A Two-Stage Constrained Submodular Maximization

Ruiqi Yang[1], Shuyang Gu[2], Chuangen Gao[3], Weili Wu[2], Hua Wang[3], and Dachuan Xu[1(✉)]

[1] Department of Operations Research and Scientific Computing, Beijing University of Technology, Beijing 100124, People's Republic of China
yangruiqi@emails.bjut.edu.cn, xudc@bjut.edu.cn

[2] Department of Computer Science, University of Texas at Dallas, Richardson, TX 75080, USA
{Shuyang.Gu,weiliwu}@utdallas.edu

[3] School of Computer Science and Technology, Shandong University, Jinan 250101, People's Republic of China
gaochuangen@gmail.com, wanghua@sdu.edu.cn

Abstract. We consider a two-stage submodular maximization under p-matroid (or p-extendible) constraints. In the model, we are given a collection of submodular functions and some p-matroid (or extendible) system constraints for each of these functions, one need to choose a representative set with a cardinality constraint and simultaneously select a series of subsets that are restricted to the representative set for all functions, the aim is to maximize the average of the summarization of these function values. We extend the two-stage submodular maximization under single matroid to handle p-matroid (or p-extendible) constraints, and derive constant approximation ratio algorithms for the two problems, respectively. In the end, we empirically demonstrate the efficiency of our method on some datasets.

Keywords: Submodular maximization · Approximation algorithms · Independence system constraints

1 Introduction

The submodular maximization has many applications, such as document summarization [5,11], recommender systems [13,14,17], and other applications [9,19], etc. Formally, it can be modeled as $\max_{S\subseteq\Omega:S\in\mathcal{I}} f(S)$, where f is a submodular function defined on ground set Ω and $\mathcal{I}$ is some specific constraint. In the text, we will give a brief summary of the submodular maximization.

For the submodular maximization under a cardinality constraint, [15] provided a $(1-1/e)$-approximation. Under more general p-matroid constraint, [7] got a deterministic $1/(p+1)$-approximation algorithm. In particular, if $p=1$, it reduced to a 1/2-approximation algorithm for monotone submodular maximization under a single matroid constraint (SMMC). [10] improved the ratio from

D.-Z. Du et al. (Eds.): AAIM 2019, LNCS 11640, pp. 329–340, 2019.
https://doi.org/10.1007/978-3-030-27195-4_30

$1/(p+1)$ to $1/(p+\epsilon)$ for monotone submodular maximization under p-matroid constraints (SMMC-p). Combining continuous greedy process and pipage rounding technique, [4] obtained a random $(1-1/e)$-approximation algorithm for the monotone SMMC. As the hardness of approximation ratio of the above models is $(1-1/e+\epsilon)$, it is still a long history of closing the gap of approximation ratio for the monotone SMMC by a deterministic algorithm. The breakthrough result was presented by [3], who gave the first deterministic 0.5008-approximation algorithm for the monotone SMMC-p. [12] introduced a more general p-extendible system constraints, which captures a class of constraints, such as p-matroid constraints, b-matching, maximum profit scheduling and maximum asymmetric traveling salesman problem. He presented a $1/p$-approximation algorithm based on greedy for a monotone submodular maximization under p-extendible constraint. For a non-monotone submodular maximization under p-extendible system constraint, [6] presented a $p/(p+1)^2$-approximation algorithm.

