Cognitive Modeling and Design of Shared Autonomy

Vaibhav Srivastava
Department of Mechanical & Aerospace Engineering
Princeton University
www.princeton.edu/~vaibhavs

April 19, 2016
Department of Aerospace Engineering
University of Maryland

Shared autonomy is the technology that brings together
- Memory and computational power of machines
- Cognitive abilities of humans
- Mitigates inefficiencies and leverages abilities
- When is shared autonomy helpful?
- Can we integrate human with an arbitrary design of autonomy?

Impact areas
- Environmental monitoring
- Healthcare
- Manufacturing
- Defense

Autonomy Paradox and Information Overload

- Autonomy Paradox: More human resource needed in autonomous operations than manual operations
- Single-operator-multiple-automata architecture
- Information overload
**Autonomy Paradox and Information Overload**

- **Autonomy Paradox:** More human resource needed in autonomous operations than manual operations
- Single-operator-multiple-automata architecture
- Information overload

**Shared Human-Robot Autonomy: My Research Directions**

- Decision support system to mitigate information overload
- Optimal attention allocation suggestions


**Realization of Shared Human-Robot Autonomy**


Shared Human-Robot Autonomy: My Research Directions

- Algorithms for distributed decision-making in sensor networks
- Fusion of heterogeneous sensor data with human input


Outline

1. Introduction
2. Stochastic Multi-armed Bandit Problems
   - Features of human decision-making
   - Upper Credible Limit algorithm
   - Data from experiments with human participants
3. Distributed Decision-Making in Multi-armed Bandit Problems
4. Conclusions and Future Directions

Stochastic Multi-armed Bandit Problems

- $N$ options with unknown mean rewards $m_i$
- the obtained reward is corrupted by noise
- distribution of noise is known $\sim \mathcal{N}(0, \sigma^2)$
- can play only one option at a time

**Objective:** maximize expected cumulative reward until time $T$
Stochastic Multi-armed Bandit Problems

- $N$ options with unknown mean rewards $m_i$
- the obtained reward is corrupted by noise
- distribution of noise is known $\sim \mathcal{N}(0, \sigma^2)$
- can play only one option at a time

Objective: maximize expected cumulative reward until time $T$

---

Multi-armed Bandit Problem

Mersereau, Rusevichientong, and Tsitsiklis. A structured MAB problem and the greedy policy. IEEE TAC, 2009

---

Human Decision-Making in Multi-armed Bandit Tasks

Cohen et al. Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. 2007
Best possible performance and State of the art

**Lai-Robbins Bound**
- Cumulative Regret $\geq K_{\min} \log T$, $T$ = Horizon length

**Upper confidence bound algorithm (Auer et al.’00)**
- play each option once, then at each time $t$ pick arm
  $$\arg\max_i \hat{m}_t^i + C \sqrt{\frac{\log T}{n_t^i}}$$
  - frequentist estimator
  - uncertainty measure
- Cumulative Regret $\leq K_{\text{ucb}} \log T$ for bounded rewards

**Bayesian UCB algorithm (Kauffman et al.’12)**
- at each time $t$ pick arm with maximum $(1 - \frac{1}{T})$-upper confidence bound
- Cumulative Regret $\leq K_{\min} \log T$ for Bernoulli rewards

Outline
- Introduction
- Stochastic Multi-armed Bandit Problems
- Modeling Human Performance in Multi-armed Bandit Tasks
  - Features of human decision-making
  - Upper Credible Limit algorithm
  - Data from experiments with human participants
- Distributed Decision-Making in Multi-armed Bandit Problems
- Conclusions and Future Directions
Spatially Embedded Gaussian Multi-armed Bandits

- reward at option $i \sim \mathcal{N}(m_i, \sigma^2_i)$
- prior on reward surface $m \sim \mathcal{N}((\mu_0, \Sigma_0)$
- spatial structure captured through $\Sigma_0$, e.g., \( \sigma_{ij}^0 = \sigma_0 \exp(-d_{ij}/\lambda) \)

Mean reward surface

Inference Algorithm: Kalman Filter

- $\phi_t = \text{indicator vector of the arm selected at time } t$
- $r_t = \text{reward obtained at time } t$

Posterior Precision: $\Lambda_t = \phi_t \phi_T + \Lambda_{t-1}$

Posterior Mean: $\mu_t = \phi_t \mu_T + \Lambda_{t-1} \mu_{t-1}$
The UCL Algorithm

Upper Credible Limit (UCL) Algorithm
- value of option $i$ at time $t$: $Q^i_t = \mu^i_t + \sigma^i_t \Phi^{-1}\left(1 - \frac{1}{K_t}\right)$

$\mu^i_t$ = posterior mean
$\sigma^i_t)^2$ = posterior variance
- pick option with maximum value $Q^i_t$ at each time

Stochastic UCL Algorithm and Human Decision-Making

Stochastic Upper Credible Limit (UCL) Algorithm:
- pick option $i$ with probability $\propto \exp(Q^i_t / \nu_t)$, $\nu_t = \nu / \log(t)$
- similar performance can be established

Stochastic UCL Algorithm

Cumulative Regret $\leq K_{\text{ucl}} \log T + o(\log T)$
Does stochastic UCL algorithm explain human experiment data?

