Operations Research Letters 44 (2016) 180–185

Contents lists available at ScienceDirect

Operations Research Letters

journal homepage: www.elsevier.com/locate/orl

Maximizing expected utility over a knapsack constraint

Jiajin Yu[a], Shabbir Ahmed[b,*]

[a] College of Computing, Georgia Institute of Technology, Atlanta, GA 30332, United States
[b] School of Industrial & Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, United States

ARTICLE INFO

Article history:
Received 19 March 2015
Accepted 26 December 2015
Available online 13 January 2016

Keywords:
Knapsack problem
Utility maximization
Submodularity
Sample average approximation
Approximation algorithm

ABSTRACT

The expected utility knapsack problem is to pick a set of items with random values so as to maximize the expected utility of the total value of the items picked subject to a knapsack constraint. We devise an approximation algorithm for this problem by combining sample average approximation and greedy submodular maximization. Our main result is an algorithm that maximizes an increasing submodular function over a knapsack constraint with an approximation ratio better than the well known $(1-1/e)$ factor.

1. Introduction

This paper develops an approximation algorithm for the expected utility knapsack problem. Given a ground set of n items $\mathcal{U} = \{1, \ldots, n\}$; a random non-negative vector of values $\tilde{a}$ for the items; a positive integer vector b of weights for the items; a positive integer capacity of B; and a *utility function* $f : \mathbb{R}_+ \mapsto \mathbb{R}_+$; the expected utility knapsack problem is to pick a subset S of items so to

$$\max_{S \subseteq \mathcal{U}} \left\{ F(S) := \mathrm{E}\left[f(\tilde{a}(S))\right] \mid b(S) \le B \right\}, \tag{SP}$$

where $x(S) := \sum_{i \in S} x_i$. Note that the expectation above is with respect to the distribution of $\tilde{a}$. Throughout the paper, we assume $f(0) = 0$ and $f(a(S)) \ge 1$ for any $a \sim \tilde{a}$ and $S \ne \emptyset$. Therefore $F(\emptyset) = 0$ and $F(S) \ge 1$ for $S \ne \emptyset$. We further assume that the utility function f is strictly increasing and concave which corresponds to risk-averse preferences [17,10]. Commonly used utility functions such as log-utility $f(t) = \log t$, exponential utility $f(t) = 1 - e^{-\alpha t}$ for $\alpha > 0$, and power utility $f(t) = t^p$ for $0 < p < 1$, all satisfy this assumption.

Concavity of f along with the non-negativity of $\tilde{a}$ implies that the expected utility F is a submodular function of the selected set S (cf. [1]). Accordingly, (SP) is a submodular maximization problem with a knapsack constraint. It is well known that in general the approximation ratio for such problems is bounded by $(1-1/e)$ [6]. Moreover a variant of the greedy algorithm achieves this bound [13]. However these results assume a value oracle model where the underlying submodular function is general and can be evaluated exactly.

In (SP) evaluation of F requires evaluating a multidimensional integral over the distribution of $\tilde{a}$. Moreover, the distribution of $\tilde{a}$ may not be explicitly available, but only available through a sampling oracle. In such a setting, exact evaluation of F is impossible. In this paper we adopt the sample average approximation (SAA) framework [11] towards approximately evaluating F. In SAA the original distribution of the uncertain parameters is replaced by an empirical distribution by sampling a certain number of scenarios.

The sample average approximation of (SP) is

$$\max_{S \subseteq \mathcal{U}} \left\{ F^N(S) = \frac{1}{N} \sum_{i=1}^{N} f(a_i(S)) \mid b(S) \le B \right\} \tag{SA}$$

where $\{a_1, \ldots, a_N\}$ is an i.i.d sample of $\tilde{a}$. Note that F^N is a submodular function and (SA) is a deterministic knapsack constrained submodular maximization problem. It follows from classical SAA theory [11] that by solving (SA) corresponding to a sufficient number of samples N using an approximation algorithm of a given absolute error δ, with high probability, we can obtain a solution to the original problem (SP) whose absolute error is not too large compared to δ. Moreover the required sample size N is polynomial with respect to problem dimension.

