Revisiting Modified Greedy Algorithm for Monotone Submodular Maximization with a Knapsack Constraint

Jing Tang
National University of Singapore
isejtang@nus.edu.sg

Xueyan Tang
Nanyang Technological University
asxytang@ntu.edu.sg

Andrew Lim
National University of Singapore
isealim@nus.edu.sg

Kai Han
University of Sci. and Tech. of China
hankai@ustc.edu.cn

Chongshou Li
National University of Singapore
iselc@nus.edu.sg

Junsong Yuan
SUNY Buffalo
jsyuan@buffalo.edu

ABSTRACT

Monotone submodular maximization with a knapsack constraint is NP-hard. Various approximation algorithms have been devised to address this optimization problem. In this paper, we revisit the widely known modified greedy algorithm. First, we show that this algorithm can achieve an approximation factor of 0.405, which significantly improves the known factors of 0.357 given by Wolsey [10] and $(1 - 1/e)/2 \approx 0.316$ given by Khuller et al. [4]. More importantly, our analysis closes a gap in Khuller et al.'s proof for the extensively mentioned approximation factor of $(1 - 1/\sqrt{e}) \approx 0.393$ in the literature to clarify a long-standing misconception on this issue. Second, we enhance the modified greedy algorithm to derive a data-dependent upper bound on the optimum, which enables us to obtain a data-dependent ratio typically much higher than 0.405 between the solution value of the modified greedy algorithm and the optimum. It can also be used to significantly improve the efficiency of algorithms such as branch and bound.

CCS CONCEPTS

• **Mathematics of computing** → Submodular optimization and polymatroids; • **Theory of computation** → Approximation algorithms analysis; Submodular optimization and polymatroids; Branch-and-bound; • **Computing methodologies** → Artificial intelligence.

KEYWORDS

Submodular; Greedy Algorithm; Approximation Guarantee

ACM Reference Format:

Jing Tang, Xueyan Tang, Andrew Lim, Kai Han, Chongshou Li, and Junsong Yuan. 2021. Revisiting Modified Greedy Algorithm for Monotone Submodular Maximization with a Knapsack Constraint. In *Abstract Proceedings of the 2021 ACM SIGMETRICS / International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS '21 Abstracts)*, June 14–18, 2021, Virtual Event, China. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3410220.3453925>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SIGMETRICS '21 Abstracts, June 14–18, 2021, Virtual Event, China

© 2021 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-8072-0/21/06.

<https://doi.org/10.1145/3410220.3453925>

1 INTRODUCTION

A set function $f: 2^V \rightarrow \mathbb{R}$ is *submodular* [7] if for all $S, T \subseteq V$, it holds that $f(S) + f(T) \geq f(S \cup T) + f(S \cap T)$. Alternatively, defining $f(v | S) := f(S \cup \{v\}) - f(S)$ as the marginal gain of adding an element $v \in V$ to a set $S \subseteq V$, an equivalent definition of a submodular set function f is that for all $S \subseteq T$ and $v \in V \setminus T$, $f(v | S) \geq f(v | T)$. The latter form of definition describes the concept of *diminishing return* in economics. The function f is *monotone* nondecreasing if and only if $f(S) \leq f(T)$ for all $S \subseteq T$ (or equivalently $f(v | S) \geq 0$). Monotone submodular maximization covers many well known combinatorial optimization problems, such as maximum coverage [2, 4] and maximum facility location [1], and a growing number of real-world applications in artificial intelligence, including sensor placement [5], viral marketing [3], etc.

In this paper, we study monotone submodular maximization with a knapsack constraint, which is defined as follows:

$$\max_{S \subseteq V} f(S) \text{ s.t. } c(S) \leq b, \quad (1)$$

where f is a monotone nondecreasing submodular set function¹, $c(S) = \sum_{v \in S} c(v)$ and $c(v)$ represents the cost of element v .

Since this optimization problem is NP-hard in general, various approximation algorithms have been proposed. For a special cardinality constraint where the costs of all elements are identical, i.e., $c(v) = 1$ for every element $v \in V$, Nemhauser et al. [7] proposed a simple hill-climbing greedy algorithm that can provide $(1 - 1/e)$ -approximation. For the general knapsack constraint, Wolsey [10] developed a modified greedy algorithm, referred to as MGREEDY, that can provide an approximation ratio of $(1 - 1/e^\beta) \approx 0.357$, where β is the unique root of the equation $e^x = 2 - x$. Khuller et al. [4] utilized MGREEDY to study the budgeted maximum coverage problem, and derived two approximation factors, i.e., $(1 - 1/e)/2 \approx 0.316$ and $(1 - 1/\sqrt{e}) \approx 0.393$. We note that the factor of $(1 - 1/\sqrt{e})$ is extensively mentioned in the literature, but unfortunately their proof was flawed as pointed out by Zhang et al. [11]. It becomes an open question whether MGREEDY can achieve an approximation ratio at least $(1 - 1/\sqrt{e})$.

