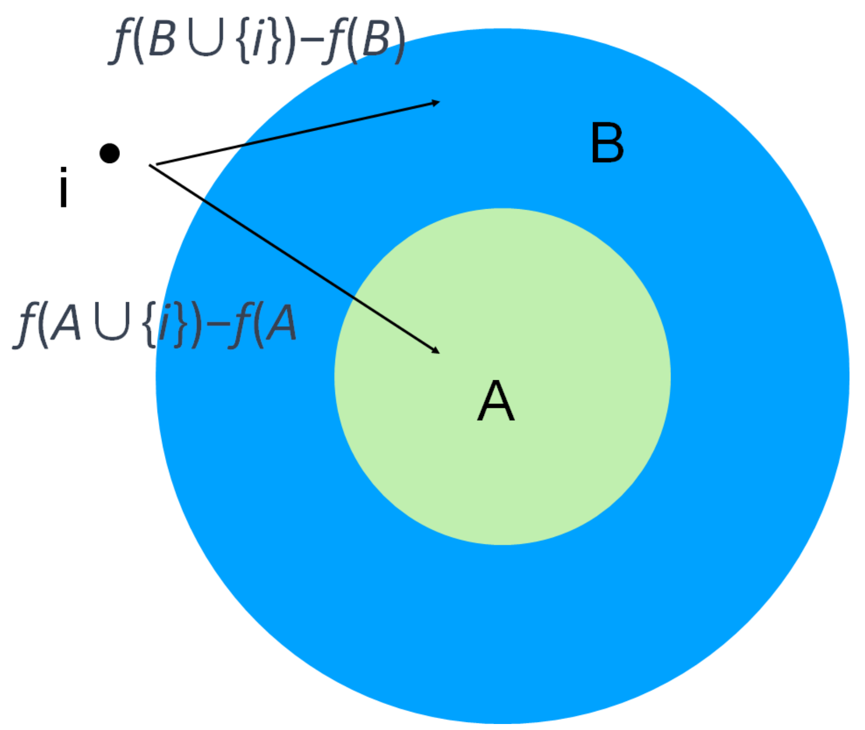
Submodular Cost Submodular Cover with Fairness.

Wenjing Chen, Zhitong Chen, and Shuo Xing

Intro to Submodular Function and Fairness Problem

Submodularity

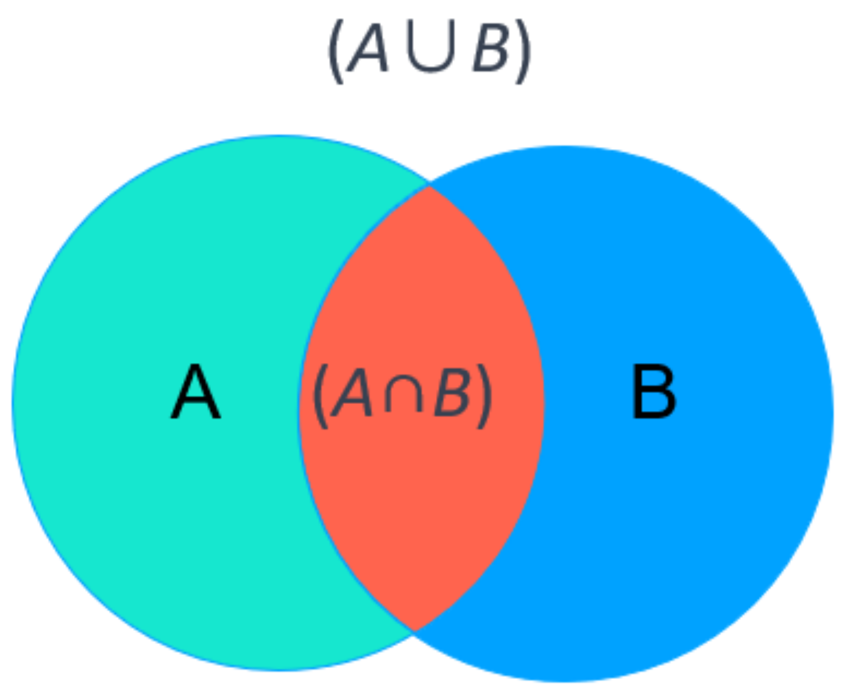
- A discrete function $f : 2^U \to \mathbb{R}$ is submodular if $\forall A \subseteq B \subseteq U$, and $\forall i \in U/B$, it holds
- $$f(A \cup \{i\}) - f(A) \geq f(B \cup \{i\}) - f(B)$$
- (Diminishing return property)



