A Common Misassumption in Online Experiments with Machine Learning Models*

Abstract

Olivier Jeunen

ShareChat, United Kingdom

Abstract

Online experiments such as Randomised Controlled Trials (RCTs) or A/B-tests are the bread and butter of modern platforms on the web. They are conducted continuously to allow platforms to estimate the causal effect of replacing system variant "A" with variant "B", on some metric of interest. These variants can differ in many aspects. In this paper, we focus on the common use-case where they correspond to machine learning models. The online experiment then serves as the final arbiter to decide which model is superior, and should thus be shipped.

The statistical literature on causal effect estimation from RCTs has a substantial history, which contributes deservedly to the level of trust researchers and practitioners have in this "gold standard" of evaluation practices. Nevertheless, in the particular case of machine learning experiments, we remark that certain critical issues remain. Specifically, the assumptions that are required to ascertain that A/B-tests yield unbiased estimates of the causal effect, are seldom met in practical applications. We argue that, because variants typically learn using pooled data, a lack of model interference cannot be guaranteed. This undermines the conclusions we can draw from online experiments with machine learning models. We discuss the implications this has for practitioners, and for the research literature.

Keywords

Online Evaluation, A/B Testing, Stable Unit Treatment Value Assumption (SUTVA)

Perspectives on the Evaluation of Recommender Systems Workshop (PERSPECTIVES 2023), September 19th, 2023, co-located with the 17th ACM Conference on Recommender Systems, Singapore, Singapore.

*The full-text version of this work was published as: Olivier Jeunen. 2023. A Common Misassumption in Online Experiments with Machine Learning Models. SIGIR Forum 57, 1 (June 2023), 9 pages. Available at arxiv.org/abs/2304.10900. A short teaser video can be found at youtu.be/piUDxYc_UMk.

jeunen@sharechat.co (O. Jeunen)
https://olivierjeunen.github.io/ (O. Jeunen)
0000-0001-6256-5814 (O. Jeunen)