

Human and Optimal Exploration and Exploitation in Bandit Problems

Shunan Zhang

University of California, Irvine

Michael Lee

University of California, Irvine

Miles Munro

University of California, Irvine

Abstract: We consider a class of bandit problems in which a decision-maker must choose between a set of alternatives each of which has a fixed but unknown rate of reward to maximize their total number of rewards over a short sequence of trials. Solving these problems requires balancing the need to search for highly-rewarding alternatives with the need to capitalize on those alternatives already known to be reasonably good. Consistent with this motivation, we develop a new model that relies on switching between latent exploration and exploitation states. We test the model over a range of two-alternative bandit problems, varying the number of trials, and the distribution of reward rates. By making inferences about the latent states from optimal decision-making behavior, we characterize how people should switch between exploration and exploitation. By making inferences from human data, we attempt to characterize how people actually do switch. We find some important regularities in and similarities optimal and human decision-making, but also some interesting individual variation. We discuss the implications of these findings for understanding and measuring the competing demands of exploration and exploitation in decision-making.