Human-Based Learning and Decision-Making

Angela Yu
University of California San Diego
Brain << Computer
Brain >> Computer

Vision

Social interactions
Overview

• **Questions**
  – What kind of problems is the brain good/bad at?
  – Implications for underlying computational principles?
Overview

• Questions
  – What kind of problems is the brain good/bad at?
  – Implications for underlying computational principles?

• Utility
  – Human-in-the-loop systems
  – Inexpensive/effective heuristics for AI systems
Overview

• Questions
  – What kind of problems is the brain good/bad at?
  – Implications for underlying computational principles?

• Utility
  – Human-in-the-loop systems
  – Inexpensive/effective heuristics for AI systems

• Approach
  – Bayesian inference + decision/control theory
  – Sensitivity to environmental statistics, context, task goals
  – Tractable approximations matching human behavior
  – Predictions for behavior in novel contexts, neurobiology
Project 1: Human Active Learning

Introduction

Bayesian Inference Model

Human Data

• reward rates fixed, iid from Beta(2, 2)

Learning component:

$S - \text{chooses option that maximizes the future cumulative}$

assuming next step being the last exploratory choice

Bayesian iterative inference assuming local patterns in

pulling one arm

Kickoff

$q F \rightarrow \text{selecting an arm to gain more information}$

at

$\tau \leftarrow \text{Pr}(t \cdot 1 \cdot k)$

$\cdot \Rightarrow \text{points this game 0}$

$\cdot \text{trial 1 of 15}$

$\cdot \text{game 1 of 20}$

The Optimal Algorithm

Fixed Belief Model

• computed via Bellman's dynamic programming principle

$\text{Pr}(M \cdot t \cdot 0 \cdot \tau)$

$\cdot q E \cdot D \cdot t - 0 \cdot \tau$\linebreak

$\cdot \leftarrow \text{if } (t + 1)$

$\cdot q E \cdot D \cdot t - 0 \cdot \tau$\linebreak

$\cdot \text{otherwise}$

$\cdot \text{first arm has produced one failure, and the second and third arms have both produced two}$

of each panel. In this example: seven trials have been completed, with the first, second

previous rewards, the ratio of successes to failures—if defined—is also shown at the top

References

Citations within the text should be numbered consecutively

Second level headings are lower case (except for first word an d others: proper nouns), flush left, bold and in

Side by side.

Corresponding address. The lead author's name is to be listed first
Multi-arm bandit problem

- Trial onset
- Choice: pulling one arm
- Outcome: success or failure to gain reward
- Feedback updated
- New trial onset

Points this game 0        Trial 1 of 15        Game 1 of 20
Points this game 0        Trial 1 of 15        Game 1 of 20
Points this game 1        Trial 1 of 15        Game 1 of 20
Choice: pulling one arm
Success!
Multi-arm bandit problem

- **Design:** 20 games, 15 trials each
Multi-arm bandit problem
- **Design:** 20 games, 15 trials each
- **Learning:** reward statistics
Multi-arm bandit problem
- **Design:** 20 games, 15 trials each
- **Learning:** reward statistics
- **Decision:** exploration vs. exploitation
Multi-arm bandit problem

- **Design:** 20 games, 15 trials each
- **Learning:** reward statistics
- **Decision:** exploration vs. exploitation
- **New approach:** trial-wise knowledge/choice, not average stationary stats (Stewart et al, 2010)
Multi-arm bandit problem
- **Design:** 20 games, 15 trials each
- **Learning:** reward statistics
- **Decision:** exploration vs. exploitation
- **New approach:** trial-wise knowledge/choice, not average stationary stats (Stewart et al, 2010)

**Results**
(Zhang & Yu, NIPS, 2013; see poster)
Multi-arm bandit problem
- **Design:** 20 games, 15 trials each
- **Learning:** reward statistics
- **Decision:** exploration vs. exploitation
- **New approach:** trial-wise knowledge/choice, not average stationary stats (Stewart et al, 2010)

**Results**
- **Semi-myopic planning** (KG) + **forgetful learning** (DBM) best match human behavior

(Zhang & Yu, NIPS, 2013; see poster)
Multi-arm bandit problem
- **Design:** 20 games, 15 trials each
- **Learning:** reward statistics
- **Decision:** exploration vs. exploitation
- **New approach:** trial-wise knowledge/choice, not average stationary stats (Stewart et al, 2010)

Results
- **Semi-myopic planning** (KG) + **forgetful learning** (DBM) best match human behavior
- Knowledge gradient (KG) closer to optimal compared to previously proposed heuristics

