Title: Advancing Seismic Data Exploration with Agentic AI Tools and LLMs

Ignacio Sánchez-Gendriz, Tiago Barros, Victor Gomes, José Augusto Oliveira, Silvan Ferreira, Adriane Praxedes, Gustavo Leitão, Ivanovitch Silva, Luiz Affonso Guedes

This paper was prepared for presentation during the 4th Joint SBGf/SEG Workshop on Machine Learning held in Rio de Janeiro, Brazil, 03-04 December 2024.

Abstract:

Contextualization: Seismic data analysis presents significant challenges due to the vast volume and complex structure of datasets, which include both seismic signal samples and detailed metadata stored in headers. These headers contain essential information for data interpretation, such as source and receiver positions, trace numbers, and offsets. The rise of Machine Learning (ML) and Artificial Intelligence (AI), particularly through agentic tools like PandasAI and LangChain, introduces new possibilities for simplifying this process. By leveraging Large Language Models (LLMs) to interact with seismic records, these tools democratize data exploration, making it available to users without extensive programming knowledge or access to expensive proprietary software.

Main Objectives: This work aims to develop an efficient analysis pipeline capable of loading and manipulating SEGY seismic data headers. The pipeline allows users to extract key structural information, visualize the spatial and temporal geometry of seismic data, and facilitate easier exploration of complex geophysical datasets.

Material and Methods:

The proposed pipeline follows a systematic approach:

1. SEGY files are loaded using Segy-SAK to extract header information.

2. Header data, such as trace counts, source and receiver coordinates, and offsets, are scraped and processed.

3. Advanced analysis and interactive exploration of the header data are conducted using agentic AI tools (PandasAI, LangChain), allowing users to interact with the data through LLMs by issuing commands in natural language.

4. Visualizations, including source and receiver coverage plots and CDP histograms, are generated to analyze acquisition geometry and data distribution.

Results: Initial tests demonstrate the potential of the proposed agentic AI pipeline in automating the extraction and analysis of SEGY headers. Early findings indicate the feasibility of using the presented tools to visualize key variables, such as source-receiver offsets and CDP distributions, providing valuable insights into the spatial organization of seismic data. These visualizations are expected to improve the efficiency of data exploration, particularly for users with limited access to proprietary software or programming expertise, yet faced with the challenge of analyzing large volumes of data.

Conclusions: The integration of agentic AI tools with traditional seismic data processing techniques presents a promising solution for simplifying geophysical data exploration. By automating the extraction of headers and offering intuitive, interactive analysis capabilities, this approach has the potential to greatly improve both the accessibility and efficiency of seismic data analysis. The developed pipeline could benefit both geophysicists and non-programmers, enabling easier data exploration while reducing reliance on costly or complex software solutions.