Khuller et al. [4] also developed a partial enumeration greedy algorithm that improves the approximation factor to $(1 - 1/e)$, which was later shown to be applicable to the general problem (1) [8]. However, this algorithm requires $O(n^5)$ (where $n = |V|$ is the total number of elements in the ground set V) function value computations, which is not scalable. We focus on the scalable MGREEDY

¹We assume that f is normalized, i.e., $f(\emptyset) = 0$, and is given via a value oracle.

algorithm of $O(n^2)$ and conduct a comprehensive analysis on its worst-case approximation guarantee.

Main Contributions. We establish an approximation ratio of 0.405 for MGREEDY, which improves the factors of $(1 - 1/e^\beta) \approx 0.357$ given by Wolsey [10] and $(1 - 1/e)/2 \approx 0.316$ given by Khuller et al. [4]. More importantly, our analysis fills a critical gap in the proof for the factor of $(1 - 1/\sqrt{e}) \approx 0.393$ given by Khuller et al. [4] to clarify a long-standing misconception in the literature.

In addition, we enhance the MGREEDY algorithm to derive a data-dependent upper bound on the optimum. The bound enables us to obtain a data-dependent ratio typically much higher than 0.405 between the solution value of MGREEDY and the optimum. It can also be used to significantly improve the efficiency of algorithms such as branch and bound.

Interested readers are referred to our full paper [9] for more details, including the formal proofs of all theoretical results and some experimental evaluations.

2 MODIFIED GREEDY ALGORITHM AND APPROXIMATION GUARANTEES

It is natural to apply a cost-effective greedy algorithm by picking in each iteration the element that maximizes the ratio $\frac{f(v|S_g)}{c(v)}$ based on the selected element set S_g . Unfortunately, this simple greedy algorithm has an unbounded approximation factor [10]. Interestingly, a small modification to the greedy algorithm, namely MGREEDY, achieves a constant approximation factor [4, 10]. Specifically, in addition to S_g obtained from the greedy heuristic, the algorithm also finds an element v^* that maximizes $f(\{v\})$, and then chooses the better one between S_g and $\{v^*\}$. A main result of this paper is to establish an improved approximation factor of 0.405 for MGREEDY.

THEOREM 2.1. *Let $\alpha^\perp$ be the root of $(1 - \alpha^\perp) \ln(1 - \alpha^\perp) + (2 - 1/e)(1 - 2\alpha^\perp) = 0$ satisfying $\alpha^\perp > 0.405$. The MGREEDY algorithm achieves an approximation factor of $\alpha^\perp$.*

To our knowledge, this is the first work giving a constant factor achieved by MGREEDY that is even larger than $(1 - 1/\sqrt{e}) \approx 0.393$, which not only significantly improves the known factors of $(1 - 1/e^\beta) \approx 0.357$ given by Wolsey [10] and $(1 - 1/e)/2 \approx 0.316$ given by Khuller et al. [4] but also clarifies a long-term misunderstanding regarding the factor of $(1 - 1/\sqrt{e})$ in the literature.

3 DATA-DEPENDENT UPPER BOUND

The constant approximation factor 0.405 established above gives a lower bound on the worst-case solution quality over all problem instances. In this section, we enhance MGREEDY to derive a data-dependent upper bound on the optimum. The upper bound allows us to obtain a potentially tighter data-dependent ratio between the solution value of MGREEDY and the optimum for individual problem instances. Specifically, given a set S , let $v_1, v_2, \dots$ be the sequence of elements in $V \setminus S$ in the descending order of $\frac{f(v|S)}{c(v)}$. Let r be the lowest index such that the total cost of the elements $\{v_1, v_2, \dots, v_r\}$ is larger than b , i.e., $c^* := \sum_{i=1}^{r-1} c(v_i) \leq b$ and $c^* + c(v_r) > b$. Define $\Delta(b | S) := \sum_{i=1}^{r-1} f(v_i | S) + f(v_r | S) \cdot \frac{b - c^*}{c(v_r)}$, which is an upper bound on the largest marginal gain on top of S subject to the budget b . We choose the smallest upper bound Λ over all the intermediate

Figure 1: Relationship among $f(S_m)$, Λ , and $f(OPT)$.

sets constructed by the greedy heuristic, i.e., $\Lambda := \min_i \{f(S_i) + \Delta(b | S_i)\}$, where S_i contains the first i elements added to S_g by the greedy heuristic. We show that Λ is a good upper bound of $f(OPT)$.

