Large-scale Open Dataset, Pipeline, and Benchmark for Bandit Algorithms

Yuta Saito[1,2], Shunsuke Aihara[3], Megumi Matsutani[3], Yusuke Narita[1,4]

[1]Hanjuku-kaso Co, Ltd. [2]Tokyo Institute of Technology, [3]ZOZO Technologies, Inc. [4]Yale University

Objectives

- We release **Open Bandit Dataset** and **Open Bandit Pipeline** to enable realistic and reproducible experimens on off-policy evaluation of bandit algorithms.
- Our data, pipeline, and benchmark experiments are open-sourced to advance future OPE research. Our implementations help practitioners use logged bandit data to compare different estimators and find an appropriate one to improve their bandit systems.

Introduction

We build and publicize the *Open Bandit Dataset* to facilitate scalable and reproducible research on bandit algorithms. It is especially suitable for *off-policy evaluation* (OPE), which attempts to estimate the performance of hypothetical policies using data generated by a different policy. We construct the dataset based on experiments and implementations on a large-scale fashion e-commerce platform, ZOZOTOWN. The data contain the ground-truth about the performance of several bandit policies and enable fair comparisons of different OPE estimators. We also build a Python package called the *Open Bandit Pipeline* to streamline implementations of bandit algorithms and OPE estimators. Our open data and pipeline will allow researchers and practitioners to easily evaluate and compare their bandit algorithms and OPE estimators with others in a large, real-world setting. Using our data and pipeline, we provide extensive benchmark experiments of existing OPE estimators. Our experiments open up essential challenges and new avenues for future OPE research. Our pipeline and example data are available at **https://github.com/st-tech/zr-obp**. You can follow the updates of the whole project at **https://groups.google.com/g/open-bandit-project**.

Figure 1:Fashion items as actions displayed in ZOZOTOWN.

Open Bandit Dataset

Our open-source data is logged bandit feedback data provided by ZOZO, Inc., the largest Japanese fashion e-commerce company with a market capitalization of over 5 billion USD (as of May 2020). The company recently started using multi-armed bandit algorithms to recommend fashion items to users in their large-scale fashion e-commerce platform called ZOZOTOWN. We present examples of displayed fashion items in Figure 1. We collected the data in a 7-day experiment in late November 2019 on three "campaigns," corresponding to "all", "men's", and "women's" items, respectively. Each campaign randomly uses either the Random policy or the Bernoulli Thompson Sampling (Bernoulli TS) policy for each user impression. Note that we pre-trained Bernoulli TS for over a month before the data collection process and the policy well converges to a fixed one. Thus, we suppose that our data are generated by a fixed policy and apply the standard OPE formulation that assumes static behavior and evaluation policies. The dataset is large and contains many millions of recommendation instances. They also include the true action choice probabilities by behavior policies computed by Monte Carlo simulations based on the policy parameters used during the data collection process. The number of actions is also sizable, so this setting is challenging for bandit algorithms and their OPE. We share the full version of our data at **https://research.zozo.com/data.html**.

OPEN BANDIT PIPELINE™

Figure 2:Open Bandit Pipeline Logo

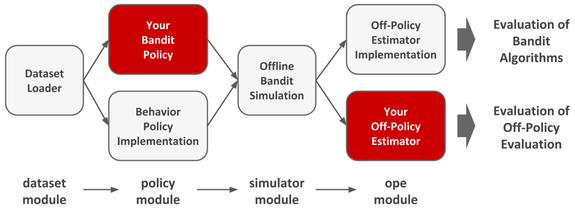

Figure 3:Overview of the Open Bandit Pipeline

Open Bandit Pipeline

To facilitate the usage of OPE, we also build a Python package called the *Open Bandit Pipeline*. Our pipeline contains four main modules (dataset/policy/simulator/ope) as depicted in Figure 3. This pipeline allows researchers to focus on building their OPE estimator and to easily compare it with other methods in realistic and reproducible ways.

To our knowledge, our open-source is the first to include logged bandit datasets collected by running *multiple* different policies, policy implementations used in production, and their ground-truth policy values. These features enable the "***realistic and reproducible evaluation of OPE***" for the first time.

Conclusion and Future Work

To enable realistic and reproducible evaluation of off-policy evaluation, **we have publicized the Open Bandit Dataset–a benchmark logged bandit dataset collected on a large-scale fashion e-commerce platform. The data comes with the Open Bandit Pipeline, a collection of implementations that makes it easy to evaluate and compare different OPE estimators.** We expect them to facilitate the understanding of the empirical properties of the OPE techniques and address experimental inconsistencies in the literature. In addition to building the data and pipeline, we have presented extensive benchmark experiments on OPE. Our experiments highlight that the current OPE methods are inaccurate for estimating out-of-distribution performance of a new policy. It is also evident from the results that it is necessary to develop a data-driven method to tune or select an appropriate estimator for each given environment.

As future work, we aim to constantly expand and improve the Open Bandit Dataset to include more data and tasks. For example, we will add additional log data generated by contextual policies on the platform. Moreover, we assume that the reward of an item at a position does not depend on other simultaneously presented items. This assumption might not hold, as an item's attractiveness can have a significant effect on the expected reward of another item in the same recommendation list. Thus, it is valuable to compare the standard OPE estimators and those for other settings such as the slate recommendation. We plan to allow our pipeline to implement bandit policies and OPE estimators for the slate recommendation setting.

Contact Information

- Yuta Saito
- Website: http://usaito.github.io/
- Email: saito.y.bj@m.titech.ac.jp