Explainable AI for Source Code Applications

sdmay25-30

Overview

- Client/Company: Dr. Ali Jannesari/ISU SWAPP Lab
- Abstract:
  - Focus on auto-labeling code datasets using AST tools, regular expressions, and LLM-generated labels.
- Goal:
  - Enhance model interpretability by evaluating learned concepts against human-defined code properties.

User Needs

- Software:
  - All of these users require a working software library to run their workloads
  - It must be compatible with the HPC platform they are using
- Hardware:
  - High performance computing clusters

Requirements

Activation Extraction

Requirement: Add methods for extracting activation data from models.

Clustering

Requirement: Refactor existing clustering algorithms to enhance efficiency and maintainability.

Visualization

Requirement: Refactor visualization tools to provide enhanced clarity and user interaction.

Requirements (cont.)

Alignment/Metrics

Requirement: Provide tools for:

- Evaluating the alignment of clusters with known concepts.

- Applying lexical and contextual criteria.

- Supporting enhanced alignment metrics (specifications to be provided).

Analysis

Requirement: Enhance existing analysis functionalities to provide deeper insights into the results of latent concept analysis.

Engineering Standards

- From ISO and IEEE, we can follow the Software and systems engineering -

Software testing - Part 1: General Concepts to obtain the following

    - Specifically related to AI and neural network-incorporated projects, we must follow model-based testing and ensure reproducible results by saving models and providing initial conditions for any processes.

    - On the testing end, various tests for our pipeline must be implemented. Some example testing formats are: A/B testing, cause-effect testing, and some dynamic testing.

- We will be using the PEP 8 style guide for Python code:

https://peps.python.org/pep-0008/

Conclusions

- Comparison:
  - Our project, due to the nature of its users and goals, requires strict adherence to requirements.
  - Different users will be able to use the features that they need, while not being encumbered by the tools they will not be using.