Motivated by the tasks of multi-objective summarization, [2] introduced the *two-stage submodular maximization problem*. In the model, we are given a ground set Ω of size n, integers ℓ, k, and multiple submodular functions $f_1, ..., f_m$, the goal is to choose a subset S of $|S| \leq \ell$ such that the sum of $\max_{T_i \subseteq S, |T_i| \leq k} f_i(T_i)$ is maximum. Obviously, this problem reduces to classical cardinality submodular maximization problem if $m=1$, or if $\ell = k$. Combining the techniques of continuous greedy and dependent rounding, they firstly presented an approximation arbitrary close to $1-1/e$ as $k \to \infty$. Secondly, under the case of that each f_i, $i \in [m]$ is a coverage function, they obtained a $1/2(1-1/e)$-approximation by a local search, while the query complexity is bounded by $O(km\ell n^2 \log n)$. [17] considered a the two-stage submodular maximization with general matroid constraint, that is, $\sum_{i=1}^{m} \max_{T_i \in \mathcal{I}(S)} f_i(T_i)$, where $(S, \mathcal{I}(S))$ is a matroid. They derived a $1/2(1-1/e^2)$-approximation algorithm, and its query complexity is at most $O(rm\ell n)$, where r is the matroid rank. [14] first studied this problem under streaming and distributed settings, respectively. In the streaming setting, they derived a one pass, $1/7$-approximation algorithm, while its memory complexity is bounded by $O(\ell \log(\ell)/\epsilon)$ and the query complexity is at most $O(kmn \log(\ell)/\epsilon)$. In the distributed setting, they got two $1/4(1-1/e^2)$ and 0.107-approximation algorithms, respectively. The query complexities of the above two distributed algorithms are bounded by $O(kmn\ell/M + Mkm\ell^2)$ and $O(kmn \log \ell/M + Mkm\ell^2 \log \ell)$, respectively, where M is the number of the machines. We first consider the two-stage submodular maximization under more general constraints. Specifically, we aim to maximize $\sum_{i=1}^{m} \max_{T_i \in \mathcal{I}^i(S)} f_i(T_i)$, where $(S, \mathcal{I}^i(S))$ is a p-matriod (or p-extendible) for any i. For the two-stage submodular maximization under p-matroid system constraint, we propose a $1/(p+1)(1-1/e^2)$-approximation algorithm, while its query complexity is bounded by $O(\ell mnr^p)$, where r is the maximum independence set size. Under more general p-extendible constraints, we yield a $1/(r+1)(1-1/e^2)$-approximation algorithm with the same query complexity. Finally, we demonstrate the efficiency of our algorithm on some datasets.

The rest of our paper is organized as follows. We present some necessary preliminaries in Sect. 2. In Sect. 3, we introduce the two-stage submodular maximization under p-matroid system constraints and provide a $1/(p+1)(1-1/e^2)$-approximation algorithm. In addition, we present a $1/(r+1)(1-1/e^2)$-approximation algorithm for two-stage submodular maximization under p-extendible system constraints in Sect. 4. In Sect. 5, we show the results of some numerical experiments of our algorithm. Finally, we give a conclusion in Sect. 6.

2 Preliminaries

In our setting, we are given an element ground set Ω of size n, and a collection $\mathcal{F} = \{f_1, ..., f_m\}$ of non-negative monotone submodular functions that are defined on the ground set Ω. For any $i \in [m] = \{1, ..., m\}$, $f_i : 2^{\Omega} \rightarrow R_+$ is a *submoduar function*, i.e.,

$$f_i(A) + f_i(B) \geq f_i(A \cup B) + f_i(A \cap B), \forall A, B \subseteq \Omega.$$

For any $i \in [m]$, there exists a constraint $\mathcal{I}^i$, the objective is to find a representative set $S \subseteq \Omega$ with $|S| \leq \ell(\ll n)$, such that the average of the summarization of the optimum of $f_i, i \in [m]$ restricted to S is maximum. Let $G_m(S) = \frac{1}{m}\sum_{i=1}^{m} \max_{T_i \in \mathcal{I}^i(S)} f_i(T_i)$. Then our model can be defined as

$$\max_{S \subseteq \Omega, |S| \leq \ell} G_m(S) = \max_{S \subseteq \Omega, |S| \leq \ell} \frac{1}{m} \sum_{i=1}^{m} \max_{T_i \in \mathcal{I}^i(S)} f_i(T_i), \tag{1}$$

where $\mathcal{I}^i(S)$ denotes the constraint $\mathcal{I}^i$ restricted to S for any $i \in [m]$.

In order to have a better understand of our model and constraints, we restate some necessary notations and definitions as follows. Given a finite element set Ω, and a collection $\mathcal{I}$ of subsets of Ω. A two-tuples $M = (\Omega, \mathcal{I})$ is defined as an *independent system*, if it has for any subset $T \in \mathcal{I}$, then any subset $S \subseteq T$ such that $S \in \mathcal{I}$. Each subset of $\mathcal{I}$ is named as *independence set*. The independent system $M = (\Omega, \mathcal{I})$ is a *matroid* if it also satisfies that if for any independence sets $S, T \in \mathcal{I}$ with $|S| > |T|$, then there exists an element $e \in S \setminus T$ such that $T \cup \{e\} \in \mathcal{I}$. A maximal independent subset $A \in \mathcal{I}$ is called a *base* of the independent system $M = (\Omega, \mathcal{I})$. Given an integer p, Let $M_j = (\Omega, \mathcal{I}_j)$ be a matroid according to $j \in [p]$, then we call the intersection of these p matroids $(\Omega, \cap_{j=1}^{p} \mathcal{I}_j)$ as *p-matroid*. Given any subsets $S, T \in \mathcal{I}$, we say T is an *extension* of S if $S \subseteq T$. We restate the definition of p-extendible system as follows.

