ESG-GPT:GPT4-based Few-Shot Prompt Learning for Multi-lingual ESG News Text Classification

Ke Tian, Hua Chen
Rakuten Group Inc, JiangXi Normal University
Tokyo, Japan, Jiangxi Province, China
tianke0711@gmail.com, hua.chen@jxnu.edu.cn

Abstract

Environmental, Social, and Governance (ESG) factors for company assessment have gained great attention from finance investors to identify companies' risks and growth opportunities. ESG Text data regarding the company like sustainable reports, media news text, and social media text are important data sources for ESG analysis like ESG factors classification. Recently, FinNLP has proposed several ESG-related tasks. One of the tasks is Multi-Lingual ESG Issue Identification 3(ML-ESG-3) which is to determine the duration or impact level of the impact of an event in the news article regarding the company. In this paper, we mainly discussed our team: KaKa's solution to this ML-ESG-3 task. We proposed the GPT4 model based on few-shot prompt learning to predict the impact level or duration of the impact of multi-lingual ESG news for the company. The experiment result demonstrates that GPT4-based few-shot prompt learning achieved good performance in leaderboard quantitative evaluations of ML-ESG-3 tasks across different languages.

Keywords: ESG, GPT-4, Few-shot Learning, Prompt Learning

1. Introduction

Recently "Environment, Social, and Governance (ESG)" related issues in the financial domain have gained more and more attention with the goal of building a sustainable environment. ESG evaluation is considered an essential tool for investors to assess a company's sustainability and ethical performance. ESG Text data regarding the company like sustainable reports, media news text, and social media text are important data sources for ESG evaluation like ESG impact level or ESG score prediction. Recently, the FinNLP organizers proposed several ESG-related shared tasks for this topic. In FinNLP-2022 (FinNLP-2022, 2022), they proposed a FinSim4-ESG shared task which is related to the ESG topic detection of ESG term words and sustainable sentence classification. Moreover, multi-lingual ESG identification tasks: ML-ESG-1(FinNLP-2023, 2022). ML-ESG-1 is to classify the ESG-related news into 35 ESG key issues. Furthermore, In real-world application scenarios such as making financial decisions, the opportunity (risk) of short-term or long-term impact of ESG news regarding the company should be taken into account. The ML-ESG task organizers defined the following definitions and categories regarding the impact type and impact duration.

- Impact Type Identification:This single-choice question aims to ascertain the type of impact a news article might have on the company. The possible labels are "Opportunity", "Risk", and "Cannot Distinguish".
- Impact Duration (Length) Inference:This single-choice question seeks to determine the duration of the impact a news article might have on the company. Based on the distinction between short-term and long-term defined above, we present three labels: "Less than 2 years", "2 to 5 years", and "More than 5 years".

Considering the importance of impact type and duration, the FinNLP organizer proposed ML-ESG-2(FinNLP-2023, 2023a) and ML-ESG-3(FinNLP-2023, 2023b) subsequently. ML-ESG-2 is to detect the ESG impact type (opportunity or risk) of ESG news regarding the company. In ML-ESG-3, the goal of this task is to determine the duration or length of the impact an event in the multi-lingual ESG news might have on the company.

To challenge these ESG tasks, a variety of methods have been proposed. NLP and deep learning are the dominant techniques(Ke Tian, 2019)(Ke Tian, 2021)(Ke Tian, 2022). However, it is difficult to apply one trained deep learning or ML model across different language ESG texts, it is required to train multi-models to solve the multi-lingual dataset task. Recently, the emergence of Large Language Models (LLM), represented by ChatGPT, has exhibited great performance in general Natural Language Processing (NLP) tasks and across different language texts. These LLMs can complete various tasks by transforming them into generative paradigms using prompt learning. For example, ChatGPT using prompt learning can perform well on text classification, text generation, sentiment detection, NER extraction, etc. As for the ML-ESG-3 task, there are 5 languages of ESG news tasks, we listed the details of each sub-task in Table 1.

In this paper, we presented our solution to the ML-ESG-3 task. Considering the multi-lingual datasets

Proceedings of the Joint Workshop of the 7th FinNLP, the 5th KDF, and the 4th ECONLP, pages 279–282

Language	Task goal
Chinese	Classify the news text into impact duration labels:"Less than 2 years", "2 to 5 years", and "More than 5 years"
English	There are two tasks in this dataset: Impact Level and Impact Length classification.
French	There are two tasks in this dataset: Impact Level and Impact Length classification.
Japanese	Classify the news text into impact duration labels: "Less than 2 years", "2 to 5 years", and "More than 5 years".
Korean	The are two tasks in this dataset: Impact Type and Impact Length classification.

