arXiv:2302.01546v2 [cs.LG] 9 Mar 2023

# 非单调子模最大化中的群体公平性

Jing Yuan[1][0000−0001−6407−834X] 和 Shaojie Tang[2][0000−0001−9261−5210]

[1] 北德克萨斯大学计算机科学与工程系 jing.yuan@unt.edu[2] 德克萨斯大学达拉斯分校纳文·金达尔管理学院 shaojie.tang@utdallas.edu

摘要。最大化次模函数在机器学习和数据挖掘中有广泛的应用。其中一项应用是数据摘要，目标是从大型数据集中选取一小组具有代表性且多样化的数据项。然而，数据项可能包含种族或性别等敏感属性。在这种情况下，设计 *fairness-aware* 算法以减轻可能导致特定组代表性过高或不足的潜在算法偏见至关重要。受此启发，我们提出并研究了带有新型群体公平性约束的经典非单调次模最大化问题。我们的目标是选取一组项目，使非单调次模函数最大化，同时确保从每个组中选取的项目数量与该组的规模成比例，具体比例由决策者指定。我们为该问题开发了首个常数因子近似算法。我们还将基本模型扩展为纳入关于所选项目总数的额外全局规模约束。

## 1 引言

次模函数指满足自然的边际收益递减性质的一大类函数：向一个较大的已有子集添加一个额外元素，带来的收益较小。许多机器学习和人工智能问题，包括基于代表点的聚类 [7], 特征选择 [6], 主动学习 [12],社交网络中的影响力最大化 [20], 推荐系统 [9], 以及多样化数据摘要 [19], ，都可以建模为次模最大化问题。该问题旨在选择一组项目，使次模函数最大化；其各种变体 [14,18] 也已在文献中得到广泛研究，并受到多种约束的限制，包括基数约束、拟阵约束或背包型限制。

我们注意到，在实践中，元素或个体通常会根据性别、种族、年龄、宗教或其他因素等不同属性被划分到不同的组中。如果不加以约束，现有算法可能会表现出偏见，例如，某些组在最终选定的子集中可能代表过多或过少。因此，设计能够缓解此类问题的公平性算法变得日益重要。为此，我们提出并

研究受新型群体公平性约束的经典非单调次模最大化问题。我们的目标是选择一个均衡的元素集合，使其最大化非单调次模函数，并确保每个组中所选元素数量与该组规模之比处于决策者确定的期望范围内。非单调次模最大化有多种重要应用，例如特征选择 [6], 利润最大化 [21],最大割 [13] 和数据摘要 [16]。形式上，我们考虑一个包含 $V$ 个元素（例如数据点）的集合，这些元素被划分为 $m$ 个组：$V_1$、$V_2$、$\cdots$、$V_m$，同一组中的元素具有相同的属性（例如性别）。如果一个包含 $S \subseteq V$ 个元素的集合满足：对于所有组 $i \in [m]$，均有 $\lfloor \alpha |V_i| \rfloor \leq |S \cap V_i| \leq \lfloor \beta |V_i| \rfloor$，我们称其为（$\alpha$，$\beta$）-公平的。利用我们的模型，决策者可以通过为 $\alpha$ 和 $\beta$ 设置适当的值，指定所需的公平性水平。具体而言，设置 $\alpha = \beta$ 会带来最高程度的公平性，即所选元素的数量与其所在组的规模严格成比例；如果设置 $\alpha = 0$ 和 $\beta = 1$，则不存在公平性约束。我们的目标是找到这样一个（$\alpha$，$\beta$）-公平的元素子集，使其最大化次模目标函数。我们对公平性的定义通过敏感属性来平衡解，已获得学术界的广泛认可，这一点从其在以往研究中的频繁使用中可以看出 [4,10,5]。我们的表述还可以涵盖其他几种公平性概念，例如 80%规则 [1],统计均等 [8] 和比例代表制 [17]。

### 1.1 我们的贡献

– 我们的研究开创性地探讨了在 $(\alpha, \beta)$-公平性约束下的经典（非单调）子模最大化问题。通过调整 $\alpha$ 和 $\beta$ 的值，我们的模型能够灵活地按照决策者的需求，体现不同程度的公平性。

– 我们为这一问题设计了首个常数因子近似算法。我们发现，参数 $\alpha$ 与求解 $(\alpha, \beta)$-公平非单调次模最大化问题的复杂度密切相关。具体而言，当 $\alpha \leq 1/2$ 时，我们设计了一个 $\frac{\gamma}{2}$ - 近似算法；当 $\alpha > 1/2$ 时，我们开发了一个 $\frac{\gamma}{3}$ - 近似算法，其中 $\gamma$ 是当前拟阵约束次模最大化最佳算法的近似比。我们还将基本模型扩展为纳入一项额外的全局规模约束，以限制所选项目的总数。对于这一扩展模型，我们提供了近似比为常数因子的近似算法。

### 1.2 其他相关工作

近年来，人们日益认识到公平、无偏的决策系统的重要性。这推动了人们对开发公平算法的兴趣不断增长，其应用范围广泛，包括影响力最大化 [25], 分类 [26], 投票 [4], 老虎机学习 [15],和数据摘要 [3]。根据具体情境以及试图缓解的

偏见类型，现有研究采用了不同的公平性度量。这可能会导致不同的优化问题，以及针对应用具体需求量身定制的不同公平算法。我们对公平性的定义足够通用，能够涵盖许多现有定义，例如 80%规则 [1], 统计均等 [8] 和比例代表制 [17]。与大多数公平次模最大化研究不同 [4] ，这些研究的目标是最大化单调次模函数，[10] 我们在流式非单调次模最大化的背景下开发公平算法。他们提出的公平性定义比我们的更通用，因此优化问题更具挑战性，且不存在任何常数因子近似算法。 [24,23] 旨在开发满足平均公平性约束的随机化算法。最近， [22] 将公平算法的研究拓展至更复杂的自适应设置，并提出了一种名为群体平等的新公平性度量。

## 2 预备知识与问题陈述

我们考虑一个包含 $V$ 个 $n$ 的集合。存在一个非负子模效用函数 $f: 2^V \to \mathbb{R}_+$。用 $f(e \mid S)$ 表示在 $S \subseteq V$ 的基础上加入 $e \in V$ 所产生的边际效用，即 $f(e \mid S) = f(\{e\} \cup S) - f(S)$。如果对于任意两个集合 $X$、$Y \subseteq V$，满足 $X \subseteq Y$，以及任意元素 $e \in V \setminus Y$，函数 $f: 2^V \to \mathbb{R}_+$ 满足以下条件，则称其为子模函数：

$$f(e \mid Y) \leq f(e \mid X).$$

假设 $V$ 被划分为 $m$ 个互不相交的组：$V_1$、$V_2$、$\cdots$、$V_m$。我们假设，对于每个组中必须包含在可行解中的元素比例，都给定了一个下界和上界。这两个界，即 $\alpha$ 和 $\beta$，代表群体公平性约束。$(\alpha, \beta)$-公平次模最大化问题（标记为 P.0）可以表述如下。