(Zhang & Yu, NIPS, 2013; see poster)
Multi-arm bandit problem

- **Design**: 20 games, 15 trials each
- **Learning**: reward statistics
- **Decision**: exploration vs. exploitation
- **New approach**: trial-wise knowledge/choice, not average stationary stats (Stewart et al, 2010)

Results

- **Semi-myopic planning** (KG) + **forgetful learning** (DBM) best match human behavior
- Knowledge gradient (KG) closer to optimal compared to previously proposed heuristics
- Exploration vs. exploitation sensitive to horizon

(Zhang & Yu, NIPS, 2013; see poster)
Project 1: Human Active Learning

Multi-arm bandit problem
- **Design:** 20 games, 15 trials each
- **Learning:** reward statistics
- **Decision:** exploration vs. exploitation
- **New approach:** trial-wise knowledge/choice, not average stationary stats (Stewart et al, 2010)

Results
- **Semi-myopic planning (KG) + forgetful learning (DBM)** best match human behavior
- Knowledge gradient (KG) closer to optimal compared to previously proposed heuristics
- Exploration vs. exploitation sensitive to horizon
- Exploration targeted rather than random (can predict up to 80% of human choices)

(Zhang & Yu, NIPS, 2013; see poster)
Project 1: Human Active Learning

Multi-arm bandit problem
- **Design**: 20 games, 15 trials each
- **Learning**: reward statistics
- **Decision**: exploration vs. exploitation
- **New approach**: trial-wise knowledge/choice, not average stationary stats (Stewart et al, 2010)

Results
- **Semi-myopic planning** (KG) + **forgetful learning** (DBM) best match human behavior
- Knowledge gradient (KG) closer to optimal compared to previously proposed heuristics
- Exploration vs. exploitation sensitive to horizon
- Exploration targeted rather than random (can predict up to 80% of human choices)

Related work

(Zhang & Yu, NIPS, 2013; see poster)
Multi-arm bandit problem

- **Design:** 20 games, 15 trials each
- **Learning:** reward statistics
- **Decision:** exploration vs. exploitation
- **New approach:** trial-wise knowledge/choice, not average stationary stats (Stewart et al, 2010)

**Results**

- **Semi-myopic planning** (KG) + **forgetful learning** (DBM) best match human behavior
- Knowledge gradient (KG) closer to optimal compared to previously proposed heuristics
- Exploration vs. exploitation sensitive to horizon
- Exploration targeted rather than random (can predict up to 80% of human choices)

**Related work**

- Extension to optimizing unknown **continuous** function (Bayesian optimization)
Project 1: Human Active Learning

Multi-arm bandit problem
- **Design**: 20 games, 15 trials each
- **Learning**: reward statistics
- **Decision**: exploration vs. exploitation
- **New approach**: trial-wise knowledge/choice, not average stationary stats (Stewart et al, 2010)

Results
- **Semi-myopic planning** (KG) + **forgetful learning** (DBM) best match human behavior
- Knowledge gradient (KG) closer to optimal compared to previously proposed heuristics
- Exploration vs. exploitation sensitive to horizon
- Exploration targeted rather than random (can predict up to 80% of human choices)

Related work
- Extension to optimizing unknown **continuous** function (Bayesian optimization)
- Cost-sensitivity: sampling cost, distance-dependent cost between successive queries

(Zhang & Yu, NIPS, 2013; see poster)
**Project 1: Human Active Learning**

**Multi-arm bandit problem**
- **Design:** 20 games, 15 trials each
- **Learning:** reward statistics
- **Decision:** exploration vs. exploitation
- **New approach:** trial-wise knowledge/choice, not average stationary stats (Stewart et al, 2010)

**Results**
- **Semi-myopic planning** (KG) + **forgetful learning** (DBM) best match human behavior
- Knowledge gradient (KG) closer to optimal compared to previously proposed heuristics
- Exploration vs. exploitation sensitive to horizon
- Exploration targeted rather than random (can predict up to 80% of human choices)

**Related work**
- Extension to optimizing unknown **continuous** function (Bayesian optimization)
- Cost-sensitivity: sampling cost, distance-dependent cost between successive queries
- KG outperforms EI
Project 2: Active Sensing

Active search of a visual target

- **Design**: non-uniform target distribution (1:3:9)
- **Learning**: spatial statistics of target
- **Decision**: sensing location & duration
- **Approach**: Bayesian inference + risk minimization
- **Related to**: sensor management (Hero & Cochran, 2011)
Active search of a visual target

- **Design:** non-uniform target distribution (1:3:9)
- **Learning:** spatial statistics of target
- **Decision:** sensing location & duration
- **Approach:** Bayesian inference + risk minimization
- **Related to:** sensor management (Hero & Cochran, 2011)

