CONTEXTUAL MULTI-ARMED BANDITS WITH COMMUNICATION CONSTRAINTS

Anonymous authors
Paper under double-blind review

ABSTRACT

We consider a remote Contextual Multi-Armed Bandit (CMAB) problem, in which the decision maker observes the context and the reward, but must communicate the actions to be taken by the agents over a rate-limited communication channel. This can model, for example, a personalized ad placement application, where the content owner observes the individual visitors to its website, and hence, has the context information, but must convey the ads that must be shown to each visitor to a separate entity that manages the marketing content. In this Rate-Constrained CMAB (RC-CMAB) problem, the constraint on the communication rate between the decision maker and the agents imposes a trade-off between the number of bits sent and the accuracy of the actions taken by the agents. We are particularly interested in the scenario in which the number of agents and the number of possible actions are large, while the communication budget is limited. Consequently, it can be considered as a policy compression problem, where the distortion metric is induced by the learning objectives. We first study the fundamental information theoretic limits of this problem by letting the number of agents go to infinity, while imposing an average per-agent bit constraint. Then, we propose a practical coding scheme, and provide numerical results.