P.0 max $f(S)$
满足：
$\lfloor \alpha |V_i| \rfloor \leq |S \cap V_i| \leq \lfloor \beta |V_i| \rfloor, \forall i \in [m].$

通过选择适当的 $\alpha$ 和 $\beta$ 的值，可以调整可行解中的群体公平性程度。也就是说，当 $\alpha = \beta$ 时，可实现严格的群体公平性，此时每个可行解都必须包含来自每个组的相同比例的项目；如果设定 $\alpha = 0$ 和 $\beta = 1$，则不存在群体公平性约束。接下来，我们给出该问题的 NP 难度结果。

引理 1. 问题 P.0 是 NP 难的。

证明：我们通过归约到经典的基数约束次模最大化问题来证明，下面给出该问题的定义。

定义 1. 基数约束次模最大化问题的输入包括一组元素 $U$、一个次模函数 $h: 2^U \to \mathbb{R}_+$ 和一个基数约束 $b$；我们的目标是选择一组元素 $S \subseteq U$，使得 $h(S)$ 最大化，并且 $|S| \leq b$。

接下来，我们证明可以将基数约束次模最大化问题归约为 P.0。给定任意一个基数约束次模最大化问题实例，我们按如下方式构造对应的 P.0 实例：令 $V = U$、$f = h$，假设只有一个组，即 $V = V_1$，并令 $\alpha = 0$、$\beta = b/|U|$。很容易验证这两个实例是等价的。至此，归约的证明完成。□

## 3 带群体公平性的非单调次模最大化

预热：单调效用函数 如果 $f$ 是单调子模的，我们可以很容易地确认，移除下界约束后，P.0 可以简化为 P.1。这是因为在这种情况下，通过添加更多项目来增大解的规模不会降低其效用。因此，只要向解中添加足够多的项目，就始终能够满足 P.0 中的下界约束，即 $\lfloor \alpha |V_i| \rfloor \leq |S \cap V_i|$ 对所有 $i \in [m]$ 均成立。

P.1 max $f(S)$
满足以下条件：
$|S \cap V_i| \leq \lfloor \beta |V_i| \rfloor, \forall i \in [m]$。

由于 $f$ 是单调次模函数，P.1 是一个众所周知的问题，即在拟阵约束[3]下最大化单调次模函数。该问题存在一个（$1 - 1/e$）近似算法。

接下来，我们将为非单调函数开发近似算法。我们将考察两种情形，具体来说，分别是 $\alpha \leq 1/2$ 和$\alpha > 1/2$。

### 3.1 当 $\alpha \leq 1/2$时的情形

在 $\alpha \leq 1/2$ 的情形下，我们将 P.1 的解作为构建算法的基础。首先，很容易验证，P.1 是 P.0 的松弛版本，其中移除了 P.0 中的下界约束 $\lfloor \alpha |V_i| \rfloor \leq |S \cap V_i|$。由于 $f$ 是次模函数，P.1 是一个经典问题：在拟阵约束下最大化（非单调）子模函数。针对 P.1 已经有有效的解法。现在，我们可以介绍如下算法设计。

[3] 拟阵是一个二元组 $\mathcal{M} = (V, \mathcal{I})$，其中 $\mathcal{I} \subseteq 2^V$ 且 1. $\forall Y \in \mathcal{I}, X \subseteq Y \to X \in \mathcal{I}$。2. $\forall X, Y \in \mathcal{I}$； $|X| < |Y| \to \exists e \in Y \setminus X$； $X \cup \{e\} \in \mathcal{I}$。

1. 应用用于拟阵约束下的次模最大化的最先进算法 $\mathcal{A}$ 来求解 P.1，并获得一个解 $A^{P.1}$。

2. 注意，$A^{P.1}$ 不一定是 P.0 的可行解，因为它可能违反某些组的下界约束 $\lfloor \alpha |V_i| \rfloor \leq |S \cap V_i|$。为使其可行，我们向 $A^{P.1}$ 中添加额外的项目。具体而言，对于满足 $|A^{P.1} \cap V_i| < \lfloor \alpha |V_i| \rfloor$ 的每个组 $i \in [m]$，我们的算法通过从 $V_i \setminus A^{P.1}$ 中随机采样 $\lfloor \alpha |V_i| \rfloor - |A^{P.1} \cap V_i|$ 个项目，选择一个大小为 $\lfloor \alpha |V_i| \rfloor - |A^{P.1} \cap V_i|$ 的备用集 $B_i$。如果 $|A^{P.1} \cap V_i| \geq \lfloor \alpha |V_i| \rfloor$，则定义 $B_i = \emptyset$。

3. 最后，将 $\cup_{i \in [m]} B_i$ 添加到 $A^{P.1}$ 以构建最终解 $A^{approx}$，即 $A^{approx} = A^{P.1} \cup (\cup_{i \in [m]} B_i)$。

**算法 1**：当 $\alpha \leq 1/2$ 时求解 P.0 的近似算法

```
1: 应用 A 求解 P.1，并获得一个解 A^{P.1}
2: 对于每个组 i ∈ [m]，执行
3:     如果 |A^{P.1} ∩ V_i| < ⌊α|V_i|⌋，则
4:         从 V_i \ A^{P.1} 中随机选择一个大小为 ⌊α|V_i|⌋ − |A^{P.1} ∩ V_i| 的随机备用集 B_i
5:     else
6:         B_i ← ∅
7: A^{approx} ← A^{P.1} ∪ (∪_{i∈[m]} B_i)
8: 返回 A^{approx}
```

该近似算法的伪代码如算法 1 所示。注意，$A^{P.1}$ 是 P.1 的一个可行解，因此，$A^{P.1}$ 满足 P.1 的上界约束，进而也满足 P.0 的上界约束，即 $|S \cap V_i| \leq \lfloor \beta |V_i| \rfloor$、$\forall i \in [m]$。根据 $B_i$ 的构造，可以很容易地验证，将 $\cup_{i \in [m]} B_i$ 添加到 $A^{P.1}$ 不会违反上界约束，因为只有当某些组不满足 P.0 的下界约束（即 $\lfloor \alpha |V_i| \rfloor \leq |S \cap V_i|$）时，才会向这些组补充 $\cup_{i \in [m]} B_i$。此外，将 $\cup_{i \in [m]} B_i$ 添加到 $A^{P.1}$ 后，便可使其满足 P.0 的下界约束。因此，$A^{approx}$ 是 P.0 的一个可行解。

