Prompt Design for Code Generation in Data Analysis Using Large Language Models

Authors : Lu Song Ma Li Zhi

Abstract : With the rapid advancement of artificial intelligence technology, large language models (LLMs) have become a milestone in the field of natural language processing, demonstrating remarkable capabilities in semantic understanding, intelligent question answering, and text generation. These models are gradually penetrating various industries, particularly showcasing significant application potential in the data analysis domain. However, retraining or fine-tuning these models requires substantial computational resources and ample downstream task datasets, which poses a significant challenge for many enterprises and research institutions. Without modifying the internal parameters of the large models, prompt engineering techniques can rapidly adapt these models to new domains. This paper proposes a prompt design strategy aimed at leveraging the capabilities of large language models to automate the generation of data analysis code. By carefully designing prompts, data analysis requirements can be described in natural language, which the large language model can then understand and convert into executable data analysis code, thereby greatly enhancing the efficiency and convenience of data analysis. This strategy not only lowers the threshold for using large models but also significantly improves the accuracy and efficiency of data analysis. Our approach includes requirements for the precision of natural language descriptions, coverage of diverse data analysis needs, and mechanisms for immediate feedback and adjustment. Experimental results show that with this prompt design strategy, large language models perform exceptionally well in multiple data analysis tasks, generating high-quality code and significantly shortening the data analysis cycle. This method provides an efficient and convenient tool for the data analysis field and demonstrates the enormous potential of large language models in practical applications.