If (SA) is solved using a *relative* error approximation algorithm (such as those in the submodular optimization literature) we need to adapt the SAA theory to recover a corresponding relative error for the true problem (SP). We make this adaptation. Further

* Corresponding author.
E-mail addresses: jiajinyu@gatech.edu (J. Yu), sahmed@isye.gatech.edu (S. Ahmed).

http://dx.doi.org/10.1016/j.orl.2015.12.016
0167-6377/

we develop an approximation algorithm for solving (SA) based on maximizing increasing submodular functions over a knapsack constraint. Specifically, the contribution of this paper is two-fold:

SAA analysis under relative error: We prove that with high probability only a polynomial number of samples is enough for an approximation algorithm that solves the SAA problem with relative error to give an approximate solution to the true stochastic problem of similar relative error. The works by Shmoys and Swamy [14], and Charikar et al. [3] are most relevant to our work as they both considered approximation algorithms with relative error for 2-stage stochastic optimization, rather than the absolute error usually considered in stochastic programming. However the polynomial sample size in their results depends on a ratio between the cost of the first stage and the cost of the second stage, which is not applicable to our single stage setting. A result similar to ours is mentioned in [15, pp. 21] but without detailed proof.

Increasing submodular maximization over a knapsack: The increasing and concavity properties of common utility functions imply that (SA) involves maximizing an increasing submodular function over a knapsack constraint. Sviridenko [13] recently developed a greedy algorithm to maximize an increasing submodular function over a knapsack constraint with approximation ratio $(1-1/e)$. We adapt this algorithm and its analysis exploiting the strict monotonicity of the utility function and show an approximation ratio *better than* the $(1-1/e)$ bound. For power utility functions, we explicitly characterize the approximation ratio as a function of the budget B and the exponent of the power function. Some other works that have improved on the $(1-1/e)$ bound are by Conforti and Cornuéjols [4], and Vondràk [16]. However these consider cardinality constraints and matroid constraints, respectively, and are not applicable in our knapsack setting.

We close this section with a brief discussion of some additional related literature. Li and Deshpande [8] study the problem of maximizing expected utility for various combinatorial optimization problems. They assume that the random coefficients are independent to simplify the expectation operation and use an approximation of the utility function. We allow more general distribution but are restricted to the knapsack setting. Klastorin [7] studies a similar problem but he assumes exact evaluation of the expectation objective and gives an algorithm that solves a continuous relaxation of the problem and then uses that in a branch-and-bound algorithm. Asadpour et al. [2] study maximizing a stochastic submodular function under matroid constraints. They assume exact evaluation of the expectation objective and do not consider increasing submodular functions. Mehrez and Sinuany-Stern [9] study a variation of the problem arising in resource allocation applications, but in their model the utility of items is separable which is different from our setting. Dean et al. [5] studied a variation where sizes of the elements are independent random variables while the values of the elements are deterministic. They devised both "nonadaptive" and "adaptive" policies to choose items in order to maximize the expected value of items that can be fit in the knapsack.

2. Sample average approximation

In this section, we adapt the classical SAA theory (cf. [12]) which corresponds to an absolute error setting to our required setting of relative error. We consider a generalization of (SP):

$$\max_{S\subseteq\mathcal{U}} \left\{ F(S) = \mathrm{E}\left[f(\tilde{a}, S)\right] \mid S \in X \right\}, \tag{SP0}$$

where X is the constraint set (e.g. knapsack constraint) and $f : 2^{\mathcal{U}} \mapsto \mathbb{R}_+$, parameterized by a, is a nonnegative set function. The sample average approximation of (SP0) is

$$\max_{S\subseteq\mathcal{U}} \left\{ F^N(S) = \frac{1}{N}\sum_{i=1}^{N} f(a_i, S) \mid S \in X \right\}, \tag{SA0}$$

where $\{a_1, \ldots, a_N\}$ is an i.i.d sample of $\tilde{a}$. Then (SP) is a special case of (SP0) and (SA) is a special case of (SA0). Let S^* be an optimal solution of (SP0). We make the following assumption on $f(a, S)$ and $\mathrm{E}\left[f(\tilde{a}, S)\right]$.

