A Multiple-Choice Secretary Algorithm with Applications to Online Auctions

Robert Kleinberg*

Abstract

In the classical secretary problem, a set S of numbers is presented to an online algorithm in random order. At any time the algorithm may stop and choose the current element, and the goal is to maximize the probability of choosing the largest element in the set. We study a variation in which the algorithm is allowed to choose k elements, and the goal is to maximize their sum. We present an algorithm whose competitive ratio is $1-O(\sqrt{1/k})$. To our knowledge, this is the first algorithm whose competitive ratio approaches 1 as $k \to \infty$. As an application we solve an open problem in the theory of online auction mechanisms.

1 Introduction

Consider the following game between two parties, Alice and Bob. Alice writes down a set S consisting of n non-negative real numbers. Bob knows the value of n but otherwise has no information about S. The elements of S are revealed to Bob one by one, in random order. After each element is revealed, he may decide whether to select it or discard it. Bob is allowed up to k selections, and seeks to maximize their sum. If $k = 1$, then Bob may use the classical *secretary algorithm*: observe the first $\lfloor n/e \rfloor$ elements and note their maximum, then select the next element which exceeds this value if such an element appears. As is well known, this algorithm succeeds in selecting the maximum element of S with probability tending to $1/e$ as $n \to \infty$; therefore, its competitive ratio tends to $1/e$. For general k, what is the optimum competitive ratio? Does it approach 1 as $k \to \infty$ or remain bounded away from 1? Here, we answer this question by presenting an algorithm whose competitive ratio is $1 - O(\sqrt{1/k})$. (In fact, we have also derived a matching lower bound demonstrating that the competitive ratio of any algorithm is $1-\Omega(\sqrt{1/k})$. The lower bound is omitted from this paper for space reasons.)

This algorithm contributes to a rich body of work on secretary problems and related topics, beginning in the 1960's [2, 4] with the determination of the optimal stopping rule for the best-choice or "secretary" problem, i.e. the problem of stopping at the maximum element of a randomly-ordered sequence. In a long history of extensions and generalizations of this result, several "multiple choice secretary problems" — in which the algorithm is allowed to select $k > 1$ items, as in this paper — have been studied. In many of these problems, the algorithm's performance is judged according to a criterion which depends only on the ranks of the selected items and not their numerical values. For example, one may seek to select k items so as to maximize the probability that one of them is the best among all n items [3, 4], or to maximize the probability that the k chosen items are the k best ones [3], or to minimize the expected sum of the z-th powers of the ranks of the chosen items [1]. Another body of work (e.g. [4, 7, 10]) has considered multiple choice secretary problems in which the elements are real-valued samples from some probability distribution which is *known to the algorithm*. (The so-called "full information" case.) Such problems are also known as house-selling problems (e.g. [6, 9]), and generalizations have appeared under the name "dynamic and stochastic knapsack problems" [8]. Our problem combines aspects of both of these strands of research, in that the objective (maximizing the expected sum of the chosen items) depends not on the ranks of the items but on their numerical values, but the algorithm has no information about the distribution of these values other than the fact that it is invariant under permutations. To our knowledge, no algorithm existing in the literature attains a competitive ratio of $1 - o(1)$ for this problem without having knowledge of the distribution from which the values were sampled.

Apart from its appeal as a natural problem in optimal stopping theory, this question is motivated by an open problem from the theory of online auctions, i.e. auctions in which agents arrive and depart over time. Recent work by Hajiaghayi, Kleinberg, and Parkes [5] demonstrated a relation between online mechanism design and optimal stopping algorithms such as the secretary algorithm. They indicated a method for transforming certain online stopping rules into mechanisms which are *temporally strategyproof*, in the sense that a dominant strategy for each agent is to truthfully reveal its arrival and departure times as well as its valuation information. Thus, for example, the classical secretary algorithm leads to a temporally strategyproof (henceforth, simply "strategyproof") online single-item auction which is $(1/e)$-competitive with the efficiency of the optimal allocation, assuming that the agents arrive in random order but that they may have arbitrarily overlapping arrival and departure intervals. Using a similar technique, [5] obtains a strategyproof constant-competitive online auction mechanism for $k > 1$ identical items. The competitive ratio of this mechanism is bounded below by $1/48$ and above by $2/3$, leaving open the question of whether there exist mechanisms whose competitive ratio approaches 1 as $k \to \infty$. Herein, we answer this question affirmatively.

*MIT Department of Mathematics. Supported by a Fannie and John Hertz Foundation Fellowship. e-mail: rdk@csail.mit.edu

2 Algorithm

The algorithm is defined recursively as follows. If $k = 1$ then we use the classical secretary algorithm: observe the first $\lfloor n/e \rfloor$ elements, then select the next element which exceeds all prior observations, if such an element occurs. If $k > 1$ then let m be a random sample from the binomial distribution $B(n, 1/2)$ (i.e. the distribution of the number of heads in n tosses of a fair coin). Recursively apply the algorithm to select up to $\ell = \lfloor k/2 \rfloor$ elements among the first m samples. Let $y_1 > y_2 > \ldots > y_m$ be the first m samples, ordered from largest to smallest. After the m-th sample, select every element which exceeds y_ℓ, until we have selected k items or have seen all elements of S.

