Unsupervised Learning in Decision Making

Domagoj Fizulic Felix Gutmann

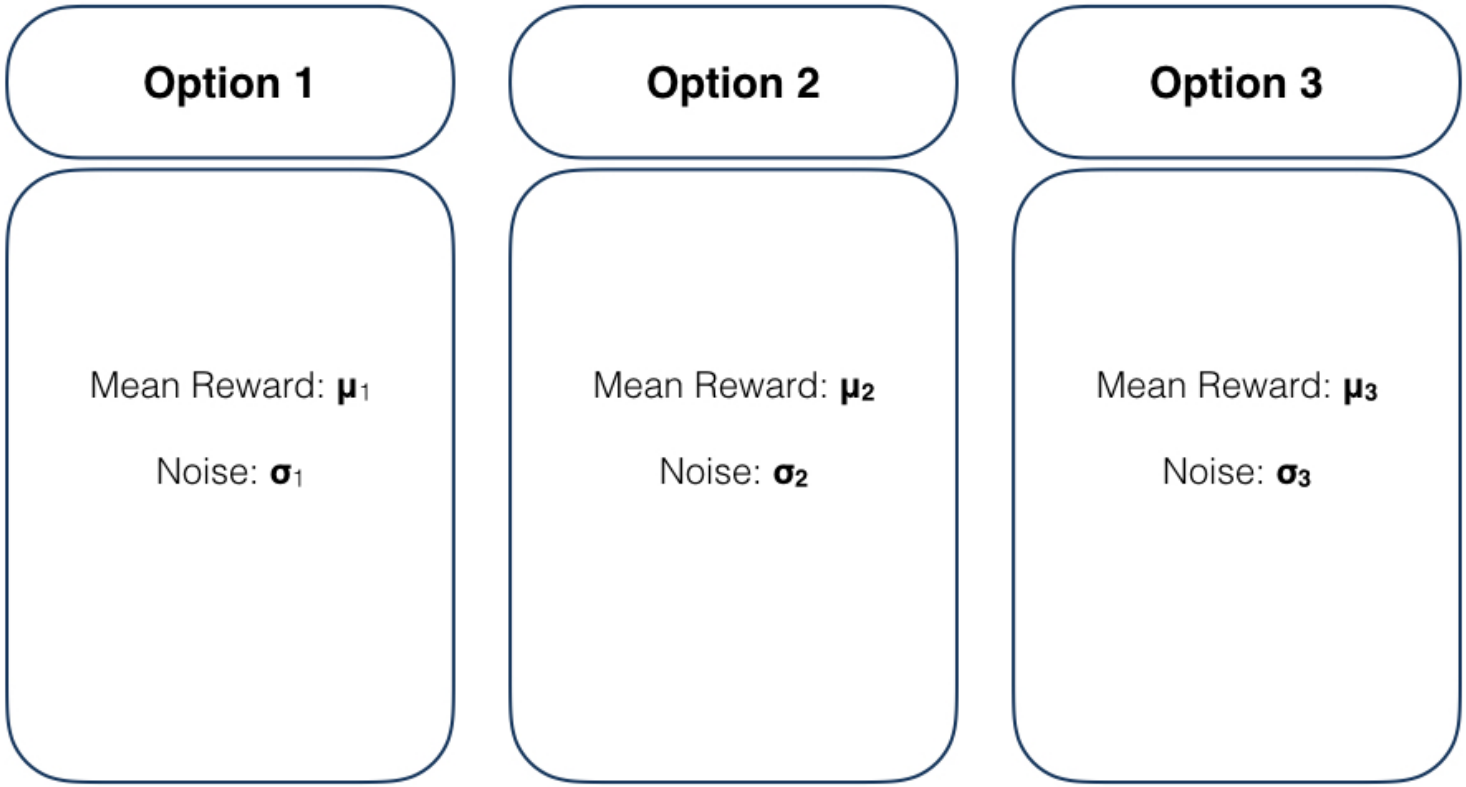
June 30, 2016

Introduction

- Introduction
- Reinforcement learning and simulation
- Experiment Data
- Experiment results

Reinforcement Learning

Multi arm bandit experiment



- Choose sequentially from set of choices
- Objective: Maximize revenues

Reinforcement Learning

Essential functions and agent modeling

Softmax Decision Function:

$$P(a)_{t+1} = \frac{e^{\frac{Q_t(a)}{\tau}}}{\sum_i^N e^{\frac{Q_t(i)}{\tau}}}$$