Definition 1 [6,12]. *An independence system $M = (\Omega, \mathcal{I})$ is p-extendible system if for every independent set $S \in \mathcal{I}$, an extension T of S and an element $e \notin S$ obeying $S \cup \{e\} \in \mathcal{I}$ there must exist a subset $Y \subseteq T \setminus S$ with $|Y| \leq p$ such that $T \setminus Y \cup \{e\} \in \mathcal{I}$. Specially, if the independent set S is maximal i.e., $T = S$, then we can reduce the definition by setting $Y = \emptyset$.*

In our p-matroid constraints model, for $i \in [m]$ and $S \subseteq \Omega$, any subset $T_i \subseteq S$ is feasible if $T_i \in \mathcal{I}^i = \cap_{j=1}^{p} \mathcal{I}_j^i$, where $M^i = (S, \cap_{j=1}^{p} \mathcal{I}_j^i(S))$ is a p-matroid system restricted to S. Similarly, for the p-extendible system constraint, $M^i = (S, \mathcal{I}^i)$ is defined as p-extendible system. We say subset T_i is feasible if $T_i \in \mathcal{I}^i$. We also assume there are value and independence oracles, i.e., for any $i \in [m]$ and subset A, we can obtain the value of $f_i(A)$ and know if $A \in \mathcal{I}^i$ or not.

3 *P*-Matroid System Constraints

In this section, we extend the ReplacementGreedy algorithm introduced by [17] (for comparison, we say their algorithm as One-to-One ReplacementGreedy) for a single matroid constraint to address the two-stage submodular maximization under p-matroid system constraints.

Algorithm 1. One-to-Many ReplacementGreedy

$S \leftarrow \emptyset$, $T_i \leftarrow \emptyset$ for all $i \in [m]$
for $t \in [\ell]$ **do**
  $x^* \leftarrow \arg\max_{x \in \Omega} \frac{1}{m} \sum_{i=1}^{m} \nabla_i'(x, T_i)$
  $S \leftarrow S \cup x^*$
  for all $i \in [m]$ **do**
    if $\nabla_i'(x, T_i) > 0$ **then**
      $T_i \leftarrow T_i \cup \{x^*\} \setminus \mathrm{Rep}_i'(x^*, T_i)$
    end if
  end for
  $t \leftarrow t + 1$
end for
Return S and $\{T_i\}_{i \in [m]}$

3.1 Algorithm

In order to have a better understand of our One-to-Many ReplacementGreedy, we investigate the One-to-One ReplacementGreedy in the first. The *replacement gain* is denoted as $\nabla_i(x, A)$, who characterizes how much they can increase the value of $f_i(A)$ by either adding x to A or replacing x with one element of A while preserving the independence of A. We restate the related notations as follows. Set $\Delta_i(x, A) = f_i(A \cup \{x\}) - f_i(A)$ as the *marginal gain* of adding x to A for any $i \in [m]$. We restate the *replace gain* of deleting an element $y \in A$ and replacing it with x as $\nabla_i(x, y, A) = f_i(A \cup \{x\} \setminus \{y\}) - f_i(A)$. As maintaining the independence of solution in each iteration is a very important point in One-to-One ReplacementGreedy algorithm, let $\mathcal{I}(x, A)$ be the set of feasible candidate $y \in A$, i.e., $\mathcal{I}(x, A) = \{y \in A : A \cup \{x\} \setminus \{y\} \in \mathcal{I}\}$. Finally, we formally redefine the replacement gain as

$$\nabla_i(x, A) = \begin{cases} \Delta_i(x, A), \text{ if } A \cup \{x\} \in \mathcal{I} \\ \max\{0, \max_{y \in \mathcal{I}(x,A)} \nabla_i(x, y, A)\}, \text{o.w.} \end{cases}$$

To specific say the element with the maximum replacement gain due to x, they define

$$\text{Rep}_i(x, A) = \begin{cases} \emptyset, \text{ if } A \cup \{x\} \in \mathcal{I} \\ \arg\max_{y \in \mathcal{I}(x,A)} \nabla_i(x, y, A), \text{o.w.} \end{cases}$$

