CDT Summer School

- Find Github
- RL Post-Mortem
  - Motivation (10 min)
  - Lecture (30 min)
- Break (10 min)
- HeartSteps (20 min)
- Panel Discussion on Future (20 min)

1

https://github.com/StatisticalReinforcementLearningLab/Stat-ML-CDT-2023/tree/main

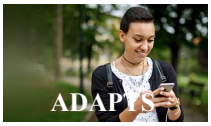
•Lecture 6 (Post-Mortem Analyses): 1 hour 30 min. 11.00 to 12:30 Wed Morning

Reinforcement Learning Post-Mortem

Susan A Murphy

&

Raaz Dwivedi



Oralytics

HeartSteps (PI Klasnja)

Goal: Develop a mobile activity coach for individuals who are at high risk of coronary artery disease

Three iterative studies:

- V1: 42-day micro-randomized study with people who are sedentary
- V2/V3: 90-day + 270-day micro-randomized, personalized, study with people who have Stage 1 Hypertension. n=91 people

3

N=42 from V2 and N=49 from V3 users—total =91

blood pressure that falls in the stage 1 hypertension range

The systolic pressure is 140 to 159 mm Hg or your diastolic pressure is 90 to 99 mm Hg

Changing your lifestyle can go a long way toward controlling high blood pressure.

Eating a heart-healthy diet with less salt

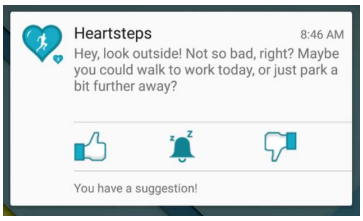
Getting regular physical activity

Maintaining a healthy weight or losing weight if you're overweight or obese

Limiting the amount of alcohol you drink

HeartSteps: Activity Suggestions Component

- 5 time points per day (set according to user's schedule)



- Treatment (Action): Contextually tailored activity suggestion (deliver or not deliver)

- Outcome (Reward): 30-minute step count following each time point.

4

Many components (good morning message, anti-sedentary message, self-monitoring…)

morning, mid-day, mid-afternoon, early evening, after dinner

Reward: Frequently the actions are primarily designed to have a near-term effect on the individual. E.g. Help then manage current craving/stress, help them manage or be aware of the impact of their social setting on their craving/stress

Truth in Advertising

- To "personalize" the actions, A_t, during the HeartSteps trial, the online bandit algorithm was run separately on each of $n = 91$ users.

- The bandit algorithm stochastically selects treatments at each time t during the trial.

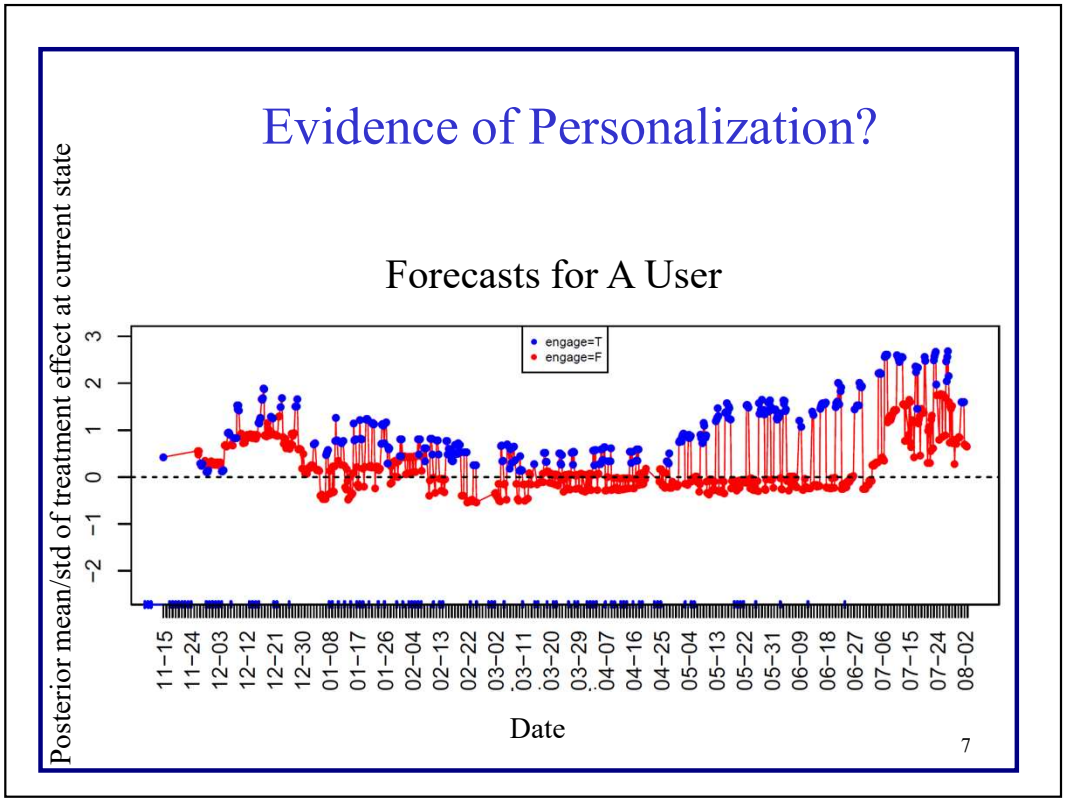
5

Truth in Advertising

- To "personalize" the actions, A_t, during the HeartSteps trial, the online bandit algorithm was run separately on each of $n = 91$ users.
- The bandit algorithm stochastically selects treatments at each time t during the trial.

True or False "Evidence of personalization" ??

False if due to stochastic selection of treatments



Advantage==treatment effect

Engagement: Binary indicator of whether the number of screens encountered in app from prior day from 12am to 11:59pm is greater than the 40% quantile of the screens collected:

1) during MRT week: over the previous days (see the initialization on the right for the first day)

2) after MRT week: over the last seven days

Exploratory Data Analysis

Challenges

1. How to characterize a graph that visually indicates the algorithm is learning/personalizing for the user?

2. How to assess if there are more users (than would occur by chance) with such graphs?

Chance due to stochastic algorithm [8]