引理 2. $A^{approx}$ 是 P.0 的可行解。

性能分析 接下来，我们分析算法 1 的性能。首先介绍一个来自 [2] 的有用引理。

引理 3. 如果 $f$ 是子模的，且 $S$ 是 $V$ 的随机子集，并且 $V$ 中的每个元素以至多 $p$ 的概率包含在 $S$ 中，那么 $\mathbb{E}_S[f(S)] \geq (1-p) f(\{v13\})$。

下一个引理指出，如果 $A^{P.1}$ 是 P.1 的 $\gamma$-近似解，那么 $f(A^{P.1})$ 至少是 P.0 的最优解的 $\gamma$ 倍。

引理 4. 假设 $\mathcal{A}$ 是一个针对拟阵约束下非单调子模最大化问题的 $\gamma$-近似算法。令 $OPT$ 表示 P.0 的最优解，则我们有 $f(A^{P.1}) \geq \gamma f(OPT)$。

证明：由于 $\mathcal{A}$ 是一个针对拟阵约束下非单调子模最大化问题的 $\gamma$-近似算法，因此我们有 $f(A^{P.1}) \geq \gamma f(O^{P.1})$，其中 $O^{P.1}$ 表示 P.1 的最优解。此外，由于 P.1 是 P.0 的松弛版本，因此我们有 $f(O^{P.1}) \geq f(OPT)$。因此，$f(A^{P.1}) \geq \gamma f(OPT)$。□

接下来我们证明，用随机集合 $\cup_{i\in[m]}B_i$ 中的元素增补 $A^{P.1}$，在期望意义下至多会使其效用降低 $1/2$ 倍。这里的期望是对 $\cup_{i\in[m]}B_i$ 的分布取的。

引理 5。假设 $\alpha \leq 1/2$，我们有 $\mathbb{E}_{A^{approx}}[f(A^{approx})] \geq \frac{1}{2}f(A^{P.1})$，其中 $A^{approx} = A^{P.1} \cup (\cup_{i\in[m]}B_i)$。

证明：回顾一下，对于所有满足 $|A^{P.1} \cap V_i| \geq \lfloor\alpha|V_i|\rfloor$ 的 $i \in [m]$，都有 $B_i = \emptyset$，因此，将这些 $B_i$ 加入 $A^{P.1}$ 不会影响其效用。在证明的其余部分，我们关注满足以下条件的 $B_i$

$$|A^{P.1} \cap V_i| < \lfloor\alpha|V_i|\rfloor. \tag{1}$$

回顾一下，对于每个满足 $|A^{P.1} \cap V_i| < \lfloor\alpha|V_i|\rfloor$ 的 $i \in [m]$，$B_i$ 都是从 $V_i \setminus A^{P.1}$ 中抽取的、大小为 $\lfloor\alpha|V_i|\rfloor - |A^{P.1} \cap V_i|$ 的随机集合。因此，$V_i \setminus A^{P.1}$ 中的每个元素以至多以下概率包含在 $B_i$ 中

$$\frac{\lfloor\alpha|V_i|\rfloor - |A^{P.1} \cap V_i|}{|V_i \setminus A^{P.1}|}. \tag{2}$$

接下来，我们给出式 (2) 的一个上界。首先，

$$\lfloor\alpha|V_i|\rfloor - |A^{P.1} \cap V_i| \leq \lfloor\alpha|V_i|\rfloor \leq |V_i|/2, \tag{3}$$

其中，第二个不等式依据假设 $\alpha \leq 1/2$。此外，

$$|V_i \setminus A^{P.1}| = |V_i| - |A^{P.1} \cap V_i| = (\lfloor\alpha|V_i|\rfloor - |A^{P.1} \cap V_i|) + (|V_i| - \lfloor\alpha|V_i|\rfloor) \tag{4}$$

$$\geq (\lfloor\alpha|V_i|\rfloor - |A^{P.1} \cap V_i|) + |V_i|/2, \tag{5}$$

其中，该不等式依据假设 $\alpha \leq 1/2$。

因此，

$$(2) \leq \frac{\lfloor\alpha|V_i|\rfloor - |A^{P.1} \cap V_i|}{(\lfloor\alpha|V_i|\rfloor - |A^{P.1} \cap V_i|) + |V_i|/2} \leq \frac{|V_i|/2}{|V_i|/2 + |V_i|/2} = 1/2, \tag{6}$$

其中，第一个不等式依据式 (5)；第二个不等式依据式 (3) 以及假设 $\lfloor\alpha|V_i|\rfloor - |A^{P.1} \cap V_i| > 0$（列于式 (1)）。也就是说，$V_i \setminus A^{P.1}$ 中每个元素都包含在 $B_i$ 中的概率至多为 $1/2$。由此可得，$V \setminus A^{P.1}$ 中每个元素都包含在 $\cup_{i\in[m]}B_i$ 中的概率至多为 $1/2$。此外，引理 3 指出，如果 $f$ 是子模的，且 $S$ 是 $V$ 的一个随机子集，并且 $V$ 中的每个元素以至多 $p$ 的概率出现在 $S$ 中，那么 $\mathbb{E}_A[f(A)] \geq (1-p)f(\emptyset)$。结合上述讨论以及 $f(A^{P.1} \cup \cdot)$ 是子模函数这一观察，可得 $\mathbb{E}_{A^{approx}}[f(A^{approx})] = \mathbb{E}_{\cup_{i\in[m]}B_i}[f(A^{P.1} \cup (\cup_{i\in[m]}B_i))] \geq (1 - \frac{1}{2})f(A^{P.1} \cup \emptyset) = \frac{1}{2}f(A^{P.1})$。□

我们的主定理如下，由引理 4 和引理 5 推出。

定理1. 假设 $\mathcal{A}$ 是一个用于在拟阵约束和 $\alpha \leq 1/2$条件下进行非单调次模最大化的 $\gamma$-近似算法，则有$\mathbb{E}_{A^{approx}}[f(A^{approx})] \geq \frac{\gamma}{2} f(\{v7)$。

$\mathcal{A}$ 的一种选择是 [11]提出的连续双贪心算法，该算法给出一个 $1/e - o(1)$-近似解，即 $\gamma \geq 1/e - o(1)$。结合定理1可得， $\mathbb{E}_{A^{approx}}[f(A^{approx})] \geq \frac{1/e-o_{(1)}}{2} f(OPT)$。