In our p-matroid system constraint setting, we define $\nabla_i(x, Y, A) = f_i(A \cup \{x\} \setminus Y) - f_i(A)$ as the new *replacement gain*. To keep the independence in each loop, we set $\mathcal{I}'^{,i}(x, A) = \{Y \subseteq A : |Y| \leq p, A \cup \{x\} \setminus Y \in \mathcal{I}^i (= \cap_{j=1}^p \mathcal{I}_j^i)\}$ as the new collection of candidate subsets. Let

$$\nabla_i'(x, A) = \begin{cases} \Delta_i(x, A), \text{ if } A \cup \{x\} \in \mathcal{I}^i \\ \max\{0, \max_{Y \subseteq \mathcal{I}'^{,i}(x,A)} \nabla_i(x, Y, A)\}, \text{o.w.} \end{cases}$$

Similarly, we set

$$\text{Rep}_i'(x, A) = \begin{cases} \emptyset, \text{ if } A \cup \{x\} \in \mathcal{I} \\ \arg\max_{Y \subseteq \mathcal{I}'^{,i}(x,A)} \nabla_i(x, Y, A), \text{o.w.} \end{cases}$$

In the One-to-Many ReplacementGreedy, we greedily choose an element x^* with the maximum average new replacement gain in each iteration until the size of S increases to ℓ. In each iteration, for any current substitute solution set T_i, $i \in [m]$, if the new replacement gain $\nabla_i'(x, A) > 0$, we will update T_i by removing $\text{Rep}_i'(x^*, T_i)$. The main pseudo codes are presented by Algorithm 1.

3.2 Theoretical Analysis

In this section, we will analyze the performance ratio of One-to-Many ReplacementGreedy. For clarity, we adopt the notations provided by [2,14,16]. Let

$$S^{m,\ell} = \arg \max_{S \subseteq \Omega, |S| \leq \ell} \left\{ \frac{1}{m} \sum_{i=1}^{m} \max_{T_i \subseteq S, T_i \in \mathcal{I}^i(S)} f_i(T_i) \right\}$$

be any optimal solution, and set $S_i^{m,\ell} = \arg\max_{T \in \mathcal{I}^i(S^{m,\ell})} f_i(T)$ to be the independent subset of $S^{m,\ell}$ with maximum f_i value for $i \in [m]$. Let T_i^t be the solution at the end of iteration t. Then our main result can be summarized as following theorem.

Theorem 1. *For any fixed $p \geq 1$, the One-to-Many ReplacementGreedy is a $1/(p+1)(1-1/e^2)$-approximation algorithm for the two-stage submodular maximization with a p-matroid system constraint $M^i = (\Omega, \cap_{j=1}^p \mathcal{I}_j^i)$ for each $i \in [m]$.*

Proof. Let X_t be the total value $\frac{1}{m}\sum_{i=1}^{m} f_i(T_i^t)$ in the end of t iteration, then the increment of value during t iteration can be lower bounded by as follows.

$$X_{t+1} - X_t \geq \frac{G_m(S^{m,\ell})}{\ell} - \frac{(p+1)X_t}{\ell}. \tag{2}$$

From inequality (2), we could complete the main proof by induction. We assume

$$X_t \geq \frac{1}{p+1}\left(1 - (1-\frac{1}{\ell})^{2t}\right) G_m(S^{m,\ell}).$$

In basis step, if $t = 1$, we have $X_1 \geq \frac{1}{\ell} G_m(S^{m,\ell}) \geq \frac{2}{(p+1)\ell} G_m(S^{m,\ell})$, where the first inequality follows from the line 3 in Algorithm 1. In the reduction step, as

$$\begin{aligned}
X_{t+1} &\geq \frac{1}{\ell} G_m(S^{m,\ell}) + \left(1 - \frac{p+1}{\ell}\right) X_t \\
&\geq \frac{1}{\ell} G_m(S^{m,\ell}) + \left(1 - \frac{p+1}{\ell}\right) \cdot \left[\frac{1}{p+1}\left(1 - (1-\frac{1}{\ell})^{2t}\right) G_m(S^{m,\ell})\right] \\
&= \frac{1}{p+1} \cdot \left[1 - (1-\frac{1}{\ell})^{2t}(1 - \frac{p+1}{\ell})\right] \cdot G_m(S^{m,\ell}) \\
&\geq \frac{1}{p+1}\left(1 - (1-\frac{1}{\ell})^{2t+2}\right) \cdot G_m(S^{m,\ell}).
\end{aligned}$$

By the above reduction process, we complete the assumption. At the end of ℓ iteration, we have

$$\begin{aligned}
\frac{1}{m}\sum_{i=1}^{m} f_i(T_i^\ell) = X_\ell &\geq \frac{1}{p+1}(1 - (1-\frac{1}{\ell})^{2\ell}) \cdot G_m(S^{m,\ell}) \\
&\geq \frac{1}{p+1}(1 - \frac{1}{e^2}) \cdot G_m(S^{m,\ell}),
\end{aligned}$$

where the second inequality is derived by the inequality of $e^{-x} \geq 1 - x$.