Table 1: Details of ML-ESG-3 task.

in this task, we created firstly four prompts for each language task goal, then utilized the GPT-4(OpenAI) few-shot learning method to predict the impact type or duration of news text in each task. Our approach achieved 1st place for the French language's impact level and performed well in other sub-tasks.

2. Method

2.1. In-Context Learning

Recently, much research work on large language models (LLMs) has explored the phenomenon of in-context learning (ICL). In this paradigm, an LLM learns to solve a new task at inference time (without any change to its weights) by being fed a prompt with examples of that task. For example, if you ask ChatGPT to categorize different, you might first give it example pieces with their correct categorizations in the prompt, then ask ChatGPT to classify the input text with provided label categories in the prompt. In our proposed method, the GPT-4(Open AI) model is utilized, and GPT-4 is a multi-modal model able to consume 32,768 tokens.

2.2. Few-shot Prompt Learning

Few-Shot Learning is the method where a machine learning model is trained with a minimal set of data to shape its predictions, using only a handful of examples at the time of inference, unlike traditional fine-tuning methods that demand a considerable volume of data for the pre-trained model to fine-tune itself accurately to a new task. Recently, with the advent of cutting-edge Language Models such as OpenAI's GPT-3 and GPT-4, its application has broadened to encompass Natural Language Processing (NLP). Within NLP, Few-Shot Learning is applicable to Large Language Models which have, during their pre-training phase on extensive textual datasets, inherently acquired the capability to undertake a diverse array of tasks. This pre-training equips the models with the ability to generalize, or understand and perform tasks that are similar yet not previously encountered, with a few examples serving as guidance. The method is exemplified through the task of translating between English and French, as illustrated in Fig 1.

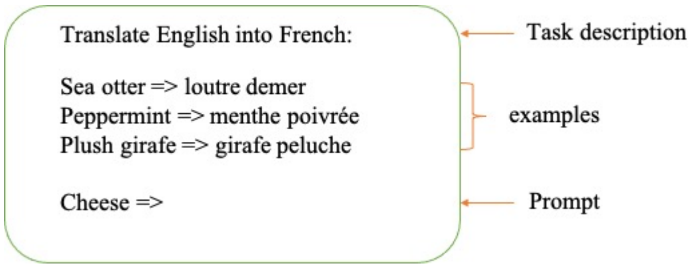

Figure 1: Few-shot prompt learning Example.

2.3. Proposed Method

We proposed the GPT4-based In-context learning with few shot learning for our task, the overall procedure of the proposed method is shown in Fig 2.

Firstly, we utilized the training dataset to create few-shot learning examples for GPT4 model learning. Secondly, the prompt engineering integrates the few-shot learning example, instruction, and test dataset (news title & content) for creating a prompt. Finally, the prompt text as input to call GPT-4 API to predict the input prompt to obtain the result.The schema of the prompt consists of the following components:

- Task description: explain the purpose of the task content.

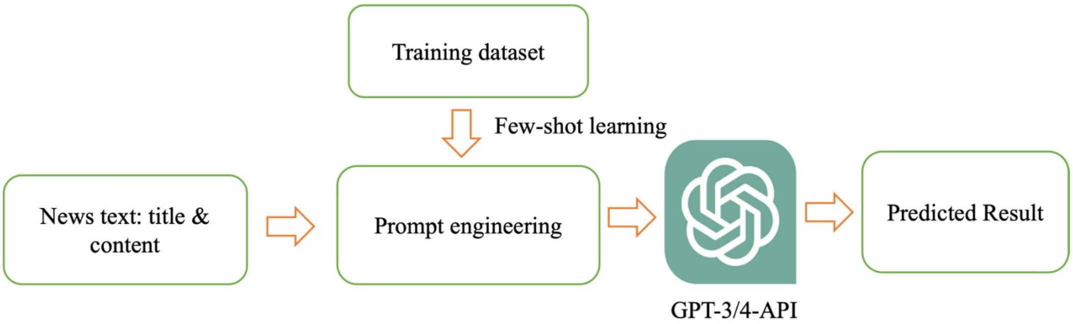

Figure 2: Overall procedure of method.