Update rule for value function of an action:

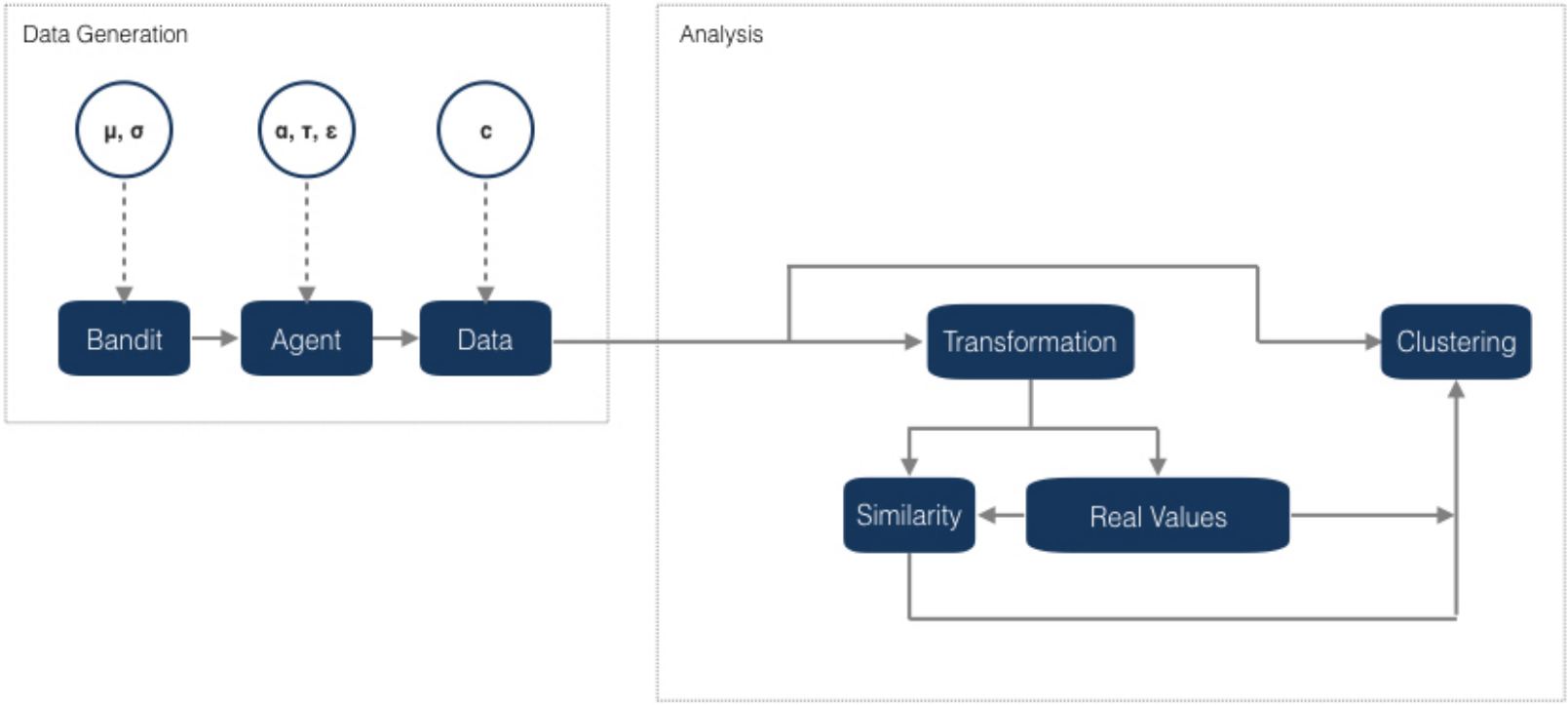
$$Q(a)_{t+1} = Q(a)_t + \alpha \left[R(a)_t - Q(a)_t\right]$$

where

- $Q(a)$ is the value function of action a
- $R(a)$ is the reward for action a
- τ is the *"temparature"*, controlling randomised behaviour
- α is the learning rate ($\alpha \in [0, 1]$)

Reinforcement Learning

Experiment design and simulation results



Key - findings:

- Clustering based on τ differences possible ($\Delta \approx 0.7$)
- Clustering based on α difficult. Good clustering only with very high differences ($\Delta \approx 0.99$)

Data Analysis

Analysed data sets

20-Arm Bandit Experiment:

- Four experiment types with different means and noises
- 429 participants

Data based on Iowa Gambling Task

- IGT data for 96 participants with 11 criminal profiles
- IGT for cocaine abusers

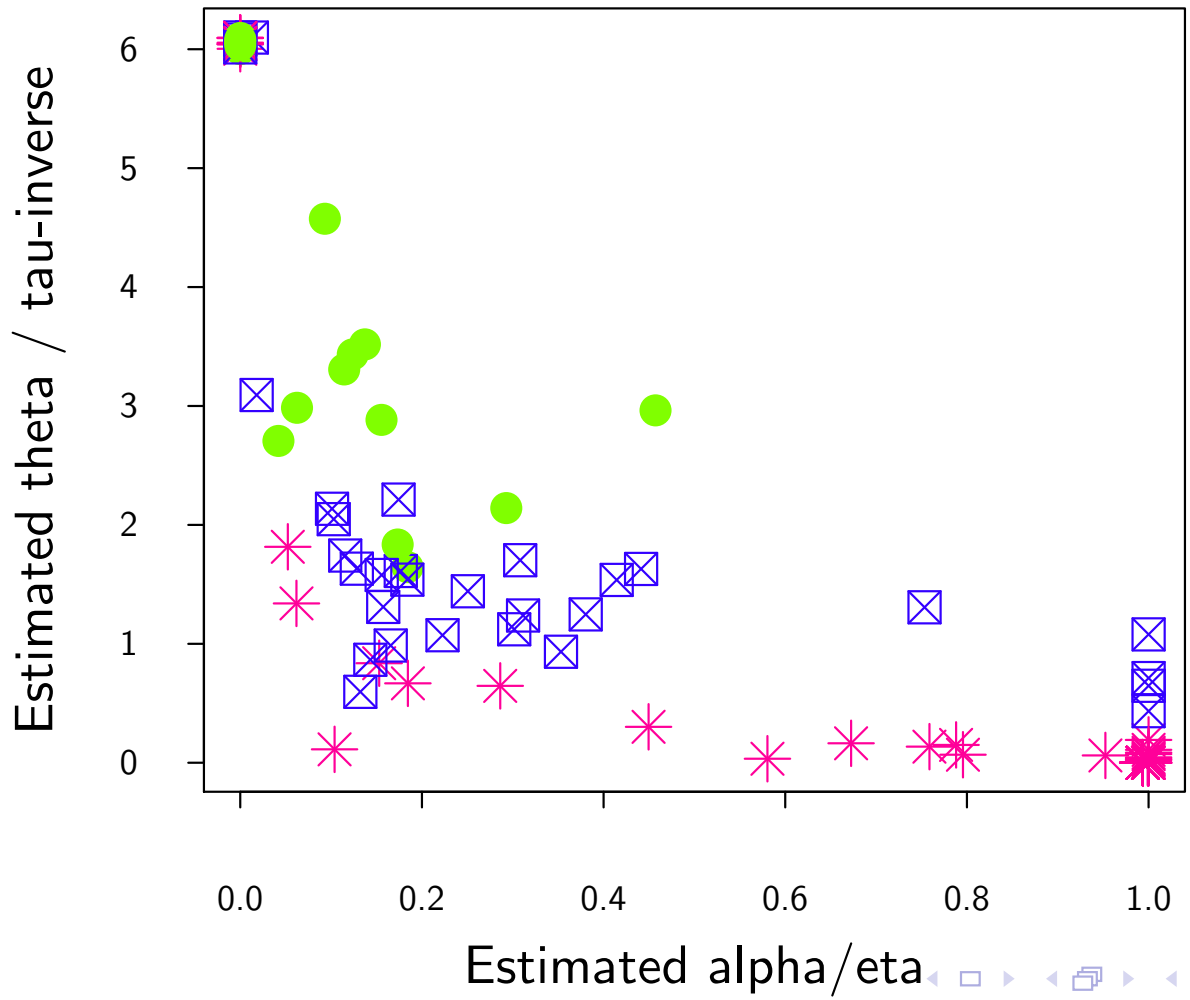
Experiment Results

Multi arm bandit experiments

- Use data to estimate clusters
- Estimated parameters from soft max equation using numerical optimization
- Try to see if unsupervised learning is recovering results from cognitive science