In order to prove the inequality (2), we provide two technical lemmas as follows.

Lemma 1. *For any $i \in [m]$, given any two independent sets $A, B \in \mathcal{I}^i = \cap_{j=1}^{p} \mathcal{I}_j$, there exists a mapping of elements in $B \setminus A$ to $[A \setminus B]^{\leq p}$ (namely, a collection of subsets included into $A \setminus B$ of size at most p), such that each element $u \in A \setminus B$ appears in at most p subsets.*

Proof. Refer to the full version of this paper.

Lemma 2. *For any $i \in [m], t \in [\ell]$, let $\pi_t^i : S_i^{m,\ell} \setminus T_i^t \to [T_i^t \setminus S_i^{m,\ell}]^{\leq p}$ be the mapping derived by Lemma 1, then we have*

$$\sum_{x \in S_i^{m,\ell} \setminus T_i^t} \Delta_i(\pi_t^i(x), T_i^t \setminus \pi_t^i(x)) \leq p \cdot f_i(T_i^t).$$

Proof. Refer to the full version of this paper.

To prove the increment of value during t iteration we have the following

$$\sum_{i=1}^{m} \nabla'_i(x^*, T_i^t) \geq \frac{1}{\ell} \sum_{x \in S^{m,\ell}} \sum_{i=1}^{m} \nabla'_i(x, T_i^t) \tag{3}$$

$$\geq \frac{1}{\ell} \sum_{i=1}^{m} \sum_{x \in S_i^{m,\ell} \setminus T_i^t} \nabla'_i(x, T_i^t) \tag{4}$$

$$\geq \frac{1}{\ell} \sum_{i=1}^{m} \sum_{x \in S_i^{m,\ell} \setminus T_i^t} f_i(T_i^t \cup \{x\} \setminus \pi_t^i(x)) - f_i(T_i^t) \tag{5}$$

$$= \frac{1}{\ell} \sum_{i=1}^{m} \sum_{x \in S_i^{m,\ell} \setminus T_i^t} \Delta_i(x, T_i^t) - \Delta_i(\pi_t^i(x), T_i^t \cup \{x\} \setminus \pi_t^i(x))$$

$$\geq \frac{1}{\ell} \sum_{i=1}^{m} \sum_{x \in S_i^{m,\ell} \setminus T_i^t} \Delta_i(x, T_i^t) - \Delta_i(\pi_t^i(x), T_i^t \setminus \pi_t^i(x)) \tag{6}$$

$$\geq \frac{1}{\ell} \sum_{i=1}^{m} f_i(S_i^{m,\ell}) - (p+1) f_i(T_i^t). \tag{7}$$

The inequality (3) is obtained by the selection of x^* from the ground set Ω in each iteration. The inequality (4) follows from the fact of on-negativity of $\nabla'_i(x, T_i)$ for all $i \in [m]$. The inequality (5) is derived by

$$\frac{1}{\ell} \sum_{i=1}^{m} \sum_{x \in S_i^{m,\ell}} \nabla'_i(x, T_i^t) = \frac{1}{\ell} \sum_{i=1}^{m} \sum_{x \in S_i^{m,\ell}} f_i(T_i^t \cup \{x\} \setminus \mathrm{Rep}'_i(x, T_i^t)) - f_i(T_i^t)$$

$$\geq \frac{1}{\ell} \sum_{i=1}^{m} \sum_{x \in S_i^{m,\ell}} f_i(T_i^t \cup \{x\} \setminus \pi_t^i(x)) - f_i(T_i^t),$$

where the inequality holds because the $\mathrm{Rep}'_i(x, T_i)$ is the maximum subset and the $\pi_t^i(x)$ is an feasible subset of $\mathcal{I}^i(x, T_i^t)$. The inequality (6) is implied by the submodularity. By the additivity of submodularity of any $f_i, i \in [m]$, we have

$$\sum_{x \in S_i^{m,\ell} \setminus T_i^t} \Delta_i(x, T_i^t) \geq f_i(S_i^{m,\ell}) - f_i(T_i^t). \tag{8}$$

Combining Lemma 2 and inequality (8), we obtain the inequality (7). The inequality (2) can be directly obtained by the above process.

The main result can be described as the following theorem.