**Results**

(Ahmad & Yu, UAI, 2013; Ahmad, Huang, & Yu, NIPS 2013)
Project 2: Active Sensing

Active search of a visual target

- **Design**: non-uniform target distribution (1:3:9)
- **Learning**: spatial statistics of target
- **Decision**: sensing location & duration
- **Approach**: Bayesian inference + risk minimization
- **Related to**: sensor management (Hero & Cochran, 2011)

Results

(Ahmad & Yu, UAI, 2013; Ahmad, Huang, & Yu, NIPS 2013)

- Humans readily learn spatial statistics (forgetful Bayes)
Project 2: Active Sensing

Active search of a visual target

- **Design:** non-uniform target distribution (1:3:9)
- **Learning:** spatial statistics of target
- **Decision:** sensing location & duration
- **Approach:** Bayesian inference + risk minimization
- **Related to:** sensor management (Hero & Cochran, 2011)

Results

(Ahmad & Yu, UAI, 2013; Ahmad, Huang, & Yu, NIPS 2013)

- Humans readily learn spatial statistics (forgetful Bayes)
- Context-sensitive optimization (relative costs of time, accuracy, sensor repositioning) explains human eye movements better than pure information maximization
Active search of a visual target

- **Design**: non-uniform target distribution (1:3:9)
- **Learning**: spatial statistics of target
- **Decision**: sensing location & duration
- **Approach**: Bayesian inference + risk minimization
- **Related to**: sensor management (Hero & Cochran, 2011)

**Results**

(Ahmad & Yu, UAI, 2013; Ahmad, Huang, & Yu, NIPS 2013)

- Humans readily learn spatial statistics (forgetful Bayes)
- Context-sensitive optimization (relative costs of time, accuracy, sensor repositioning) explains human eye movements better than pure information maximization
- Model reproduces “confirmation bias” (alternative: belief quantization, w/ Ertin)
Project 2: Active Sensing

Active search of a visual target

- **Design:** non-uniform target distribution (1:3:9)
- **Learning:** spatial statistics of target
- **Decision:** sensing location & duration
- **Approach:** Bayesian inference + risk minimization
- **Related to:** sensor management (Hero & Cochran, 2011)

### Results

(Ahmad & Yu, UAI, 2013; Ahmad, Huang, & Yu, NIPS 2013)

- Humans readily learn spatial statistics (forgetful Bayes)
- Context-sensitive optimization (relative costs of time, accuracy, sensor repositioning) explains human eye movements better than pure information maximization
- Model reproduces “confirmation bias” (alternative: belief quantization, w/ Ertin)
- Different approximations to optimal (expensive) policy:
Project 2: Active Sensing

Active search of a visual target
- **Design**: non-uniform target distribution (1:3:9)
- **Learning**: spatial statistics of target
- **Decision**: sensing location & duration
- **Approach**: Bayesian inference + risk minimization
- **Related to**: sensor management (Hero & Cochran, 2011)

Results
(Ahmad & Yu, UAI, 2013; Ahmad, Huang, & Yu, NIPS 2013)
- Humans readily learn spatial statistics (forgetful Bayes)
- Context-sensitive optimization (relative costs of time, accuracy, sensor repositioning) explains human eye movements better than pure information maximization
- Model reproduces “confirmation bias” (alternative: belief quantization, w/ Ertin)
- Different approximations to optimal (expensive) policy:
  * Approximate Q-factors in dynamic programming
Project 2: Active Sensing

Active search of a visual target
- **Design**: non-uniform target distribution (1:3:9)
- **Learning**: spatial statistics of target
- **Decision**: sensing location & duration
- **Approach**: Bayesian inference + risk minimization
- **Related to**: sensor management (Hero & Cochran, 2011)

Results

(Ahmad & Yu, UAI, 2013; Ahmad, Huang, & Yu, NIPS 2013)

- Humans readily learn spatial statistics (forgetful Bayes)
- Context-sensitive optimization (relative costs of time, accuracy, sensor repositioning) explains human eye movements better than pure information maximization
- Model reproduces “confirmation bias” (alternative: belief quantization, w/ Ertin)
- Different approximations to optimal (expensive) policy:
  * Approximate Q-factors in dynamic programming
  * Forego DP by adopting semi-myopic planning (KG)
Project 2: Active Sensing

Active search of a visual target
- **Design:** non-uniform target distribution (1:3:9)
- **Learning:** spatial statistics of target
- **Decision:** sensing location & duration
- **Approach:** Bayesian inference + risk minimization
- **Related to:** sensor management (Hero & Cochran, 2011)

