Towards Automatic Translation of Machine Learning Visual Insights to Analytical Assertions

Arumoy Shome
a.shome@tudelft.nl
Delft University of Technology
Netherlands

Luìs Cruz
l.cruz@tudelft.nl
Delft University of Technology
Netherlands

Arie van Deursen
arie.vandeursen@tudelft.nl
Delft University of Technology
Netherlands

ABSTRACT

We present our vision for developing an automated tool capable of translating visual properties observed in Machine Learning (ML) visualisations into Python assertions. The tool aims to streamline the process of manually verifying these visualisations in the ML development cycle, which is critical as real-world data and assumptions often change post-deployment. In a prior study, we mined 54,070 Jupyter notebooks from Github and created a catalogue of 269 semantically related visualisation-assertion (VA) pairs. Building on this catalogue, we propose to build a taxonomy that organises the VA pairs based on ML verification tasks. The input feature space comprises of a rich source of information mined from the Jupyter notebooks—visualisations, Python source code, and associated markdown text. The effectiveness of various AI models, including traditional NLP4Code models and modern Large Language Models, will be compared using established machine translation metrics and evaluated through a qualitative study with human participants. The paper also plans to address the challenge of extending the existing VA pair dataset with additional pairs from Kaggle and to compare the tool's effectiveness with commercial generative AI models like ChatGPT. This research not only contributes to the field of ML system validation but also explores novel ways to leverage AI for automating and enhancing software engineering practices in ML.

KEYWORDS

SE4AI, NLP4Code, ML Testing, Visualisations, Assertions, Computational Notebooks, Automated Tool

ACM Reference Format:

Arumoy Shome, Luìs Cruz, and Arie van Deursen. 2024. Towards Automatic Translation of Machine Learning Visual Insights to Analytical Assertions. In *Proceedings of 46th International Conference on Software Engineering (ICSE 2024).* ACM, New York, NY, USA, 4 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Visualisations are employed at various stages of a Machine Learning (ML) pipeline. They are used to understand and verify data properties during the early stages, summarise metrics and fine-tune models during development, and monitor performance post-deployment. The iterative and experimental nature of building ML systems heavily relies on insights from visualisations to guide design and implementation decisions [2–4, 8–10].

However, real-world data that ML systems encounter post- deployment seldom remain static. They often change as a reflection of the world, potentially violating initial assumptions made during development [1, 7]. Every subsequent iteration of the ML development cycle used to retrain and update the ML model, therefore demands manual validation of the visualisations that were used to test ML system properties.

Assertions or analytical tests derived from ML visualisations can significantly reduce manual verification efforts. Such formal assertions record the AI practitioner's observations about the model or data at a specific moment. They also serve as a reference point for future AI practitioners to understand the interpretations made from earlier visualisations.

In a prior study currently under review, we investigate how frequently analytical tests are formulated from visualisations in ML systems and analysed the effectiveness of these tests. We mined 54,070 Jupyter notebooks from GitHub that contain assertions written in Python and developed a novel methodology to identify 1,764 notebooks which contain an assertion below a visualisation. We manually analysed the 1,764 notebooks and catalogued 269 semantically related visualisation-assertion (VA) pairs (to be released publicly with the camera-ready version of the paper). Further in-depth analysis of the VA pairs revealed that assertions often fail to capture all the information obtained from the corresponding visualisations. Our results indicate that formulating analytical assertions from visualisations is an emerging testing technique in ML. However, it also highlights the limitations of current practices, demonstrating a need for automated tools that can assist ML practitioners in validating visualisations and formulating more comprehensive analytical assertions.

In this paper, we propose our research vision to develop an automated tool to generate analytical assertions from ML visualisations. The research questions we plan to address along with the contributions we envision to make are presented below.

RQ1. How are VA pairs used to perform ML verification tasks?

We begin by creating a taxonomy that clusters the existing 269 VA pairs by specific ML verification tasks. This taxonomy not only enables us to fine-tune the data and model for the automated tool, but also serves as valuable reference for future ML practitioners.

ICSE 2024, April 2024, Lisbon, Portugal

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/XXXXXXX.XXXXXXX