THEOREM 2.1. *Let S be any set of n non-negative real numbers, and let v be the sum of the k largest elements of S. The expected value of the elements selected by the algorithm is at least $(1 - 5/\sqrt{k})v$.*

Proof sketch. Let $T \subseteq S$ denote the k largest elements of S. Define the *modified value* of an element $x \in S$ to be equal to its value if $x \in T$ and zero otherwise; the modified value of a set is the sum of the modified values of its elements. We will prove that the expected modified value of the elements selected by the algorithm is at least $(1 - 5/\sqrt{k})v$. The proof is by induction on k, paralleling the recursive structure of the algorithm.

Let $y_1 > y_2 > \ldots > y_m$ be the first m samples, and let $z_1 > z_2 > \ldots > z_{n-m}$ be the remaining samples; denote these sets by Y and Z respectively. Since m has distribution $B(n, 1/2)$, it follows that Y is a sample from the uniform distribution on all 2^n subsets of S. In particular, $Y \cap T$ is a uniform random subset of T. This has the following consequences. First, the random variable $|Y \cap T|$ has the distribution $B(k, 1/2)$. Second, conditional on the event $|Y \cap T| = r$, the expected modified value of Y is $(r/k)v$. Third, the expected modified value of the top $\ell = k/2$ elements of Y is bounded below by

$$\sum_{r=1}^{k} \Pr(|Y \cap T| = r) \cdot (\min(r, \ell)/k)v \geq \left(1 - \frac{1}{2\sqrt{k}}\right) \frac{v}{2}.$$

The elements selected from Y have expected modified value at least $(1-5/\sqrt{k/2}) \cdot (1-1/2\sqrt{k}) \cdot (v/2)$, by the induction hypothesis.

Now we turn to estimating the modified value of the elements selected from Z. We first examine the random variable q which counts the number of elements of Z exceeding y_ℓ. Let q_i be the number of elements of Z whose value lies between y_i and y_{i-1}. The q_i are stochastically dominated by i.i.d. geometrically distributed random variables each having mean 1 and variance 2. Thus their sum $q = \sum_{i=1}^{\ell} q_i$ satisfies $\mathbf{E}(|q - \ell|) \leq \sqrt{k}$. Let $r = |q - \ell|$. An easy argument shows that the expected modified value of the elements of Z selected by the algorithm, conditional on r, is at least $(1/2 - r/k)v$. Removing the conditioning on r and recalling that $\mathbf{E}(r) \leq \sqrt{k}$, we find that the algorithm selects a subset of Z whose expected modified value is at least $(1/2 - \sqrt{1/k})v$.

Combining the estimates from the preceding two paragraphs, an easy computation verifies that the expected modified value of all elements selected by the algorithm is greater than $(1 - 5/\sqrt{k})v$, confirming the induction step.

3 Strategyproof mechanism

The algorithm presented above may be transformed into a strategyproof mechanism using a straightforward application of the ideas in [5]. The mechanism is specified as follows. For $k = 1$ we use the single-item online auction of [5]. For $k > 1$, we put $\ell = \lfloor k/2 \rfloor$ and choose a random transition time m as above. We recursively use the ℓ-item auction mechanism up until the arrival time of the m-th bidder. At that time, we set a threshold price equal to the $(\ell + 1)$-th highest bid yet seen, and allocate items to any bidders present in the system whose bid exceeds the threshold price, breaking ties in favor of earlier arrivals, but stopping when the total number of items sold reaches ℓ. At that point the threshold is raised to the ℓ-th highest bid yet seen, and subsequently items are allocated to any agents whose bid exceeds this price (including those who arrived before the transition time) until the supply of k items is exhausted. Agents who received an allocation are granted their item at the time of their departure, and their payment is the minimum of the threshold prices during the time they were present in the market. The proof of strategyproofness is based on Lemma 2 from [5] and closely resembles the proofs in that paper. The competitive ratio of $1 - O(\sqrt{1/k})$ is established using Theorem 2.1, along with an upper bound on the number of agents who are allocated an item at a transition time after their arrival.

References

[1] M. Ajtai, N. Megiddo, and O. Waarts. Improved algorithms and analysis for secretary problems and generalizations. *SIAM J. Discrete Math.*, 14(1):1–27, 2001.

[2] E. B. Dynkin. The optimum choice of the instant for stopping a markov process. *Sov. Math. Dokl.*, 4, 1963.

[3] P. Freeman. The secretary problem and its extensions: a review. *Internat. Statist. Rev.*, 51(2):189–206, 1983.

[4] J. P. Gilbert and F. Mosteller. Recognizing the maximum of a sequence. *J. Amer. Statist. Assoc.*, 61:35–73, 1966.

[5] M. T. Hajiaghayi, R. Kleinberg, and D. C. Parkes. Adaptive limited-supply online auctions. In *Proc. 5th ACM conference on Electronic commerce*, pages 71–80. ACM Press, 2004.

[6] S. Karlin. Stochastic models and optimal policy for selling an asset. *Studies in applied probability and management science*, pages 148–158, 1962.

[7] D. Kennedy. Prophet-type inequalities for multichoice optimal stopping. *Stoch. Proc. Appl.*, 24(1):77–88, 1987.

[8] A. J. Kleywegt and J. D. Papastavrou. The dynamic and stochastic knapsack problem. *Oper. Res.*, 46(1):17–35, 1998.

[9] M. Sakaguchi and V. Saario. A class of best-choice problems with full information. *Math. Japon.*, 41(2):389–398, 1995.

[10] W. Stadje. On multiple stopping rules. *Optimization*, 16(3):401–418, 1985.