

Causal Bandits

Author’s Name

Editor: Editor’s name

Abstract

The abstract

Keywords: Causality, Bandits, Regret Bounds

1. Introduction

The aim of this paper is to demonstrate theoretically the utility of having causal knowledge when solving multi-armed bandit (MAB) problems.

We do this by:

1. Providing some motivating examples as to why it is reasonable to expect to have some knowledge of causal structures when solving a sequential decision making problem such as MAB problems.
2. Formalising what we mean by “causal information” for a simple class of stochastic multi-armed bandit (MAB) problems.
3. Proposing a simple explore-then-exploit algorithm for this class of problems that makes use of the causal information and proving it has a sublinear regret bound.
4. Deriving a matching lower bound for the same class of problems for algorithms that make use of the causal structure.
5. Showing that the upper bound for the “causally informed” algorithm is better than the corresponding lower bound for algorithms on this class of problems that do not use the extra causal information.

We identify a problem-dependent constant that appears in the upper and lower bounds that can be roughly interpreted as a measure of how much actions reveal about other actions via the causal structure.

1.1. Related Work

- Bareinboim, Forney & Pearl, *Bandits with unobserved confounders*, NIPS 2015
- Salomon, Audibert & Alaoui, *Lower Bounds and Selectivity of Weak-Consistent Policies in Stochastic Multi-Armed Bandit Problems*, JMLR 2013.
- Alon, Cesa-Bianchi, Dekel, Koren, *Online Learning with Feedback Graphs: Beyond Bandits*, COLT 2015.