

Insights for Auto Port Ranking with LangChain-LLMs

Zehao Qian zehao.qian.cn@gmail.com

January 3, 2024

Contents

1	Task Schedule Model	2
2	LLM Fine-tuning and Prompt Engineering for Auto Port Ranking	2
3	Port Data Acquisition	2
4	Dealing with Outliers and Missing Values	2
4.1	Outliers	2
4.2	Missing Values	2
	References	3

1 Task Schedule Model

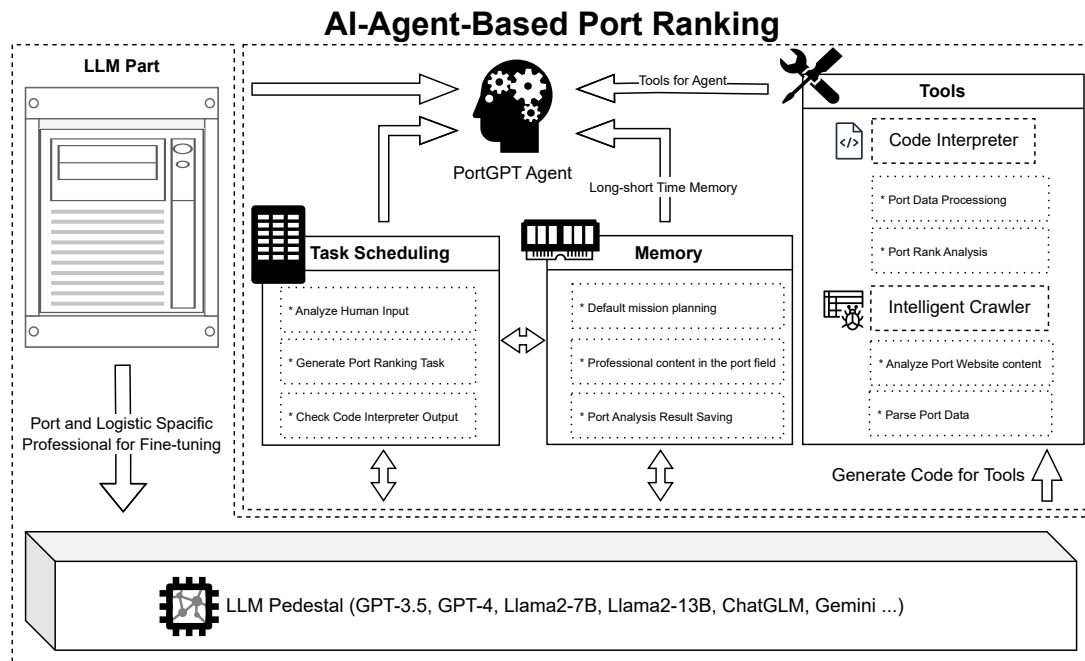


Figure 1: PortGPT Conceptual Architecture

[1]

2 LLM Fine-tuning and Prompt Engineering for Auto Port Ranking

3 Port Data Acquisition

4 Dealing with Outliers and Missing Values

4.1 Outliers

4.2 Missing Values

Delete or Impute? When deciding whether to delete or impute missing data, consider the proportion and pattern of missingness, the nature of the data, the requirements of the analysis method, and the purpose of the study. Deletion is suitable for small proportions of randomly missing data and when it won't introduce bias, while imputation is preferred for large proportions of missing data, non-random missingness, to maintain dataset size and integrity, and to retain critical information. The decision should balance the characteristics of the data, the reasons for missingness, and the objectives of the analysis.

References

- [1] Yinan Mei, Shaoxu Song, Chenguang Fang, Haifeng Yang, Jingyun Fang, and Jiang Long. Capturing semantics for imputation with pre-trained language models. In *2021 IEEE 37th International Conference on Data Engineering (ICDE)*, pages 61–72, 2021.