- Denote $\Delta f(A, i) = f(A \cup i) - f(A)$, then $\Delta f(A, i) \geq \Delta f(B, i)$

Submodularity

- submodular functions
- $$f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$$
- Proof:



What is fairness

- **Definition:**
    - Fairness ensures that individuals from different backgrounds and groups receive unbiased decision-making.
    - In the realm of algorithms, fairness involves data processing and decision-making output that is fair and non-discriminatory to all users.

What is fairness

- **Key points:**
  - Statistical equality, error rate balance, and equality of opportunity are key measures of algorithmic fairness.
- **Challenges:**
  - Including dealing with unbalanced data, correcting historical biases and ensuring group representation.

Motivation: Why study fairness

- **Background:**
  - As algorithmic decision tools proliferate in society, their influence in sensitive areas is also growing.
- **Social impact of the algorithm:**
  - Potential for bias and discrimination in automated decision especially in sensitive domains: voting, hiring, criminal justice, access to credit, etc.

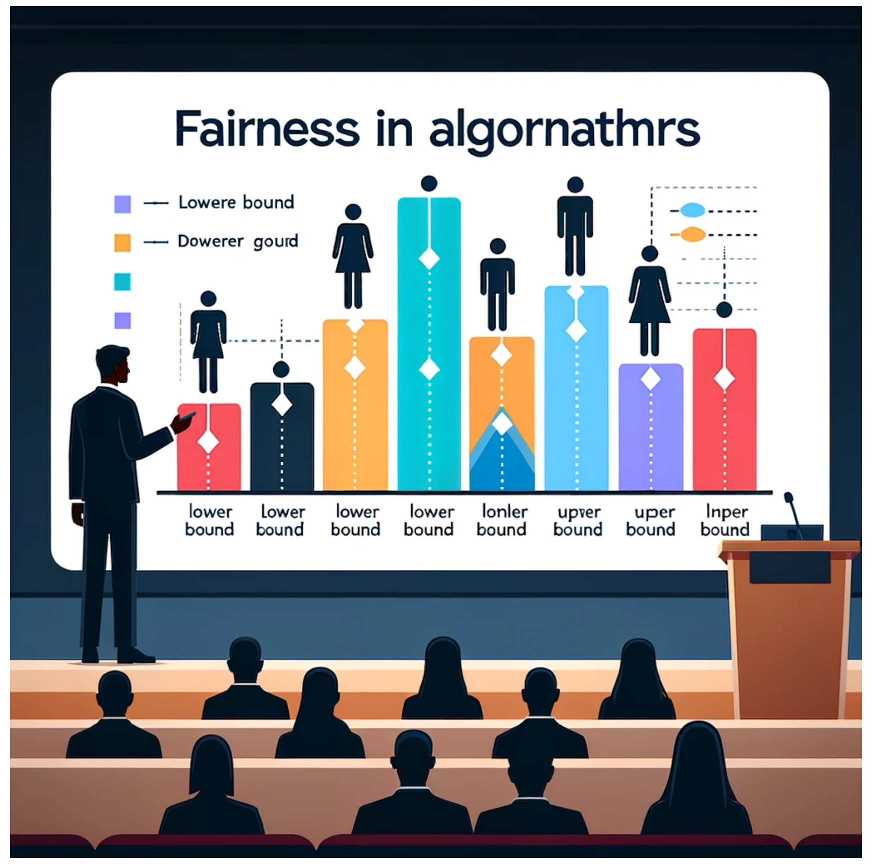
Motivation: Why study fairness

- **The Law and Ethics of fairness and research evaluation:**
  - 1. Public expectations and legal requirements.
  - 2. Enhance the fairness and trust of decision systems.