Theorem 2. *For any fixed $p \geq 1$, the query complexity of the One-to-Many ReplacementGreedy algorithm is upper bounded by $O(\ell m n r^p)$.*

Proof. It concludes that the main time computation is the greedy chosen of line 3 in Algorithm 1. Given any iteration $t \in [\ell]$ and $i \in [m]$, it needs to check at most $O(n)$ elements to find the element x^* while it also needs at most $O(r^p)$ function evaluations by enumerating all candidate subsets, where r is the maximum size of feasible subsets belong to p-matroid. Then the total query complexity (i.e., the number of function evaluations) of Algorithm 1 is bounded by $O(\ell mnr^p)$.

4 *P*-Extendible System Constraints

In this section, we extend our algorithm for p-matroid system constraints to dealing with p-extendible system constraints. As discussed in the work of [6, 12], the p-extendible system constraint is a generalization of p-matroid system constraint. In our model, we choose a set S of size at most ℓ, while we also select a set $T_i \subseteq S$ such that $T_i \in \mathcal{I}^i$ for each $i \in [m]$, where $M^i = (S, \mathcal{I}^i)$ is a p-extendible system. The aim is to maximize the average of summarization of their function values.

Algorithm 2. Generalized One-to-Many ReplacementGreedy

$S, T_i, E(T_i) \leftarrow \emptyset$ for all $i \in [m]$
for $t \in [\ell]$ **do**
  $x^* \leftarrow \arg\max_{x \in \Omega} \frac{1}{m} \sum_{i=1}^{m} \tilde{\nabla}_i(x, T_i)$
  $S \leftarrow S \cup x^*$
  for all $i \in [m]$ **do**
    if $\tilde{\nabla}_i(x, T_i) > 0$ **then**
      $T_i \leftarrow E(T_i) \cup \{x^*\} \setminus \tilde{\text{Rep}}_i(x^*, T_i)$
    end if
  end for
  compute an extension $E(T_i)$ of T_i
  $t \leftarrow t + 1$
end for
Return S and $\{E(T_i)\}_{i \in [m]}$

4.1 Algorithm

Following from the definition of p-extendible system, we have if $A \subseteq B \in \mathcal{I}$ and $A \cup \{x\} \in \mathcal{I}$, then there exists a subset $Y \subseteq B \setminus A$ with $|Y| \leq p$ such that $B \setminus A \cup \{x\} \in \mathcal{I}$. Given a p-extendible system $M^i = (\Omega, \mathcal{I}^i)$ for each $i \in [m]$. The goal is to select a set S of size at most ℓ, such that the average of the summary of the optimum of f_i restricted to S according to a p-extendible system $M^i = (\Omega, \mathcal{I}^i)$ for all $i \in [m]$ is maximum.

In our setting, to keep the independence of $T_i, i \in [m]$ in each iteration under p-extendible system constraint, we modify the One-to-Many ReplacementGreedy to a Generalized One-to-Many ReplacementGreedy, the main pseudo codes are presented in Algorithm 2.

For each $i \in [m]$, we define $\tilde{\nabla}_i(x, A)$ as the new replacement gain, in specific, $\tilde{\nabla}_i(x, Y, A) = f_i(E(A) \cup \{x\} \setminus Y) - f_i(E(A))$, where $E(A)$ is an extension of A. Let $\tilde{\mathcal{I}}^i(x, A) = \{Y \subseteq E(A) \setminus A : |Y| \leq p, A \cup \{x\} \in \mathcal{I}^i, E(A) \cup \{x\} \setminus Y \in \mathcal{I}^i\}$ be the candidate set. Let

$$\tilde{\nabla}_i(x, A) = \begin{cases} \Delta_i(x, A), \text{ if } A \cup \{x\} \in \mathcal{I}^i \\ \max\{0, \max_{Y \in \tilde{\mathcal{I}}^i(x,A)} \tilde{\nabla}_i(x, Y, A)\}, \text{o.w.} \end{cases}$$

Simultaneously, let

$$\tilde{\text{Rep}}_i(x, A) = \begin{cases} \emptyset, \text{ if } A \cup \{x\} \in \mathcal{I} \\ \arg\max_{Y \in \tilde{\mathcal{I}}^i(x,A)} \tilde{\nabla}_i(x, A), \text{o.w.} \end{cases}$$

4.2 Theoretical Analysis

In this section, we present the analyses of the Generalized One-to-Many ReplacementGreedy. The setting under p-extendible system constraints differ from the p-matroid constraints, that is, there is not such similar mapping presented by Lemma 2. We notice that there is interesting property provided by the following lemma.