### 3.2 当 $\boldsymbol{\alpha > 1/2}$ 时的情况

接下来，我们考虑 $\alpha > 1/2$的情况。首先介绍一个新的效用函数$g$： $2^V \to \mathbb{R}_+$ 如下:

$$g(\cdot) = f(V \setminus \cdot). \tag{7}$$

我们首先给出一个众所周知的结果：次模函数取补后仍保持次模性。

引理 6. 若 $f$ 是次模函数，则 $g$ 必定也是次模函数。

对于效用函数 $g$，我们提出如下新的优化问题 P.2:

**P.2** $\max g(S)$
**subject to:**
$|V_i| - \lfloor \beta |V_i| \rfloor \leq |S \cap V_i| \leq |V_i| - \lfloor \alpha |V_i| \rfloor, \forall i \in [m].$

P.2 是原问题 P.0 的一个翻转版本，因为如果存在 P.2 的一个 $\gamma$-近似解 $A^{P.2}$，很容易验证 $V \setminus A^{P.2}$ 是 P.0 的一个 $\gamma$-近似解。因此，本节余下部分将着重求解 P.2。

为求解 P.2，我们引入如下另一个问题（记为 P.3）：

P.3 最大化 $g(S)$
约束条件:
$|S \cap V_i| \leq |V_i| - \lfloor \alpha |V_i| \rfloor, \forall i \in [m]$。

P.3 是 P.2 的松弛版本，其中移除了 P.2 中的下界约束 $|V_i| - \lfloor \beta |V_i| \rfloor \leq |S \cap V_i|$。由于 $g$ 是一个次模函数，P.3 是一个经典问题：在拟阵约束下最大化次模函数。现在我们可以介绍算法的设计了。

1. 应用用于拟阵约束下的次模最大化的最先进算法 $\mathcal{A}$ 求解 P.3，并得到一个解 $A^{P.3}$。2. 请注意， $A^{P.3}$ 不一定是 P.2 的可行解，因为它可能违反某些组的下界约束 $|V_i| - \lfloor \beta |V_i| \rfloor \leq |S \cap V_i|$。我们向 $A^{P.3}$ 中添加额外元素，使其满足约束。具体而言，对于每个满足 $|A^{P.3} \cap V_i| < |V_i| - \lfloor \beta |V_i| \rfloor$的组 $i \in [m]$，我们的算法通过从 $V_i \setminus A^{P.3}$中随机采样$|V_i| - \lfloor \beta |V_i| \rfloor - |A^{P.3} \cap V_i|$个项目，选取一个大小为 $|V_i| - \lfloor \beta |V_i| \rfloor - |A^{P.3} \cap V_i|$的备用集 $B_i$。若 $|A^{P.1} \cap V_i| \geq |V_i| - \lfloor \beta |V_i| \rfloor$，则定义 $B_i = \emptyset$。

3. 将 $\cup_{i\in[m]}B_i$ 添加到 $A^{P.3}$ 中以构建 $A^{approx}$，即 $A^{approx} = A^{P.3} \cup (\cup_{i\in[m]}B_i)$。返回 $V \setminus A^{approx}$ 作为最终解。

**算法 2** 当 $\alpha > 1/2$ 时求解 P.0 的近似算法

```
1: 应用 A 求解 P.3，并得到一个解 A^{P.3}
2: 对每个组 i ∈ [m] 执行
3:     如果 |A^{P.3} ∩ V_i| < |V_i| − ⌊β|V_i|⌋，则
4:         从 V_i \ A^{P.3}中随机选择一个大小为 |V_i| − ⌊β|V_i|⌋ − |A^{P.3} ∩ V_i| 的随机备用集 B_i
5:     else
6:         B_i ← ∅
7:   A^{approx} ← A^{P.3} ∪(∪_{i∈[m]} B_i)
8: 返回     V \ A^{approx}
```

该近似算法的伪代码如算法 2 所示。注意，$A^{P.3}$ 满足 P.3 的上界约束，因此也满足 P.2 的上界约束，因为 $A^{P.3}$ 是 P.3 的一个可行解。根据 $B_i$的构造，向 $A^{P.1}$ 中添加 $\cup_{i\in[m]}B_i$ 不会违反上界约束，因为必要时会添加 $\cup_{i\in[m]}B_i$ 以满足 P.2 的下界约束。此外，向 $A^{P.3}$ 中添加 $\cup_{i\in[m]}B_i$ 使其满足 P.2 的下界约束。因此，$A^{approx}$ 是 P.2 的一个可行解。

引理 7. $A^{approx}$ 是 P.2 的一个可行解。

性能分析 我们首先介绍一个技术引理，该引理指出，如果 $A^{P.3}$ 是 P.3 的一个 $\gamma$-近似解，那么 $f(A^{P.3})$ 至少为 P.2 最优解的 $\gamma$ 倍。该引理可由以下观察得出：P.3 是 P.2 的一个松弛。

引理 8. 假设 $\mathcal{A}$ 是一个用于求解带拟阵约束的非单调子模最大化问题的 $\gamma$-近似算法。设 $O^{P.2}$ 表示 P.2 的最优解，则有 $g(A^{P.3}) \geq \gamma g(O^{P.2})$。

接下来我们证明，在期望意义下，用来自 $\cup_{i\in[m]}B_i$ 的元素扩充 $A^{P.3}$，其效用最多会降低一个 $2/3$ 因子。

引理 9 $\alpha > 1/2$ $\mathbb{E}_{A^{approx}}[g\ A^{approx}\ ] \geq \frac{1}{3}g\ A^{P.3}$ $A^{approx} =$. 假设 , ( ) ( ), 其中 $A^{P.3} \cup (\cup_{i\in[m]}B_i)$。

证明：回顾可知，对于所有满足 $|A^{P.3} \cap V_i| \geq |V_i| - \lfloor\beta|V_i|\rfloor$ 的 $i \in [m]$，$B_i = \emptyset$，因此，将这些 $B_i$ 添加到 $A^{P.3}$ 中不会影响其效用。因此，在证明的其余部分中，我们只关注满足 $|A^{P.3} \cap V_i| < |V_i| - \lfloor\beta|V_i|\rfloor$ 的组 $i \in [m]$。令 $M = \{i \mid |A^{P.3} \cap V_i| < |V_i| - \lfloor\beta|V_i|\rfloor\}$ 表示包含所有此类组的索引的集合，并假设 $M \neq \emptyset$，以避免平凡情况。接下来我们证明，不失一般性地，可以安全地假设 $\min_{i\in M}|V_i| > 1$，即 $M$ 中最小的组至少包含两个项目。为此，我们根据 $B_i = \emptyset$ 的取值分两种情况进行讨论

