Business Data Science

Reinforcement Learning 2024

MAB Assignment

You have now received data (file zozo_Context_80items.csv) sampled from a large-scale experiment ran on a very high-traffic website (Zozo). Zozo provided a description of this dataset and its variables, which I add below for your information

- timestamp: timestamps of impressions.
- item_id: index of items as arms
- position: the position of an item being recommended (1, 2, or 3 correspond to left, center, and right position of the ZOZOTOWN recommendation interface, respectively).
- click: target variable that indicates if an item was clicked (1) or not (0).
- propensity_score: the probability of an item being recommended at each position.user feature 0-4: user-related feature values.
- user feature 0-3: user-related feature values. According to Zozo, two of these variables are potentially sensitive/protected variables (e.g., age group).

The data was collected 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 TS policy for each user impression. These policies select three of the candidate fashion items for each user. The three positions were shown next each other, as shown in the image provided by the company.

They assume that the reward (click indicator) depends only on the item and its position, which is a general assumption on the click generative process used in the web industry. Each row of the data has feature vectors such as age,

gender, and past click history of the users. These feature vectors are hashed, thus the dataset does not contain any personally identifiable information.

Using these data, implement two algorithms of your choice derived from, or extensions of, TS and UCB. Use them to solve the multi-armed bandit problem of finding the best fashion item to serve to visitors, and where to place them. Questions to be tackled:

(a) Fully describe your methods of choice and show how they compare (against each other and against the method used to generate the data) in terms of their ability to generate higher CTR in this data. **[3 pts]**

(b) Show how sensitive your results are to batch size, aggregation/heterogeneity, and parameter tuning. **[3 pts]**

(c) Show how fair your methods are in terms of fairness-to-the-users and fairness-to-the-items. Propose and implement an extension to get more fair recommendations. (Because we do not know exactly which of the four variables is a sensitive variable, you can analyze the four as if they were sensitive variables) **[2 pts]**

(d) Outline model changes for a/the non-stationary case. Empirically demonstrate whether the data used in your analysis is stationary or non-stationary **[2 pts]**