LLMs as data analyst

Reyva Babtista

June 30, 2023

1 Cheng, L., Li, X., & Bing, L. (2023). Is GPT-4 a Good Data Analyst?

This papar talked about comparing GPT-4's performance of data analysis to actual data analyst. Their GPT-4 as data analyst's experiment framework are:

1. code generation (for data collection and data visualization),
2. code execution, and
3. analysis generation (for data analysis).

They used nvBench dataset to evaluate their experiments. Their evaluation metrics include figure evaluation (information correctness, chart type correctness, aesthetics) and analysis evaluation (correctness, alignment, complexity, fluency).

Although they found that GPT-4 perfomances are comparable to human data analysts, GPT-4 still suffer from hallucination or poor calculation accuracy. Furthermore, there is still a need of human intervention to:

1. Ask the right questions to be answered by ChatGPT.
2. Prepare the data and execute the code in between the experiment.

2 Hassan, M. M., Knipper, A., & Santu, S. K. K. (2023). ChatGPT as your Personal Data Scientist.

The paper proposed a new architecture utilizing ChatGPT for user-friendly automated machine learning (AutoML). Although it is an interesting idea,

it is not also clear in the paper how their proposed system is user friendly and automated. More time needed to understand the paper since it is long and complex.

3 Noever, D., & McKee, F. (2023). Numeracy from Literacy: Data Science as an Emergent Skill from Large Language Models.

The paper explore the possibility of ChatGPT as a tool for exploratory data analysis. The experiments executed CRUD operations to well known datasets such as Iris, Titanic survival, and Boston housing prices. The random sampling and CRUD operations to the dataset made sure that ChatGPT will not rely on training data but only their own numerical calculations.

Their finding were that the latest LLMs have reached sufficient scale to handle complex statistical questions.

Questions emerge from reading this paper:

1. How can we provide datasets as the "input data" to provide additional context for ChatGPT?

2. Can ChatGPT's memory capacity be increased? Would that help ChatGPT to generate better answers? `https://redis.com/blog/chatgpt-memory-project/`

4 Hassani, H., & Silva, E. S. (2023). The Role of ChatGPT in Data Science: How AI-Assisted Conversational Interfaces Are Revolutionizing the Field. Big Data and Cognitive Computing, 7(2), 62. `http://dx.doi.org/10.3390/bdcc7020062`

A review of ChatGPT in data science. No new methods introduced.

5 Chatterjee, J., & Dethlefs, N. (2023). This new conversational AI model can be your friend, philosopher, and guide .. and even your worst enemy. Patterns, 4(1), 100676. http://dx.doi.org/10.1016/j.patter.2022.100676

This short paper reviews the potential and danger of ChatGPT and show some examples of them. Not necessarily relevant to the topic of utilizing ChatGPT for data science. However, this paper can be cited for our paper's introduction.

6 Karpathy, A. (2023). State of GPT. https://build.microsoft.com/en-US/sessions/db3f4859-cd30-4445-a0cd-553c3

7 Schick, T., Dwivedi-Yu, Jane, Dess\'i, Roberto, Raileanu, R., Lomeli, M., Zettlemoyer, L., Cancedda, N., ... (2023). Toolformer: Language Models Can Teach Themselves to Use Tools.

8 Gao, J., Guo, Y., Lim, G., Zhang, T., Zhang, Z., Li, T. J., & Perrault, S. T. (2023). CollabCoder: A GPT-Powered Workflow for Collaborative Qualitative Analysis.

9 Xiao, Z., Yuan, X., Liao, Q. V., Abdelghani, R., & Oudeyer, P. (2023). Supporting Qualitative Analysis with Large Language Models: Combining Codebook with GPT-3 for Deductive Coding. In , 28th {{International Conference}} on {{Intelligent User Interfaces} (pp. 75–78). : .

10 Ahmed, T., & Devanbu, P. (2022). Few-shot training LLMs for project-specific code-summarization. In , Proceedings of the 37th {{IEEE}}/{{ACM International Conference}} on {{Automated Software Engineering} (pp. 1–5). Rochester MI USA: ACM.