。如果 $\beta = 1$，那么对于任何满足 $|V_i| = 1$ 的组 $i$， $|A^{P.3} \cap V_i| < |V_i| - \lfloor \beta |V_i| \rfloor$ 都不成立，也就是说，$\min_{i \in M} |V_i| > 1$。如果 $\beta < 1$，那么根据 P.0 中列出的群体公平性约束，我们不允许从满足 $|V_i| = 1$ 的组中选择任何项目。因此，在考虑范围内移除所有大小为 1 的组，不会影响最优解的质量。

在假设 $\min_{i \in M} |V_i| > 1$ 的情况下，我们现在可以证明这个引理。回顾一下，对于每个 $i \in M$ $B_i$ $|V_i| - \lfloor \beta |V_i| \rfloor - |A^{P.3} \cap V_i|$，都是从 $V_i \setminus A^{P.3}$ 中采样得到的特定大小的随机集合。因此，$V_i \setminus A^{P.3}$ 中的每个元素以至多 的概率出现在 $B_i$ 中

$$\frac{|V_i| - \lfloor \beta |V_i| \rfloor - |A^{P.3} \cap V_i|}{|V_i \setminus A^{P.3}|}. \tag{8}$$

接下来，我们给出式 (8) 的一个上界。由于假设 $\alpha > 1/2$，因此有 $\beta \geq \alpha > 1/2$。结合 $\min_{i \in M} |V_i| > 1$ 的假设可得，对所有 $i \in M$，

$$|V_i| - \lfloor \beta |V_i| \rfloor - |A^{P.3} \cap V_i| \leq |V_i| - \lfloor \beta |V_i| \rfloor \leq 2|V_i|/3. \tag{9}$$

此外，

$$\begin{aligned}
&|V_i \setminus A^{P.3}| = |V_i| - |A^{P.3} \cap V_i| && (10)\\
&= (|V_i| - \lfloor \beta |V_i| \rfloor - |A^{P.3} \cap V_i|) + (|V_i| - (|V_i| - \lfloor \beta |V_i| \rfloor)) && (11)\\
&= (|V_i| - \lfloor \beta |V_i| \rfloor - |A^{P.3} \cap V_i|) + \lfloor \beta |V_i| \rfloor && (12)\\
&\geq (|V_i| - \lfloor \beta |V_i| \rfloor - |A^{P.3} \cap V_i|) + |V_i|/3, && (13)
\end{aligned}$$

其中，不等式成立是因为观察到 $\beta > 1/2$。因此，

$$(8) \leq \frac{|V_i| - \lfloor \beta |V_i| \rfloor - |A^{P.3} \cap V_i|}{(|V_i| - \lfloor \beta |V_i| \rfloor - |A^{P.3} \cap V_i|) + |V_i|/3} \leq \frac{2|V_i|/3}{2|V_i|/3 + |V_i|/3} = 2/3, \tag{14}$$

其中，第一个不等式由式 (13) 得出，第二个不等式由式 (9) 以及假设 $|V_i| - \lfloor \beta |V_i| \rfloor - |A^{P.3} \cap V_i| > 0$ 得出。也就是说， $V_i \setminus A^{P.3}$ 中的每个元素以至多 $2/3$ 的概率出现在 $B_i$ 中。引理 3 以及观察到 $g(A^{P.3} \cup \cdot)$ 是子模的这一事实意味着 $\mathbb{E}_{A^{approx}}[g(A^{approx})] = \mathbb{E}_{\cup_{i \in [m]} B_i}[g(A^{P.3} \cup (\cup_{i \in [m]} B_i))] \geq (1 - \frac{2}{3}) g(A^{P.3} \cup \emptyset) = \frac{1}{3} g(A^{P.3})$。 □

引理 8 和引理 9 一起意味着

$$\mathbb{E}_{A^{approx}}[g(A^{approx})] \geq \frac{1}{3} g(A^{P.3}) \geq \frac{\gamma}{3} g(O^{P.2}).$$

根据函数 $g$ 的定义，我们有

$$\mathbb{E}_{A^{approx}}[f(V \setminus A^{approx})] = \mathbb{E}_{A^{approx}}[g(A^{approx})] \geq \frac{\gamma}{3} g(O^{P.2}) = \frac{\gamma}{3} f(OPT)$$

其中，最后一个等式是基于这样的观察：P.2 和 P.0 的最优解具有相同的值。因此，以下主定理成立。

定理 2. 假设 $\mathcal{A}$ 是一种 $\gamma$-近似算法，用于求解带有拟阵约束和 $\alpha > 1/2$ 的非单调子模最大化问题，则有 $\mathbb{E}_{A^{approx}}[f(V \setminus A^{approx})] \geq \frac{\gamma}{3} f(OPT)$。

如果我们采用连续双贪心算法 [11] 作为 $\mathcal{A}$ 来计算 $A^{P.3}$，则可得到一个 $1/e-o(1)$-近似解，即 $\gamma \geq 1/e-o(1)$。结合定理 2，可推出 $\mathbb{E}_{A^{approx}}[f(V \setminus A^{approx})] \geq \frac{1/e-o(1)}{3} f(OPT)$。

## 4 扩展：纳入全局基数约束

本节将扩展 P.0，以纳入全局基数约束。该问题的形式化定义列于 P.A。我们的目标是寻找一个最佳 $S$，使其满足群体公平约束 $(\alpha, \beta)$ 以及额外的基数约束 $c$。

P.A $\max f(S)$
约束条件:
$\lfloor \alpha |V_i| \rfloor \leq |S \cap V_i| \leq \lfloor \beta |V_i| \rfloor, \quad \forall i \in [m]$。
$|S| \leq c$。

### 4.1 当 $\alpha \leq 1/2$ 时的情形

我们首先考虑 $\alpha \leq 1/2$ 时的情形。我们引入一个新的优化问题 P.B，如下所示:

P.B $\max f(S)$
约束条件:
$|S \cap V_i| \leq \lfloor \beta |V_i| \rfloor, \forall i \in [m]$。
$\sum_{i \in [m]} \max\{\lfloor \alpha |V_i| \rfloor, |S \cap V_i|\} \leq c$。

很容易验证，P.B 是 P.A 的一种松弛，因为 P.A 的每个可行解也是 P.B 的可行解。因此，我们得到以下引理。

引理10. 设 $OPT$ 表示 P.A 的最优解，$O^{P.B}$ 表示 P.B 的最优解，则有 $f(O^{P.B}) \geq f(OPT)$。

已有研究表明，P.B 中的约束构成一个拟阵 [10]。结合假设 $f$ 是一个子模函数，可知 P.B 是一个经典问题：在拟阵约束下最大化子模函数。现在，我们准备介绍算法的设计。