- Instruction prompt: The sentences that describe the key information regarding the task that the model needs to complete based on required constraints. Moreover, also includes the few-shot learning example based on the train label data.
- Input: the input text for the model predicting Response: the predicted result regarding the input text.

We take the English dataset as an example to explain the above prompt.

As the task description: " Below is an instruction that describes the English ESG text classification task. Write a response that appropriately completes the request. Instruction: " In this task, you are presented with English news titles and their corresponding content, all of which are focused on Environmental and Social (ES) Governance (ESG) themes. Your objective is to evaluate the potential future impact level of these news items to assist in predicting the viability of investments based on their environmental and social implications. There are three categories to classify the impact level: low, medium, and high. Below are 9 examples of English news texts, including both the title and content, related to impact level in the end as follows: created few-shot examples. Your task is to classify the input text, which consists of the news title and content, into one of the three impact level categories: low, medium, and high. Please respond with only one of the three category labels: low, medium, and high, based on your analysis of the news' future impact level. Do not include any additional words or explanations in your response."

As examples in the above string, we used the following example format: input: news title: Guest Post: Carbon Trading and Transfer Pricing; News Content: In order to meet overall carbon emissions ... impact length: long.

input: news title: Guest Post: Eaton Appoints Harold Jones as Chief; News Content: Eaton Appoints Harold Jones as Chief Sustainability Officer; impact length: medium.

3. Result

To understand the model's performance, the organizer used micro-F1 and macro-F1 indicators to evaluate the performance of each submission. The performance of our submitted result is displayed in Table 2.

As for the English task result, the impact level and impact length result are obviously different which is caused by the different number of few-shot learning examples. The result shows that the more few-shot learning examples are better for helping the GPT4 model to understand the semantic meaning of the task. The French submission result achieved the best performance for the impact level in the leaderboard. Our proposed method also achieved good performance in the Korean and Japanese dataset tasks. However, there is still a gap between our solution results and other top results in the leaderboard.

4. Conclusion

We presented GPT-4 mode-based few-shot text classification for the ML-ESG-3 task. We demonstrated that generative LLMs, like GPT-3.5 and GPT-4, can perform well in solving the multi-lingual ESG text classification. Although the French submission result achieved the best performance for the impact level in the leaderboard. There is still a gap between our solution results and other top results in the leaderboard. The following direction will be conducted in the next step: Firstly optimizing the prompt engineering to obtain a better result. Secondly, fine-tuning the GPT-3/4 model or other LLMs to make the GPT model understand well the training dataset.

5. Acknowledgements

This research has been supported by the Jiangxi Double Thousand Plan (JXSQ2019101077) fund, China.

Language	Impact Length		Impact Level		Impact Type	
	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1
English	52.94%	36.36%	51.47%	51.07%	-	-
French	46.58%	47.42%	63.70%	63.29%	-	-
Japanese	34.90%	25.50%	-	-	-	-
Korean	56.00%	52.94%	-	-	63.00%	55.53%

Table 2: Result of sub-task submission in leaderboard.

6. References

FinNLP-2022. 2022. Shared task finsim4-esg. Accessed: 2024-03-08.

FinNLP-2023. 2022. Multi-lingual esg issue identification (ml-esg) shared task. Accessed: 2024-03-08.

FinNLP-2023. 2023a. Multi-lingual esg impact type identification (ml-esg-2) shared task. Accessed: 2024-03-08.

FinNLP-2023. 2023b. Multi-lingual esg impact type identification (ml-esg-3) shared task. Accessed: 2024-03-08.

Hua Chen Ke Tian. 2021. aiai at the finsim-2 task: Finance domain terms automatic classifica- tion via word ontology and embedding. In *In The 1st Work- shop on Financial Technology on the Web (FinWeb) The Web Conference*, Ljubljana, Slovenia.

Hua Chen Ke Tian, ZePeng Zhang. 2022. Automatic term and sentence classification via augmented term and pre-trained language model in esg taxonomy texts. In *The 4th Workshop on Financial Technology and Natural Language Processing (FinNLP) In conjunction with IJCAI-ECAI-2022*, Vienna, Austria.

Zi Jun Peng Ke Tian. 2019. aiai at finsbd task: Sentence boundary detection in noisy texts from financial documents using deep attention model. In *Proceedings of the First Workshop on Financial Technology and Natural Language Processing*, Macao, China.

OpenAI. 2023. *GPT-4 technical report*. Open AI.