-type prompt, it remains stable with the varying number of examples.
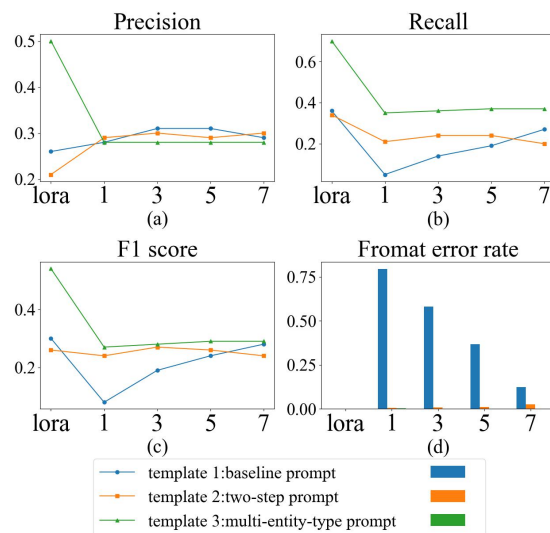

Fig.2 The performance of different prompt templates in NER task

(2) As for LoRA, the three templates exhibit distinct performance. After fine-tuned, the performance of mutli-entity-type prompt gains significant improvement, with its F1-value rising from 27% to 49%. In contrast, the performances of baseline prompt and two-step prompt are inferior to their counterparts of contextual learning.

4. Conclusion

This study proposes a framework for NER task from patent texts with local LLM. It is found that prompt template not only affects the performance of LLM but also influences the effectiveness of efficient fine-tuning method. Simple prompt templates tend to yield better results, while complicated prompts significantly increase the difficulty for both LLM and fine-tuning algorithm.

Acknowledgements

References

[1] L.Chen, S.Xu, L.Zhu, J.Zhang, X.Lei & G.Yang. A deep learning based method for extracting semantic information from patent documents. Scientometrics, 125(2020) 289–312.

[2] I.FULFORD, N.ANDREW. ChatGPT prompt engineering for developers, 2023. URL: https://learn.deeplearning.ai/chatgpt-prompt-eng/

[3] E.J.Hu, Y.Shen, P.Wallis, Z.A.Zhu, Y.Li, S.Wang, L.Wang, W.Chen. Lora: Low-rank adaptation of large language models, 2021. arXiv preprint arXiv:2106.09685.