1. 使用最先进的算法 $\mathcal{A}$ 求解拟阵约束下的次模最大化问题 P.B，并得到一个解 $A^{P.B}$。

2. 注意， $A^{P.B}$ 不一定是 P.A 的可行解，因为它可能违反某些组的下界约束 $\lfloor\alpha|V_i|\rfloor \leq |S \cap V_i|$。为了使其可行，我们向 $A^{P.B}$中添加额外的项目。具体而言，对于每个满足 $|A^{P.B} \cap V_i| < \lfloor\alpha|V_i|\rfloor$的组 $i \in [m]$，我们的算法通过从 $V_i \setminus A^{P.B}$中随机采样 $\lfloor\alpha|V_i|\rfloor - |A^{P.B} \cap V_i|$ 个项目，选取一个大小为 $\lfloor\alpha|V_i|\rfloor - |A^{P.B} \cap V_i|$的备用集 $B_i$。若 $|A^{P.1} \cap V_i| \geq \lfloor\alpha|V_i|\rfloor$，则定义 $B_i = \emptyset$。
3. 最后，将 $\cup_{i\in[m]}B_i$ 添加到 $A^{P.B}$ 中，以构建最终解 $A^{approx}$，即 $A^{approx} = A^{P.B} \cup(\cup_{i\in[m]}B_i)$。

**算法 3** 当 $\alpha \leq 1/2$ 时求解 P.A 的近似算法

1: 应用 $\mathcal{A}$ 求解 P.B，并得到一个解 $A^{P.B}$
2: 对于每个组 $i \in [m]$，
3: 　如果 $|A^{P.B} \cap V_i| < \lfloor\alpha|V_i|\rfloor$，则
4: 　　从 $V_i \setminus A^{P.B}$中随机选取一个大小为 $\lfloor\alpha|V_i|\rfloor - |A^{P.B} \cap V_i|$ 的备用集 $B_i$
5: 　**else**
6: 　　$B_i \leftarrow \emptyset$
7: $A^{approx} \leftarrow A^{P.B} \cup(\cup_{i\in[m]}B_i)$
8: 返回 $A^{approx}$

该近似算法的伪代码如算法 3 所示。注意，由于 $A^{P.B}$ 满足 P.B 中的第一组约束，因此 $A^{P.B}$ 满足 P.A 的分组上界约束。根据 $B_i$的构造，向 $A^{P.B}$ 中添加 $\cup_{i\in[m]}B_i$ 不会违反 P.A 的分组上界约束，因为必要时会添加 $\cup_{i\in[m]}B_i$ 以满足 P.A 的下界约束。此外，向 $A^{P.B}$ 中添加 $\cup_{i\in[m]}B_i$ 不会违反 P.A 的全局基数约束，因为 $A^{P.B}$ 满足 P.B 中的第二组约束。最后，容易验证，向 $A^{P.B}$ 中添加 $\cup_{i\in[m]}B_i$ 可使其满足 P.A 的下界约束。因此， $A^{approx}$ 是 P.A 的一个可行解。

引理 11. $A^{approx}$ 是 P.A 的一个可行解。

沿用定理1的证明，我们得到以下定理。

定理3. 假设 $\mathcal{A}$ 是一个用于求解拟阵约束下非单调子模最大化问题的 $\gamma$-近似算法，且满足 $\alpha \leq 1/2$，则有$\mathbb{E}_{A^{approx}}[f(A^{approx})] \geq \frac{\gamma}{2}f(OPT)$。

### 4.2 当 $\boldsymbol{\alpha > 1/2}$ 时的情形

接下来考虑 $\alpha > 1/2$ 的情况。回顾一下， $g(\cdot) = f(V \setminus \cdot)$。我们首先给出 P.A 的翻转构造如下:

P.C 最大值 $g(S)$
满足:
$|V_i| - \lfloor\beta|V_i|\rfloor \leq |S \cap V_i| \leq |V_i| - \lfloor\alpha|V_i|\rfloor$, $\forall i \in [m]$。
$|S| \geq n - c$。

假设存在一个 P.C 的 $\gamma$-近似解 $A^{P.C}$，很容易验证，$V \setminus A^{P.C}$ 是 P.A 的一个 $\gamma$-近似解。本节余下部分将重点讨论如何求解 P.C。我们首先引入一个新的优化问题（标记为 P.D），如下：

$$\begin{array}{|l|}\hline \textbf{P.D } \max g(S) \\ \textbf{subject to:} \\ |S \cap V_i| \leq |V_i| - \lfloor \alpha |V_i| \rfloor, \forall i \in [m]. \\ \hline \end{array}$$

P.D 是 P.C 的松弛版本，其中移除了 P.C 中的组级下界约束 $|V_i| - \lfloor \beta |V_i| \rfloor \leq |S \cap V_i|$ 和全局下界约束 $|S| \geq n - c$。因此，我们有以下引理。

引理 12. 令 $O^{P.C}$ 表示 P.C 的最优解，$O^{P.D}$ 表示 P.D 的最优解，则有 $g(O^{P.D}) \geq g(O^{P.C})$。

回顾一下，若 $f$ 是子模的，则 $g$ 必定也是子模的（根据引理 6）。因此，P.D 是一个经典问题：在拟阵约束下最大化次模函数。接下来，我们介绍所设计的算法。

1. 应用用于拟阵约束下次模最大化的最先进算法 $\mathcal{A}$ 求解 P.D，并得到一个解 $A^{P.D}$。2. 注意，$A^{P.D}$ 不一定是 P.C 的可行解，因为它可能违反 P.C 的组级或全局下界约束。我们向 $A^{P.D}$ 中添加额外项目，使其满足约束。具体而言，对于每个组 $i \in [m]$，我们的算法通过随机采样，从 $V_i \setminus A^{P.D}$中选取 $|V_i| - \lfloor \alpha |V_i| \rfloor - |A^{P.D} \cap V_i|$个 $|V_i| - \lfloor \alpha |V_i| \rfloor - |A^{P.D} \cap V_i|$ 项目，构成一个大小为 $|V_i| - \lfloor \alpha |V_i| \rfloor - |A^{P.D} \cap V_i|$的备用集 $B_i$。若 $|V_i| - \lfloor \alpha |V_i| \rfloor - |A^{P.D} \cap V_i| = 0$，则定义 $B_i = \emptyset$。3. 将 $\cup_{i \in [m]} B_i$ 添加到 $A^{P.D}$ 中以构造 $A^{approx}$，即 $A^{approx} = A^{P.D} \cup (\cup_{i \in [m]} B_i)$。返回 $V \setminus A^{approx}$ 作为最终解。

