Learning Index Policies for Restless Bandits with Application to Maternal Healthcare
Arpita Biswas, Gaurav Aggarwal, Pradeep Varakantham, Milind Tambe
Google Research

Motivation

Beneficiary’s Engagement (listenership)
- Call records (preventive care information)
- Healthcare worker visiting patients (interventions)
- Timely interventions ➞ Increase in engagement

Problem Statement

Goal
Learning intervention scheme under budget constraints

Post-intervention improvements in engagements are uncertain.

Uncertainty

Cost of Intervention

Human behavior

Sequential decision

The engagement dynamics of the beneficiaries are unknown.

No. of healthcare workers much less than the no. of beneficiaries.

Decisions taken at a timestep impacts the behavior of all the beneficiaries in the next timesteps.

Model: Restless Multi-Armed Bandits

\[
\max_{\pi} \lim_{t \to \infty} \frac{1}{t} \mathbb{E} \left[ \sum_{t=0}^{t-1} \sum_{i \in N} R_i X_i(h) \left( A_i^\pi(h) \right) \right]
\]

s.t. \[ \sum_{i \in N} A_i^\pi(t) \leq M \quad \text{for all } t = \{1, 2, \ldots\} \]

\[ X_i(h) := \text{State of arm } i \text{ at timestep } h. \]

\[ A_i^\pi(h) := \text{Intervention using policy } \pi \]

\[ R_i X_i(h) \left( A_i^\pi(h) \right) := \text{Reward observed.} \]

Empirical Analysis

Intervention Schemes
- WIQL
- Greedy
- Random
- Myopic

Conclusion: WIQL outperforms existing methods.

This work was done when all the authors were affiliated with Google Research.

Contact: arpita.biswas@live.in