Lemma 3. *For any $i \in [m]$, let $\{Y_j^i\}_{j=1}^q$ be a collection of $E(T_i) \setminus T_i$ such that each element of $E(T_i) \setminus T_i$ appears in at most r of these subsets, where r is the size of maximal independence set in p-extendible system. Then we have*

$$\sum_{j=1}^{q} (f(E(T_i)) - f_i(E(T_i) \setminus Y_j^i)) \leq r \cdot (f(E(T_i)) - f(T_i)).$$

Proof. Refer to the full version of this paper.

Theorem 3. *For any fixed $p \geq 1$, the Generalized One-to-Many ReplacementGreedy is a $1/(r+1)(1-1/e^2)$-approximation algorithm, while the query complexity is upper bounded by $O(\ell m n r^p)$, for the two-stage submodular maximization with p-extendible system constraints $M^i = (\Omega, \mathcal{I}^i)$ for each $i \in [m]$, where r is the size of the maximum independence set in $\mathcal{I}^i$.*

Proof. Refer to the full version of this paper.

5 Experiments

In this section, we run Algorithm 2, generalized one-to-many ReplaceGreedy (say, G-REPLACEGREEDY), on the application exemplar-based clustering with two dataset and consider the following benchmarks:

- Random selection (i.e., Random): the output is randomly k elements chosen for each function f_i, $i \in [m]$.
- Greedy-Sum (i.e., Greedy-SUM): the output is greedily k elements selected for each function f_i, $i \in [m]$, and return the union as S.

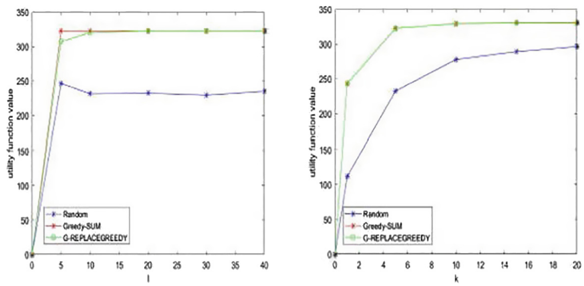

Fig. 1. Performances of Algorithm 2 comparing with Random and Geedy-Sum on Census.

We consider the application of exemplar-based clustering on Census 1990 [1], which has 24, 581 elements with 68 attributes. Let the first 10 attributes as our classified genres, such as, age, ancestry, citizen etc. In the experiment, we first choose a dataset Ω of 500 different people by reservoir sampling [18]. We denote Ω_i as the set of people containing genre $i \in [m]$. All people are expressed by their features vector, the distance of any two vectors is calculated by Euclidean distance and $d(v, S) = \min_{u \in S} d(u, v)$ denotes the distance of element v to set S. The goal is to choose a subset $S \subseteq \Omega$ of size at most ℓ, such that each genre $i \in [m]$ has a good expression of size limit k. For each genre $i \in [m]$, the utility function $f_i(S)$ is defined by Exemplar Based Clustering [1,17]. We restate as follows

$$f_i(S) = L_i(\{e_0\}) - L_i(S \cup \{e_0\}),$$

where e_0 is an auxiliary vector (w.l.o.g., $e_0 = \mathbf{0}$), $S_i = S \cap \Omega_i$ is the set of people with genre i, and $L_i(S) = \frac{1}{|\Omega_i|} \sum_{x \in \Omega_i} d(x, S_i)$. As the submodularity and non-negativity of utility function have been discussed by [1,8], we omit the proof here. The left of Fig. 1 shows the performance of Algorithm 2 on census application, when k is fixed as 5 and the right figure shows that the performance of Algorithm 2 with fixing $\ell = 20$. We observe that if k is fixed, then the function value goes to some assured values with the increasing of ℓ and our Algorithm 2 performs similarly to Greedy-Sum.

We also consider a classification application that feature vectors are generated from a random distribution. Specifically, we generate a 500×10 feature matrix which each component is randomly chosen from the range $[0, 3]$. Figure 2 shows the performance of Algorithm 2 on the classification, when k and ℓ are fixed, respectively. We observe that our Algorithm 2 still matches the Greedy-

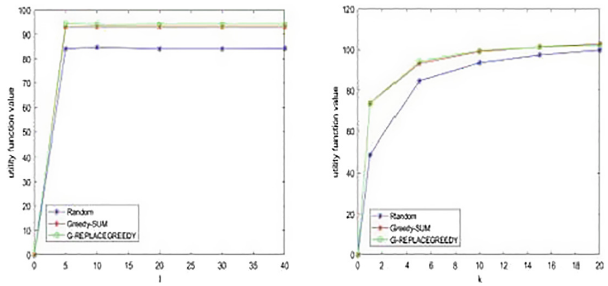

Fig. 2. Performances of Algorithm 2 comparing with Random and Geedy-Sum on Classification.