当 $\alpha > 1/2$ 时，P.A 的近似算法 4

1: 应用 $\mathcal{A}$ 求解 P.D，并得到一个解 $A^{P.D}$
2: 对于每个组 $i \in [m]$，执行
3: 　如果 $|A^{P.D} \cap V_i| < |V_i| - \lfloor \alpha |V_i| \rfloor$，则
4: 　　从 $V_i \setminus A^{P.D}$ 中随机选取一个大小为 $|V_i| - \lfloor \alpha |V_i| \rfloor - |A^{P.D} \cap V_i|$ 的备用集 $B_i$
5: 　**else**
6: 　　$B_i \leftarrow \emptyset$
7: 　$A^{approx} \leftarrow A^{P.D} \cup (\cup_{i \in [m]} B_i)$
8: 返回 　$V \setminus A^{approx}$

该近似算法的伪代码如算法 4 所示。请注意，将 $\cup_{i \in [m]} B_i$ 添加到 $A^{P.D}$ 可确保每个组恰好为解贡献$|V_i| - \lfloor \alpha |V_i| \rfloor$ $n - c \leq \sum_{i \in [m]} |V_i| -$ 个 $\lfloor \alpha |V_i| \rfloor$ $A^{P.D} \cup \cup_{i \in [m]} B_i$ 元素。由于 ( )（否则 P.C 没有可行解），( ) 必须满足 P.C 中的所有约束。因此，我们得到以下引理。

引理 13。 $A^{approx}$ 是 P.C 的可行解。

接下来我们分析 $A^{approx}$ 的性能。以下引理表明，将 $\cup_{i\in[m]}B_i$ 添加到 $A^{P.D}$ 中，其效用的期望最多会降低 $2/3$ 倍。

引理 14。假设 $\alpha > 1/2$，我们有 $\mathbb{E}_{A^{approx}}[g(A^{approx})] \geq \frac{1}{3}g(A^{P.D})$。

证明：注意，对于所有满足 $|A^{P.D}\cap V_i| = |V_i| - \lfloor\alpha|V_i|\rfloor$ 的 $i\in[m]$，都有 $B_i = \emptyset$，因此，将这些 $B_i$ 添加到 $A^{P.D}$ 不会影响其效用。因此，在证明的其余部分，我们只关注满足 $|A^{P.D}\cap V_i| < |V_i| - \lfloor\alpha|V_i|\rfloor$ 的组 $i\in[m]$。令 $Z = \{i\in[m] \mid |A^{P.D}\cap V_i| < |V_i| - \lfloor\alpha|V_i|\rfloor\}$ 表示包含所有此类组的索引的集合。我们假设 $Z\neq\emptyset$，以排除平凡情况。接下来我们证明，不失一般性，可以安全地假设 $\min_{i\in Z}|V_i| > 1$，即 $Z$ 中最小的组至少包含两个项目。为证明这一点，我们根据 $\alpha$ 的取值分两种情况讨论。如果 $\alpha = 1$，则对于任何满足 $|V_i| = 1$ 的组 $i$，$|A^{P.D}\cap V_i| < |V_i| - \lfloor\alpha|V_i|\rfloor$ 均不成立。因此，$\min_{i\in Z}|V_i| > 1$。如果 $\alpha < 1$，那么根据 P.A 中列出的群体公平性约束，我们不能从满足 $|V_i| = 1$ 的组中选择任何项目。因此，排除所有大小为一的组不会影响最优解的质量。

在假设 $\min_{i\in Z}|V_i| > 1$ 的情况下，我们现在可以证明这个引理了。回想一下，对于每个 $i\in Z$ $B_i$ $|V_i| - \lfloor\alpha|V_i|\rfloor -$，$|A^{P.D}\cap V_i|$ $V_i\setminus A^{P.D}$ $V_i\setminus A^{P.D}$ 都是从 中抽取的特定大小的随机集合。因此，中的每个元素以至多 的概率出现在 $B_i$ 中

$$\frac{|V_i| - \lfloor\alpha|V_i|\rfloor - |A^{P.D}\cap V_i|}{|V_i\setminus A^{P.D}|}. \tag{15}$$

接下来，我们给出 (15) 的一个上界。由于我们假设 $\alpha > 1/2$ 并且 $\min_{i\in Z}|V_i| > 1$，因此对所有 $i\in M$，都有

$$|V_i| - \lfloor\alpha|V_i|\rfloor - |A^{P.D}\cap V_i| \leq |V_i| - \lfloor\alpha|V_i|\rfloor \leq 2|V_i|/3. \tag{16}$$

此外，

$$\begin{aligned}
&|V_i\setminus A^{P.D}| = |V_i| - |A^{P.D}\cap V_i| && (17)\\
&= (|V_i| - \lfloor\alpha|V_i|\rfloor - |A^{P.D}\cap V_i|) + (|V_i| - (|V_i| - \lfloor\alpha|V_i|\rfloor)) && (18)\\
&= (|V_i| - \lfloor\alpha|V_i|\rfloor - |A^{P.D}\cap V_i|) + \lfloor\alpha|V_i|\rfloor && (19)\\
&\geq (|V_i| - \lfloor\alpha|V_i|\rfloor - |A^{P.D}\cap V_i|) + |V_i|/3, && (20)
\end{aligned}$$

其中，不等式由假设 $\alpha > 1/2$ 和 $|V_i| > 1$ 得出。因此

$$(15) \leq \frac{|V_i| - \lfloor\alpha|V_i|\rfloor - |A^{P.D}\cap V_i|}{(|V_i| - \lfloor\alpha|V_i|\rfloor - |A^{P.D}\cap V_i|) + |V_i|/3} \leq \frac{2|V_i|/3}{2|V_i|/3 + |V_i|/3} = 2/3, \tag{21}$$

其中，第一个不等式由 (20) 得出，第二个不等式由 (16) 以及假设 $|V_i| - \lfloor\alpha|V_i|\rfloor - |A^{P.D}\cap V_i| > 0$ 得出。也就是说， $V_i\setminus A^{P.D}$ 中的每个元素

以至多 $2/3$ 的概率出现在 $B_i$ 中。引理 3 以及 $g(A^{P.D} \cup \cdot)$ 是子模函数这一观察表明，$\mathbb{E}_{A^{approx}}[g(A^{approx})] = \mathbb{E}_{\cup_{i\in[m]}B_i}[g(A^{P.D} \cup (\cup_{i\in[m]}B_i))] \geq (1-\frac{2}{3})g(A^{P.D} \cup \emptyset) = \frac{1}{3}g(A^{P.D})$。□