Assumption 1. For any $a \sim \tilde{a}, S \in X$, and $S \neq \emptyset$, we assume $f(a, S) \geq 1$. Therefore $F^N(S) \geq 1$ and $F(S) \geq 1$. For any $S \in X$, we also assume $\mathrm{E}\left[f(\tilde{a}, S)\right]$ is well-defined and finite, and $\mathrm{E}\left[e^{tf(\tilde{a},S)}\right]$ is finite in a neighborhood of $t = 0$.

Using the above assumption and standard Large Deviation analysis (cf. [11]), we can show that if N is large enough, for every $S \in X$, $F^N(S)$ is close to $F(S)$ in a relative sense.

Lemma 2. *Given $\gamma > 0$, let $\sigma^2 = \max\left\{\mathrm{Var}\left[f(\tilde{a}, S)\right] \mid S \in X\right\}$ and S^* be an optimal solution of the problem. If $N \geq \frac{2\sigma^2}{\epsilon^2}\log\frac{|X|}{\gamma}$, then*

$$\Pr\left\{\bigcap_{S\in X} |F(S) - F^N(S)| \leq \epsilon F(S^*)\right\} \geq 1 - 2\gamma. \tag{1}$$

Proof. Let $\{a_1, \ldots, a_N\}$ be the i.i.d sample defining $F^N(S)$. Let A_1 be the event that there exists a set S such that $F(S) - F^N(S) > \epsilon F(S^*)$, and let A_2 be the event that there exists a set S such that $F^N(S) - F(S) > \epsilon F(S^*)$. Let $\sigma_S^2 = \mathrm{Var}\left[f(\tilde{a}, S)\right]$, then $\sigma^2 = \max_{S\in X}\sigma_S^2$. If we can show that when $N \geq \frac{2\sigma^2}{\epsilon^2}\log\frac{|X|}{\gamma}$ we have $\Pr\{A_1\} \leq \gamma$ and $\Pr\{A_2\} \leq \gamma$, then we have the desired inequality (1).

Let us prove that $\Pr\{A_1\} \leq \gamma$.

$$\begin{aligned}\Pr&\left\{\bigcup_{S\in X} F(S) - F^N(S) > \epsilon F(S^*)\right\}\\ &\leq \sum_{S\in X}\Pr\left\{F(S) - F^N(S) > \epsilon F(S)\right\}\\ &= \sum_{S\in X}\Pr\left\{F^N(S) < (1-\epsilon)F(S)\right\}.\end{aligned}$$

By Assumption 1, we know that $F(S)$ is finite for every S and $\mathrm{E}\left[e^{tf(\tilde{a},S)}\right]$ is finite in a neighborhood of $t = 0$. So if ϵ is sufficiently small, by Large Deviation Theory (cf. [12, Sec 7.2.8]), we have

$$\Pr\left\{F^N(S) < (1-\epsilon)F(S)\right\} \leq \exp\left(-\frac{N(\epsilon F(S))^2}{2\sigma_S^2}\right).$$

Thus

$$\begin{aligned}\sum_{S\in X}\Pr\left\{F^N(S) < (1-\epsilon)F(S)\right\} &\leq \sum_{S\in X}\exp\left(-\frac{N(\epsilon F(S))^2}{2\sigma_S^2}\right)\\ &\leq |X|\exp\left(-\frac{N\epsilon^2}{2\sigma^2}\right)\\ &\leq \gamma\end{aligned}$$

which proves $\Pr\{A_1\} \leq \gamma$. The proof for $\Pr\{A_2\} \leq \gamma$ is identical, which we omit here. □

Equipped with the lemma above, we are ready to show that we can use any algorithm that solves (SA0) approximately to solve (SP0) without losing too much.

Theorem 3. *Given an algorithm that solves (SA0) with approximation ratio β, with probability at least $1 - 2\gamma$, we can use the same algorithm to solve the stochastic problem (SP0) with approximation ratio $\beta(1-\epsilon) - \epsilon$ by sampling $\tilde{a}$ at least $\frac{2\sigma^2}{\epsilon^2}\log\frac{|X|}{\gamma}$ times.*