Abstract

Motivated by applications in recommender systems, web search, social choice and crowdsourcing, we consider the problem of identifying the set of top K items from noisy pairwise comparisons. In our setting, we are non-actively given r pairwise comparisons between each pair of n items, where each comparison has noise constrained by a very general noise model called the strong stochastic transitivity (SST) model. We analyze the competitive ratio of algorithms for the top-K problem. In particular, we present a linear time algorithm for the top-K problem which has a competitive ratio of $\tilde{O}(\sqrt{n})$; i.e. to solve any instance of top-K, our algorithm needs at most $\tilde{O}(\sqrt{n})$ times as many samples needed as the best possible algorithm for that instance (in contrast, all previous known algorithms for the top-K problem have competitive ratios of $\tilde{\Omega}(n)$ or worse). We further show that this is tight: any algorithm for the top-K problem has competitive ratio at least $\tilde{\Omega}(\sqrt{n})$.