Stochastic UCL as a model for human subjects

- Prior: $\mathcal{N}(\mu_0 1_N, \Sigma_0)$, $\Sigma_0_{ij} = \sigma_0^2 \exp(-d_{ij}/\lambda)$
- Model parameters: $(\mu_0, \lambda, \sigma_0, \nu)$

Importance is the correlation scale

- Data from Amazon Mechanical Turk bandit experiment
- $10 \times 10$ spatial grid of $N = 100$ options
- Given $T = 90$ trials: insufficient time to explore whole space
- Global mean reward $\approx 30$, Maximum mean reward $= 60$
Goal: Distributed algorithms that maximize total expected cumulative reward

The Distributed MAB Problem

- $N$ options with unknown mean rewards $m_i$
- $M$ decision-making agents with a connected communication graph
- Each agent can play only one option at a time
- Rewards corrupted by Gaussian noise $\mathcal{N}(0, \sigma_i^2)$
- No interference/collisions among agents

Goal: Distributed algorithms that maximize total expected cumulative reward

Anantharam et al. Asymptotically efficient allocation rules for the multiarmed bandit problem with multiple plays. Trans on Auto Cntrl, 1987
Anandkumar et al. Distributed algorithms for learning and cognitive medium access with logarithmic regret. J. Sel. Areas Comm., 2011
Goal:
Cooperative Decision-making in Multiarmed Bandit Problem

The Cooperative MAB Problem

Each agent receives reward corrupted by Gaussian noise
Each agent can play only one option at a time

\( M \) options with unknown mean rewards
agents communicate over a connected undirected graph

Decision Making in Multiarmed Bandit Problems

\( \mathcal{X} \): set of options
\( n_t \): number of selections of arm \( i \) at time \( t \)
\( \hat{\mu}_i(t) \): estimate of average reward from arm \( i \)

\( r_i(t) \): reward from arm \( i \)

\( g(t) \): function that controls exploration vs. exploitation

Cumulative Regret for all agents

\( \log(T) \)
**Distributed UCB Algorithm**

- **Initialization:** Agent $k$ selects each option once
- Agent $k$ estimates the value of option $i$

$$Q^k_t = \hat{\mu}^k_t + \sigma_k \sqrt{\frac{2 \log t}{\bar{N}^k_t(t)}}$$

- $g(t)$ saturates to 1 as $t \to \infty$
- Agent $k$ selects option with maximum $Q^k_t$

**Cumulative Regret for all agents** $\leq K_{d,ucb} \log T + O(1)$

---

**Putting Pieces Together: Shared Human-Robot Search**

- **A simple scheme for shared human-robot search**
  - Common decision-making framework facilitates systematic integration
  - Variable autonomy: Human expert takes control if needed
  - Learn efficient priors from human actions
  - Implement distributed algorithms with efficient prior

---

**Outline**

1. **Introduction**
2. **Stochastic Multi-armed Bandit Problems**
3. **Modeling Human Performance in Multi-armed Bandit Tasks**
   - Features of human decision-making
   - Upper Credible Limit algorithm
   - Data from experiments with human participants
4. **Distributed Decision-Making in Multi-armed Bandit Problems**
5. **Conclusions and Future Directions**
Conclusions

**Shared Autonomy**
- Machines’ memory and computation + human cognition

**MAB Framework for Human and Robot Decision-Making**
- A rich framework for search and surveillance problems
- Stochastic UCL algorithm captures human decision-making
- Allows to learn human expertise
- Distributed UCB for multiagent search and surveillance
- A common framework facilitates integration of shared human-robot system

Future Directions

**Cognitive Modeling and Shared Autonomy**
- Incorporating richer models of human decision making
- Models parameterized by task observables
- Methods for real-time learning of expertise
- Haptic interactions and shared manipulation
- Reinforcement learning
- Design of user interface for human-automata interaction
- Aerial Robotics and Human-Robot Interaction Lab

**Autonomous Systems**
- Distributed reinforcement learning
- Coverage control with explore-exploit conflict
- Explore-Exploit in SLAM problems
- Online learning for tracking and pursuit-evasion

Thanks for your attention!

Questions?