The Boltzmann Exploration and Comparisons With Other Multi-Armed Bandit Algorithms

Turi-Kováts Bálint

Abstract

In this thesis, I discuss the multi-armed bandit problem and three of its algorithms. First I present multi-armed bandit algorithms as an alternative to A/B testing. I define the multi-armed bandit environment and introduce basic bandit definitions, such as regret and the σ-subgaussian property.

The first algorithm I look at is the epsilon-Greedy method. After defining and implementing it, I state and prove that the regret of epsilon-Greedy algorithms is $O(T)$ and I calculated the lower bound for any number of arms and ϵ parameter. Then I mention its disadvantages and present an alternative to the greedy approach: the Boltzmann-exploration. I prove that the regret of the Boltzmann-exploration is $\Omega(T)$ and I show that it is difficult to tune. The third and final algorithm is the Boltzmann-Gumbel exploration. This algorithm is defined and implemented as well, along with the Gumbel distribution. I estimate the regret of the algorithm as $O(\log^2(T))$ and prove this statement.

After discussing all three algorithms, I begin to compare them through simulations. With the Monte Carlo method these random events become comparable. First I compare epsilon-Greedy algorithms with different ϵ parameter, then the two Boltzmann-exploration algorithms, finally the Boltzmann-Gumbel exploration with epsilon-Greedy. I conclude that the Gumbel approach yields the best results, both in terms of regret and accuracy. Finally I discuss how bandit algorithms are applied in the industry, looking at a case study by Skyscanner, an online travel metasearch engine.