Towards Next-Gen Machine Learning Asset Management Tools

SAMUEL O. IDOWU

Presentation:
November 20th, 2023, 13:00
Room 520, Jupiter Building
Hörselgången 5,
Chalmers University of Technology, Campus Lindholmen

Discussion leader:
Prof. Martin Monperrus,
KTH Royal Institute of Technology, Sweden

UNIVERSITY OF GOTHENBURG

Abstract

Context: The proficiency of *machine learning* (ML) systems in solving many real-world problems effectively has enabled a paradigm shift toward ML-enabled systems. In ML-enabled software, significant software code artifacts (i.e., assets) are replaced by ML-related assets, introducing multiple system development and production challenges. In particular, the need to manage extended asset types introduced by ML systems and the non-deterministic nature of ML make using traditional *software engineering* (SE) tools ineffective. The lack of supporting tools makes it demanding to address the concerns of specific aspects of ML-enabled system development, such as model experimentation. Consequently, new tool classes are being introduced to address these challenges. ML *experiment management tools* (ExMT) are examples of such tools aiming to mitigate the challenges and users' burden of managing ML-specific assets. Although these tools have recently become available, they are, unfortunately, not fully mature and have the potential for several improvements. For instance, many practitioners still consider ExMTs costly, restrictive, and ineffective. These challenges imply the need for improvements in many areas and raise research questions about the appropriate characteristics of a useful and effective ExMT for managing the development assets of ML-enabled systems.

Objective: This PhD research aims to contribute to the rapidly evolving space of new and improved ExMTs to facilitate the development of improved tools targeting combined SE and data science use cases. Consequently, we contributed to the knowledge and extended insights on ML experiment, their assets, the ExMT's landscape, and their benefits and effectiveness. We later proposed steps towards integrated ExMTs and artifacts based on the obtained insights.

Method: We addressed our objectives by adopting 1) *knowledge-seeking* research, including exploratory studies, literature reviews, feature surveys, practitioner surveys, and controlled experiments, and 2) *solution-seeking* research, including design science proposing unified concepts from multiple tools. The former was used to understand ML experiments, the challenges of managing experiment assets, the state of practice and landscape of existing ExMTs, and their effectiveness, benefits, and limitations. The acquired insights are then leveraged to propose research steps in the later part toward integrated ExMTs using design science to develop a blueprint for unified management tools.

Results: This thesis presents seven significant results. First, it provides an empirically informed overview of the challenges in ML experiment management. Second, it presents insights into the types of ML-based projects, their development activities, and evolution patterns. Third, it offers an overview of existing tools, shedding light on the state of practice and research on asset management tools for ML experiments. Fourth, it presents an empirical-based report on the benefits and challenges of ExMTs. Fifth, it establishes the effectiveness of ExMTs in improving user performance. Sixth, it proposes a step-by-step guide toward integrated ML tools for SE and data science. Seventh, it presents a prototype and blueprint for a unified ExMT.

Conclusion: This thesis highlights the significance of ML asset management as an essential discipline in facilitating experiments and asset management for ML-enabled software systems. It provides empirical data that offers crucial insights into the tooling landscape for managing ML experiment assets, including their features, benefits, limitations, and effectiveness. Additionally, the research proposes a guide and prototype to facilitate the design of new ExMTs.

Keywords