Modeling human learning and exploration in a temporal combinatorial bandit task

Guang-Yu Deng
Peking University, Beijing, China

Xi Guo
Peking University, Beijing, China

Fei Peng
Peking University, Beijing, China

Li Wang
Peking University, Beijing, China

Hang Zhang
Peking University, Beijing, China

Abstract

Life often presents choices that are not mutually exclusive, yet there has been insufficient research on human learning and directed exploration involved in combinatorial settings. We investigated human behavior in a four-armed combinatorial bandit (CB) task (N=107) where participants combined "nutrients" affecting required nurture time of virtual plants. Participants demonstrated effective learning but converged to suboptimal strategies, preferring combinations of one or two options. To model learning, two computational models were proposed and compared: a naïve extension of upper confidence bound (NaiveUCB), and a linear UCB model (LinUCB), both incorporating heuristic components. The NaiveUCB model with penalty for multiple selections, value decay, stickiness, and recency-based credit assignment best explained behavior, outperforming both LinUCB and simplified variants, suggesting that humans may navigate uncertainty through simple heuristics rather than sophisticated estimation. These findings extend our understanding of exploration and credit assignment in CB, and provide insight into daily decision making.

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*