Machine Learning Research Phase Tracking with Hypotheses Graphs

Yongzhi Ong, Vincent Pollet
Amazon Studios Technology

Abstract

A challenging Machine Learning project can be confronted by two mutually dependent ambiguities: what is feasible with science, and what is required and useful for the product. Teams set out on strategies and plans for both threads - multiple research aspects, e.g. training data selection, training data preparation, model design, fine-tuning and optimization, and multi-faceted product subjects, e.g. customer engagement, persona and journeys, business assumptions and requirements. A baseline solution often does not exist, and product clarity demands time. Front-loaded with changing exploratory factors, classic metrics are not reliable as trackers throughout the project lifecycle. We describe our 4-phases approach to project scheduling, leading to the convergence of technical solutions and product definition. We introduce the *hypotheses graph* method in early phases to track the velocity of research, resource and effort allocation. This new method provides numerical measurements of lifecycle progress, beginning with wide-scope and high-risk scientific explorations, gradually transitioning to a converged solution with product-aligned metrics.

Introduction

In a case study, we embark on research simultaneously with product discovery, confronted by ambiguous interim targets of a super-positional problem, i.e. a problem with multiple interdependencies. We face the problem where classic metrics are unreliable as trackers for measuring research progress given the volatile and interdependent (i.e. super-positional) requirements and conditions. Our strategy involves initial wide-scope research producing interim solutions. On a set of central metrics, these would not be meaningfully evaluated and would incur excessive evaluation overhead.

Our proposed solution is a hypotheses graph that captures the outcome of hypotheses on training data, training data preparation, model design, model tuning and refinement. The formulated hypotheses are either affirmed, abandoned, active or open. We show how, as product discovery uncovers requirements and product assumptions are resolved, the research scope narrows and converges towards a suitable solution, and the graph reflects progress through an increasing proportion of affirmed and abandoned hypotheses. Velocity can be measured by the amount of affirmed or abandoned hypotheses weekly. A high proportion of affirmed and abandoned hypotheses indicates the end of the project's exploratory phase, when metrics takeover under established conditions as project trackers and targets, and engineering effort is ramped up for the customer facing product. The hierarchy represents investigative threads that are broken down into child hypotheses. We demonstrate that the sub-hierarchies are useful to monitor research focus, if the right proportion of inquisitions are made in relation to use cases, and if team resource allocation is reasonable. We share our lightweight routines to update the graph, integrated with our issue management and report platforms (JIRA and Confluence), and inform regular risk assessment mechanisms with graph information.

Potential discussion points

- In practice, the hypotheses graph is weakly associated with Scrum sprint tasks, but is maintained independently and does not impose restrictions on Scrum activities. We discuss potential improvements to the method by aligning Scrum practices to the approach, e.g. by specifying Sprint tasks as hypotheses during the exploratory phases of the project, assigning effort weights to hypotheses, and tracking burn-down of hypotheses in sprints.
- The method we describe requires manual effort for graph visualizations and trend observations, although charts and plots are semi-automated via JIRA. Potential automation to alleviate manual effort, such as generating visualizations, alerts and notifications, and reporting routines are interesting to explore.

Relevance to workshop

This talk presents a case study of a new data-driven method, the hypotheses graph, for an audacious Machine Learning project lifecycle tracking and resource management. Workshop attendees will benefit from the proposed tool that mediates scientific solution search alongside product ambiguity, producing data to help inform themselves and business stakeholders of the status of exploratory projects, while giving product teams and users time for discovery, and avoiding miscommunication with premature metrics.

Biography of main presenter

Yongzhi Ong is a Technical Program Manager at Amazon running a machine-learning project portfolio on stringent timelines with scientific and product challenges. He is focused on helping research and development teams provide top value to customers. Yongzhi applies 12 years of experience in software delivery and research-to-market projects to drive effective machine-learning project execution.

Company or project portrait

Amazon Studios is an American television and film producer and distributor that is a subsidiary of Amazon. It specializes in developing television series and distributing and producing films.