Sum algorithm, and performs better than Random algorithm. As the generation process of data, we also observe that the three algorithms perform similarly as ℓ increases.

6 Conclusion

We consider the two-stage submodular maximization under p-matroid and p-extendible system constraints for each sub-functions, respectively. Specifically, for the front model, we derive a $1/(p+1)(1-1/e^2)$-approximation algorithm, which needs $O(\ell mnr^p)$ function evaluations. For the second setting, we obtain a $1/(r+1)(1-1/e^2)$-approximation algorithm with the same query complexity. In the end, we show the performance of our generalized One-to-Many algorithm on some datasets.

Acknowledgements. The first and sixth authors are supported by Natural Science Foundation of China (Nos. 11531014, 11871081). The second and fourth authors are supported by Natural Science Foundation (No. 1747818).

References

1. Badanidiyuru, A., Mirzasoleiman, B., Karbasi, A., Krause, A.: Streaming submodular maximization: massive data summarization on the fly. In: Proceedings of SIGKDD, pp. 671–680 (2014)
2. Balkanski, E., Mirzasoleiman, B., Krause, A., Singer, Y.: Learning sparse combinatorial representations via two-stage submodular maximization. In: Proceedings of ICML, pp. 2207–2216 (2016)
3. Buchbinder, N., Feldman, M., Garg, M.: Deterministic (1/2+ ε)-approximation for submodular maximization over a matroid. In: Proceedings of SODA, pp. 241–254 (2019)
4. Calinescu, G., Chekuri, C., Pál, M., Vondrák, J.: Maximizing a monotone submodular function subject to a matroid constraint. SIAM J. Comput. **40**(6), 1740–1766 (2011)

5. Dasgupta, A., Kumar, R., Ravi, S.: Summarization through submodularity and dispersion. In: Proceedings of ACL, pp. 1014–1022 (2013)
6. Feldman, M., Harshaw, C., Karbasi, A.: Greed is good: near-optimal submodular maximization via greedy optimization. arXiv: 1704.01652 (2017)
7. Fisher, M.L., Nemhauser, G.L., Wolsey, L.A.: An analysis of approximations for maximizing submodular set functions-II. In: Balinski, M.L., Hoffman, A.J. (eds.) Polyhedral Combinatorics, pp. 73–87. Springer, Heidelberg (1978). https://doi.org/10.1007/BFb0121195
8. Gomes, R., Krause, A.: Budgeted nonparametric learning from data streams. In: Proceedings of ICML, pp. 391–398 (2010)
9. Kempe, D., Kleinberg, J., Tardos, É.: Maximizing the spread of influence through a social network. In: Proceedings of SIGKDD, pp. 137–146 (2003)
10. Lee, J., Sviridenko, M., Vondrák, J.: Submodular maximization over multiple matroids via generalized exchange properties. Math. Oper. Res. **35**(4), 795–806 (2010)
11. Lin, H., Bilmes, J.: A class of submodular functions for document summarization. In: Proceedings of ACL, pp. 510–520 (2011)
12. Mestre, J.: Greedy in approximation algorithms. In: Proceedings of ESA, pp. 528–539 (2006)
13. Mitrovic, S., Bogunovic, I., Norouzi-Fard, A., Tarnawski, J.M., Cevher, V.: Streaming robust submodular maximization: a partitioned thresholding approach. In: Proceedings of NIPS, pp. 4557–4566 (2017)
14. Mitrovic, M., Kazemi, E., Zadimoghaddam, M., Karbasi, A.: Data summarization at scale: a two-stage submodular approach. arXiv preprint arXiv:1806.02815 (2018)
15. Nemhauser, G.L., Wolsey, L.A., Fisher, M.L.: An analysis of approximations for maximizing submodular setfunctions-I. Math. Program. **14**(1), 265–294 (1978)
16. Sarpatwar, K.K., Schieber, B., Shachnai, H.: Constrained submodular maximization via greedy local search. Oper. Res. Lett. **47**(1), 1–6 (2019)
17. Stan, S., Zadimoghaddam, M., Krause, A., Karbasi, A.: Probabilistic submodular maximization in sub-linear time. In: Proceedings of ICML, pp. 3241–3250 (2017)
18. Vitter, J.S.: Random sampling with a reservoir. ACM Trans. Math. Softw. **11**(1), 37–57 (1985)
19. Wu, W.L., Zhang, Z., Du, D.Z.: Set function optimization. J. Oper. Res. Soc. China (2018). https://doi.org/10.1007/s40305018-0233-3