THEOREM 3.1. *Let α' be the root of $(1 - \alpha') \cdot (\ln(1 - \alpha') + 2) - 1 = 0$ satisfying $\alpha' > 0.357$. We have $f(OPT) \leq \Lambda \leq \frac{f(S_m)}{\alpha'} \leq \frac{f(OPT)}{\alpha'}$.*

Figure 1 depicts the relationship among $f(S_m)$, Λ , and $f(OPT)$. The data-dependent ratio of $f(S_m)$ to Λ is guaranteed to be larger than 0.357 for any problem instance, which matches the best known factor given by Wolsey [10] and is usually much larger than 0.357 or 0.405 in practice [9]. In addition, when the knapsack constraint degenerates to a cardinality constraint, i.e., $\max_{S \subseteq V} f(S)$ s.t. $|S| \leq k$, our upper bound Λ is guaranteed to be smaller than $\frac{f(S_g)}{1 - 1/e}$, which matches the tight approximation factor of $(1 - 1/e)$ [6].

THEOREM 3.2. *For monotone submodular maximization with a cardinality constraint, we have $f(OPT) \leq \Lambda \leq \frac{f(S_g)}{1 - 1/e} \leq \frac{f(OPT)}{1 - 1/e}$.*

4 CONCLUSION

In this paper, we show that MGREEDY can achieve an approximation factor of 0.405 for monotone submodular maximization with a knapsack constraint. This factor not only significantly improves the known factors of 0.357 and $(1 - 1/e)/2 \approx 0.316$ but also closes a critical gap on the misunderstood factor of $(1 - 1/\sqrt{e}) \approx 0.393$ in the literature. We also derive a data-dependent upper bound on the optimum that is guaranteed to be smaller than a multiplicative factor of $\frac{1}{0.357}$ to the solution value obtained by MGREEDY. It remains an open question to study whether the approximation factor of 0.405 for MGREEDY is completely tight.

REFERENCES

- [1] Gerard Cornuejols, Marshall L Fisher, and George L Nemhauser. 1977. Location of Bank Accounts to Optimize Float: An Analytic Study of Exact and Approximate Algorithms. *Management Science* 23 (1977), 789–810.
- [2] Uriel Feige. 1998. A Threshold of $\ln n$ for Approximating Set Cover. *J. ACM* 45, 4 (1998), 634–652.
- [3] David Kempe, Jon Kleinberg, and Éva Tardos. 2003. Maximizing the Spread of Influence Through a Social Network. In *Proc. ACM KDD*. 137–146.
- [4] Samir Khuller, Anna Moss, and Joseph Naor. 1999. The Budgeted Maximum Coverage Problem. *Inform. Process. Lett.* 70, 1 (1999), 39–45.
- [5] Andreas Krause, Ajit Singh, and Carlos Guestrin. 2008. Near-Optimal Sensor Placements in Gaussian Processes: Theory, Efficient Algorithms and Empirical Studies. *Journal of Machine Learning Research* 9, 3 (2008), 235–284.
- [6] George L. Nemhauser and Laurence A. Wolsey. 1978. Best Algorithms for Approximating the Maximum of a Submodular Set Function. *Mathematics of Operations Research* 3, 3 (1978), 177–188.
- [7] George L. Nemhauser, Laurence A. Wolsey, and Marshall L. Fisher. 1978. An Analysis of Approximations for Maximizing Submodular Set Functions-I. *Mathematical Programming* 14, 1 (1978), 265–294.
- [8] Maxim Sviridenko. 2004. A Note on Maximizing a Submodular Set Function Subject to a Knapsack Constraint. *Operations Research Letters* 32, 1 (2004), 41–43.
- [9] Jing Tang, Xueyan Tang, Andrew Lim, Kai Han, Chongshou Li, and Junsong Yuan. 2021. Revisiting Modified Greedy Algorithm for Monotone Submodular Maximization with a Knapsack Constraint. *Proc. ACM Meas. Anal. Comput. Syst.* 5, 1, Article 8 (2021), 22 pages.
- [10] Laurence A. Wolsey. 1982. Maximising Real-Valued Submodular Functions: Primal and Dual Heuristics for Location Problems. *Mathematics of Operations Research* 7, 3 (1982), 410–425.
- [11] Ping Zhang, Zhifeng Bao, Yuchen Li, Guoliang Li, Yipeng Zhang, and Zhiyong Peng. 2018. Trajectory-driven Influential Billboard Placement. In *Proc. ACM KDD*. 2748–2757.