假设 $\mathcal{A}$ 是一个用于求解拟阵约束下非单调子模最大化问题的 $\gamma$ 近似算法，则有

$$\mathbb{E}_{A^{approx}}[g(A^{approx})] \geq \frac{1}{3}g(A^{P.D}) \geq \frac{\gamma}{3}g(O^{P.D})$$

其中，第一个不等式由引理 $14$ $g$ $\mathbb{E}_{A^{approx}}[g\ A^{approx}] \geq \frac{\gamma}{3}g\ O^{P.C}$ $\geq$ 得出。结合（$O^{P.D}$）（如 $g\ O^{P.C}$ $12$ 引理中所证明），可推出（） （）。根据函数 $g$ 的定义，有

$$\mathbb{E}_{A^{approx}}[f(V \setminus A^{approx})] = \mathbb{E}_{A^{approx}}[g(A^{approx})] \geq \frac{\gamma}{3}g(O^{P.C}) = \frac{\gamma}{3}f(OPT)$$

其中，最后一个等式源于以下观察：P.A 和 P.C 的最优解具有相同的值。因此，以下主定理成立。

定理 4. 假设 $\mathcal{A}$ 是一个用于求解拟阵约束下非单调子模最大化问题的 $\gamma$ 近似算法，且 $\alpha > 1/2$，则有 $\mathbb{E}_{A^{approx}}[f(V \setminus A^{approx})] \geq \frac{\gamma}{3}f(OPT)$。

## 5 结论

本文全面研究了群体公平性约束下的非单调子模最大化问题。我们的主要贡献是为该问题开发了多种常数因子近似算法。未来，我们计划拓展研究，探索其他公平性指标。

## 参考文献

1. Biddle, D.：《不利影响与测试验证：有效且可辩护的就业测试实践指南》。Routledge（2017年） 2. Buchbinder, N., Feldman, M., Naor, J., Schwartz, R.：基数约束下的子模最大化。载于：第二十五届 ACM-SIAM 离散算法研讨会论文集。第 1433–1452 页。SIAM（2014年） 3. Celis, E., Keswani, V., Straszak, D., Deshpande, A., Kathuria, T., Vishnoi, N.：基于 DPP 的公平多样化数据摘要。载于：国际机器学习会议。第 716–725 页。PMLR（2018年） 4. Celis, L.E., Huang, L., Vishnoi, N.K.：公平性约束下的多赢家投票。载于：第 27 届国际人工智能联合会议论文集。第 144–151 页（2018年） 5. Chierichetti, F., Kumar, R., Lattanzi, S., Vassilvtiskii, S.：拟阵、匹配与公平性。载于：第二十二届国际人工智能与统计会议。第 2212–2220 页。PMLR（2019年）

6. Das, A., Kempe, D.：线性回归中的子集选择算法。载于：第四十届 ACM 计算理论年度研讨会论文集。第 45–54 页（2008年） 7. Dueck, D., Frey, B.J.：用于无监督图像分类的非度量亲和传播。载于：2007年 IEEE 第 11 届国际计算机视觉会议。第 1–8 页。IEEE（2007年） 8. Dwork, C., Hardt, M., Pitassi, T., Reingold, O., Zemel, R.：通过意识实现公平性。载于：第三届理论计算机科学创新会议论文集（2012年） 9. El-Arini, K., Guestrin, C.：超越关键词搜索：发现相关科学文献。载于：第 17 届 ACM SIGKDD 知识发现与数据挖掘国际会议论文集。第 439–447 页（2011年） 10. El Halabi, M., Mitrovi´c, S., Norouzi-Fard, A., Tardos, J., Tarnawski, J.M.：流式次模最大化中的公平性：算法与难度。神经信息处理系统进展 33，13609–13622（2020年） 11. Feld man, M., Naor, J., Schwartz, R.：用于次模最大化的统一连续贪心算法。载于：2011年 IEEE 第 52 届计算机科学基础年会。第 570–579 页。IEEE（2011年） 12. Golovin, D., Krause, A.：自适应次模性：主动学习与随机优化中的理论和应用。《人工智能研究杂志》42，427–486（2011年） 13. Gotovos, A., Karbasi, A., Krause, A.：非单调自适应次模最大化。载于：第二十四届国际人工智能联合会议（2015年） 14. Gu, S., Gao, C., Wu, W.：用于非单调 DR次模最大化的二分搜索双重贪心算法。载于：《信息与管理中的算法方面》：第 16 届国际会议 AAIM 2022，中国广州，2022年8月13–14日，论文集。第 3–14 页。Springer（2022年） 15. Joseph, M., Kearns, M., Morgenstern, J. H., Roth, A.：学习中的公平性：经典老虎机与上下文老虎机。神经信息处理系统进展 29（2016年） 16. Mirzasoleiman, B., Badanidiyuru, A., Karbasi, A.：快速约束次模最大化：个性化数据摘要。载于：国际机器学习大会。第 1358–1367 页（2016年） 17. Monroe, B.L.：完全比例代表制。《美国政治科学评论》89(4)，925–940（1995年） 18. Shi, G., Gu, S., Wu, W.：带有两类约束的 k-次模最大化。《离散数学、算法与应用》13(04)，2150036（2021年） 19. Sipos, R., Swaminathan, A., Shivaswamy, P., Joachims, T.：使用子模词语覆盖进行时间语料摘要。载于：第 21 届 ACM 信息与知识管理国际会议论文集。第 754–763 页（2012年） 20. Tang, S., Yuan, J.：带部分反馈的影响力最大化。《运筹学快报》48(1)，24–28（2020年） 21. Tang, S., Yuan, J.：自适应正则化次模最大化。载于：第 32 届国际算法与计算研讨会（ISAAC 2021）。Schloss Dagstuhl-Leibniz-Zentrum f¨ur Informatik（2021年） 22. Tang, S., Yuan, J.：自适应次模最大化中的群体平等。arXiv 预印本 arXiv:2207.03364（2022年） 23. Tang, S., Yuan, J.：超越子模性：通过随机化实现公平子集选择的统一框架。载于：审稿中（2023年）

24. Tang, S., Yuan, J., Mensah-Boateng, T.：通过随机化实现次模最大化中的长期公平性。载于：审稿中（2023年） 25. Tsang, A., Wilder, B., Rice, E., Tambe, M., Zick, Y.：影响力最大化中的组公平性。arXiv 预印本 arXiv:1903.00967（2019年） 26. Zafar, M.B., Valera, I., Rogriguez, M.G., Gummadi, K.P.：公平性约束：实现公平分类的机制。载于：人工智能与统计学。第 962–970 页。PMLR（2017年）