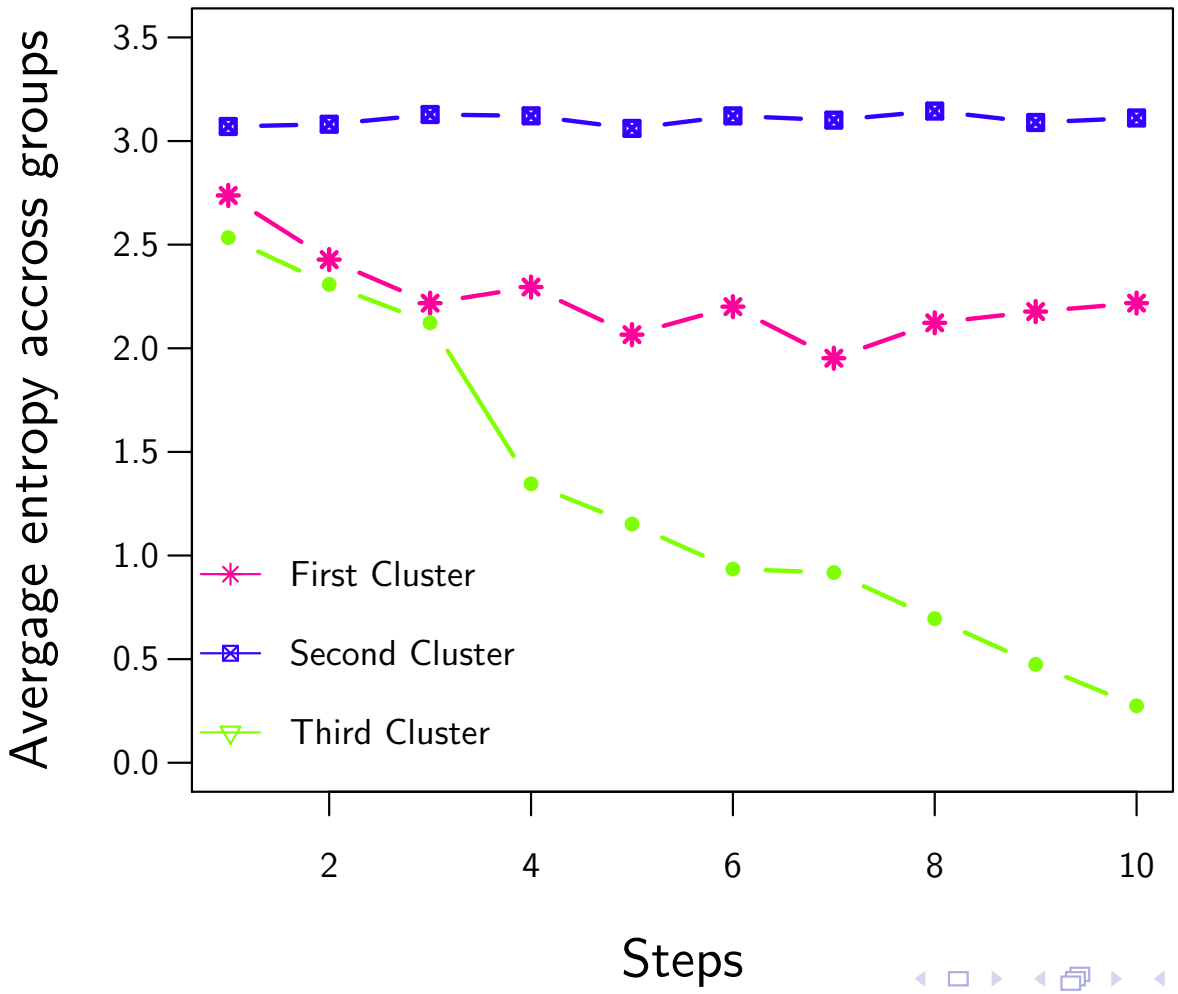
Experiment Results

Multi arm bandit experiments - 2 Clusters / Spectral RBF / Blockwise Entropy



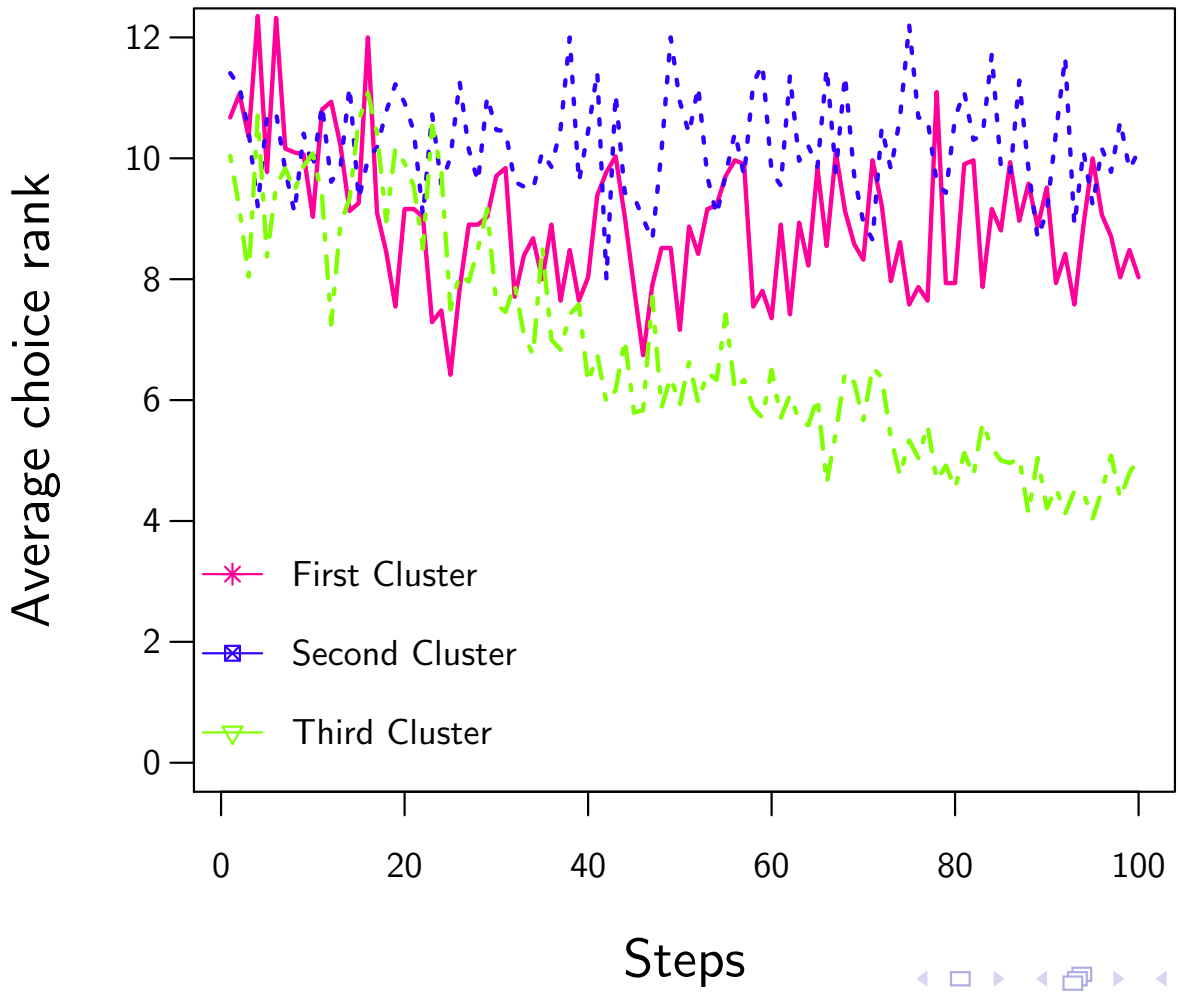
Experiment Results

Multi arm bandit experiments - 2 Clusters / Spectral RBF / Blockwise Entropy



Experiment Results

Multi arm bandit experiments - 2 Clusters / Spectral RBF / Blockwise Entropy



Experiment Results

IGT data for conviceted criminals

Authors results:

- Specialized version IGT test
- No control group
- EVM analysis shows clustering for convicted robbery and assault/murders; Other criminal groups have overlapping parameters

Our findings:

- Convicted assault/murders separate the strongest (highest clustering against forgery)
- Robbery only cluster moderately in our settings

Experiment Results

IGT data for drug abusers

Author's findings

- Cocaine abusers persistently do disadvantageous choices
- Effect still present after controlling for IQ (in general lower than within control group)

Our findings

- Only moderate clustering between control group and cocaine abusers
- Best separation criteria turned out disadvantageous behaviour

Conclusion

Key - Finings

- Clustering peoples choices in general difficult
- Algorithms can recover clustering if individuals show sufficient difference in their strategic behaviour