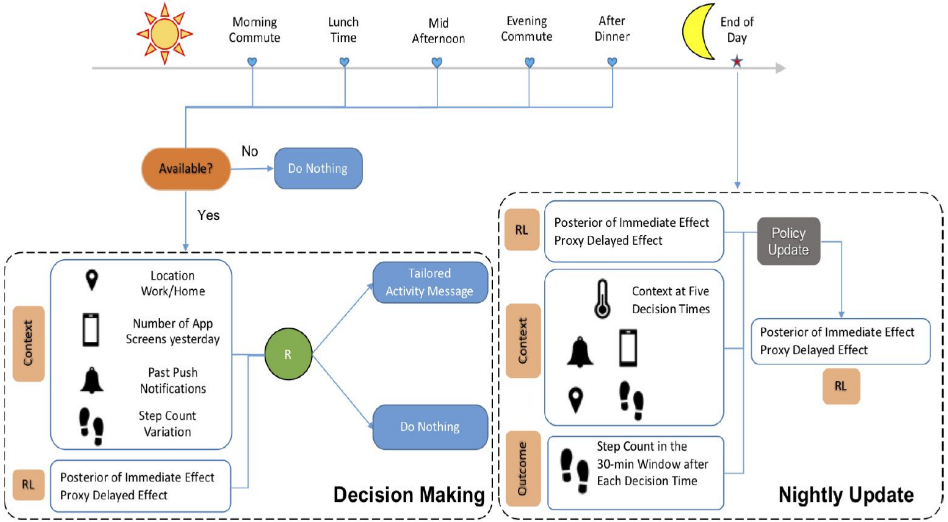
Parallel Reinforcement Learning

team07: Sam DePaolo, Michael Kielstra, Manqing Liu, Xiaohan Wu

Application in health behavioral studies

- **Question:** Whether or not to deliver a physical activity suggestion at each decision time?
- **RL framework**
  - State: user's current and past context (e.g. location, prior 30 mins step count)
  - Action: whether to deliver an activity suggestion or not?
  - Reward: step count 30 mins after decision time
- **Goal**
  - Learn the best sequence of actions (i.e. the policy) to maximize the total reward



Liao, Greenewald, K., Klasnja, P., & Murphy, S. (2020).

A Mathematical Model for RL

- A **bandit** is a random function b: {1, 2,..., n} → ℝ, mapping **actions** to **rewards**
  - We assume b(i) is normally distributed with fixed mean and variance for fixed i
- The RL algorithm can call b but not inspect the code
- Goal: maximize average reward
- This model has severe limitations!
  - Not everything is normally distributed
  - Some things depend on history

Why do we need parallel implementation for RL?

- The only way to learn which actions are best is to take each of them
  - The greater the variance, the more each action must be taken
- In principle, there is no reason why we cannot access a bandit in parallel, decreasing wall clock time to learn

How to parallelize the RL algorithm?

- Use MPI and OpenMP to synchronize data between agents
- Distributed/shared memory model:
  - Speeds up the learning process for each agent by sharing information between the agents
- Optimization:
  - Choosing synchronization frequency between agents
- Evaluating this performance will mean comparing wall clock time for *n* agents to converge on the optimal action choices

References

- Liao, Greenewald, K., Klasnja, P., & Murphy, S. (2020). Personalized HeartSteps. Proceedings of ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 4(1), 1–22. https://doi.org/10.1145/3381007
- Kretchmar, R. M. (2002, July). Parallel reinforcement learning. In The 6th World Conference on Systemics, Cybernetics, and Informatics.