Results

- Humans readily learn spatial statistics (forgetful Bayes)
- Context-sensitive optimization (relative costs of time, accuracy, sensor repositioning) explains human eye movements better than pure information maximization
- Model reproduces “confirmation bias” (alternative: belief quantization, w/ Ertin)
- Different approximations to optimal (expensive) policy:
  * Approximate Q-factors in dynamic programming
  * Forego DP by adopting semi-myopic planning (KG)
- Extensions to peripheral vision, ↑ fixation locations, ↑ hypotheses (target locations)
Project 3: Competitive Foraging

VOI

Patch A: 1 food item / minute

Patch B: 2 food items / minute

Patch C: 3 food items / minute
Project 3: Competitive Foraging

Matching in Competitive Foraging

- **Patch B**: 2 food items / minute
- **Patch C**: 3 food items / minute
- **Patch A**: 1 food item / minute
Project 3: Competitive Foraging

Matching in Competitive Foraging

- Observed in ducks, fish, humans, etc.
Project 3: Competitive Foraging

Matching in Competitive Foraging

- Observed in ducks, fish, humans, etc.
- Optimal resource allocation (Nash equilibrium)
Project 3: Competitive Foraging

Matching in Competitive Foraging

- Observed in ducks, fish, humans, etc.
- Optimal resource allocation (Nash equilibrium)
- Decentralized learning/planning + minimal social info
Project 3: Competitive Foraging

Matching in Competitive Foraging

– Observed in ducks, fish, humans, etc.
– Optimal resource allocation (Nash equilibrium)
– Decentralized learning/planning + minimal social info
– **Model:** Bayesian inference + long-term reward maximization
Project 3: Competitive Foraging

Matching in Competitive Foraging
– Observed in ducks, fish, humans, etc.
– Optimal resource allocation (Nash equilibrium)
– Decentralized learning/planning + minimal social info
– **Model:** Bayesian inference + long-term reward maximization
– **Human experiment:** reward distribution changes unpredictably over time
Project 3: Competitive Foraging

Matching in Competitive Foraging
- Observed in ducks, fish, humans, etc.
- Optimal resource allocation (Nash equilibrium)
- Decentralized learning/planning + minimal social info
- **Model**: Bayesian inference + long-term reward maximization
- **Human experiment**: reward distribution changes unpredictably over time
- **Key behavioral measures**: speed of adaptation to new reward structure, effects of different level of communication among agents (own choices & outcomes, others’ choices, others’ outcomes, others’ beliefs)
Matching in Competitive Foraging

- Observed in ducks, fish, humans, etc.
- Optimal resource allocation (Nash equilibrium)
- Decentralized learning/planning + minimal social info
- **Model**: Bayesian inference + long-term reward maximization
- **Human experiment**: reward distribution changes unpredictably over time
- **Key behavioral measures**: speed of adaptation to new reward structure, effects of different level of communication among agents (own choices & outcomes, others’ choices, others’ outcomes, others’ beliefs)
- **Novelty**: first “cognitive” model, theory of mind, trial-wise behavior
Matching in Competitive Foraging

- Observed in ducks, fish, humans, etc.
- Optimal resource allocation (Nash equilibrium)
- Decentralized learning/planning + minimal social info
- **Model:** Bayesian inference + long-term reward maximization
- **Human experiment:** reward distribution changes unpredictably over time
- **Key behavioral measures:** speed of adaptation to new reward structure, effects of different level of communication among agents (own choices & outcomes, others’ choices, others’ outcomes, others’ beliefs)
- **Novelty:** first “cognitive” model, theory of mind, trial-wise behavior
- Applications to **human-in-the-loop** (w/ Hero)
Multi-source change detection (Yu, NIPS 2007; Tsuchida & Yu, in prep; related to Dandach et al, 2010)

\[ x^1 \rightarrow \Phi_t \rightarrow x^2 \]
Other Related Work

**Multi-source change detection** (Yu, NIPS 2007; Tsuchida & Yu, in prep; related to Dandach et al, 2010)

\[ \Phi_t \]

\[ x^1 \]
\[ x^2 \]

**Contextual effects in preference choice** (Shenoy & Yu, 2013) & preference learning (Ahmad & Yu, in prep); related to Jordan & Hero
Other Related Work

**Multi-source change detection** (Yu, NIPS 2007; Tsuchida & Yu, in prep; related to Dandach et al, 2010)

![Diagram showing multi-source change detection](Image)

**Contextual effects in preference choice** (Shenoy & Yu, 2013) & **preference learning** (Ahmad & Yu, in prep); related to Jordan & Hero

![Graphs illustrating contextual effects](Image)
Multi-source change detection (Yu, NIPS 2007; Tsuchida & Yu, in prep; related to Dandach et al, 2010)

Contextual effects in preference choice (Shenoy & Yu, 2013) & preference learning (Ahmad & Yu, in prep); related to Jordan & Hero