The benefit of general reward functions

We give an example in Section 4.3, which can’t be covered in existing models. This example is a network routing problem with latency. The latency of each edge in the network follows an exponential distribution where its mean has a linear relation with its features. We assume such a relation is a linear relation. And then regard each edge as open if the delay time is smaller than some tolerance. In this problem, the expected reward function can’t be represented by disjunctive objective or conjunctive objective, discussed in previous papers. We also did an experiment on such a general case and the result is shown in Figure 3(c).

Experimental results

When we demonstrate the experimental results, we focus on the comparison aspect to show the advantage of involving contextual information and position discounts. The curves increase in linear shape is because T is not big enough. We run on the setting of L= with T= and can get a concave shape of regret. We will replace all the experiments with a larger enough T.

Assumption on monotonicity

This assumption implies if the recommended list has higher reward mean, then the expected reward of the list should also get higher. It is a relatively basic assumption.

The V\_t norm in line 376 is defined in line 362-364. Due to the length limit, we didn’t bring the definition out. Sorry for the inconvenience.