Fairness definition and constraint

- **The necessity of fairness:**
  - Fairness is introduced into the algorithm to ensure that all groups are represented fairly.
- **Define lower and upper bounds:**
  - Lower bound l_c ensure minimum representation of each group and the upper bound u_c prevents over-representation.

Fairness definition and constraint



- **Example of fairness constraints:**
  - In the recruitment process, the lower and upper bounds of gender or race representation are set according to the proportion of the applicant group.

Practical application of fairness constraints

- How can we avoid exacerbating social inequality when relying on data-driven decision-making?
    - Group—> unfairly represented —> socio-economic opportunities unequal
    - Ensure groups get a fair shot in automated decision-making: credit approval, hiring, education or health diagnostics.
- How to set upper and lower bound
    - For example: if a group is 30% of the overall population, that group should be at least 30% (lower bound), Higher will lead to over-representation.

Submodular Optimization Problem with Fairness

Fairness Set Up

- Assume the ground set V is colored so and each element has exactly one color. Index color by $c = 1,2,\cdots, C$ and denote by V_c the set of the elements of color c.

- Thus, we have partition $V = V_1 \cap V_2 \cap \cdots \cap V_C$.

- Denote by $\mathscr{F}$ the set of solutions feasible under the fairness and cardinality constraints, i.e.

    - $\mathscr{F} = \{S \subseteq V : |S| \leq k, \ |S \cap V_c| \in [l_c, u_c] \forall c \in [C]\}$, where l_c and u_c are the lower bound and upper bound on the number of elements with color c

Submodular Maximization under Fairness

- The problem of submodular maximization under fairness constraint is therefore formulated as

- $\max_{S \in E} \; f(S)$
- s.t. $l_c \leq |S \cap V_c| \leq u_c \qquad \forall c \in [C]$
- $|S| \leq k$

Extendable Set

- ***Definition 1*** $S \subseteq V$ is extendable if it is a subset $S \subseteq S'$ of some feasible solution $S' \in \mathcal{F}$.
- ***Lemma 1*** A set $S \subseteq V$ is extendable if and only if
- $|S \cap V_c| \leq u_c$ for all $c = 1, \cdots, C$ and $\sum_{c=1}^{C} \max(|S \cap V_c|, l_c) \leq k$.

Extendable Sets: Results

- ***Lemma 2.*** If f is monotone, then SM under fairness constraint is equivalent to SM under extendable sets constraint
- ***Lemma 3*** If $\mathcal{I}$ is the collection of all the extendable sets in the ground set V, then $(V, \mathcal{I})$ is a matroid.
- Therefore, the problem is equivalent to SM under matroid constraint.

Matroid

- ***Definition 2*** A matroid is a pair $(E, \mathscr{M})$ where E is a finite set (called ground set) and $\mathscr{M}$ is a family of subsets of E (called the independent sets) with the following properties:
  - 1. The empty set is in $\mathscr{M}$, I.e., $\emptyset \in \mathscr{M}$
  - 2. For any subset A, B, if $A \in \mathscr{M}$, and B is a subset of A, then $B \in \mathscr{M}$. (downward property)
  - 3. If A and B are are two independent sets and $|A| > |B|$, then there exists $x \in A \backslash B$ such that $B \cup \{x\} \in \mathscr{M}$.

Submodular Maximization Problem with Fairness

- The problem of submodular maximization under fairness is thus equivalent to:
  - $\max_{S \subseteq U} f(S)$
  - s.t. $S \in \mathcal{I}$, where $\mathcal{I}$ is the collection of all the extendable sets
- Thus, by applying a greedy algorithm, we can obtain a $1/2$ approximation (best known result), i.e. $f(S) \geq 1/2 \ OPT$.

Submodular Cover Problem

Submodular Cover: Cardinality constraint

- (Submodular Maximization)
  - $\max_{S \subseteq U} \; f(S)$
  - s.t. $|S| \leq k$

- (Submodular Cover)
  - $\min_{S \subseteq U} \; |S|$
  - s.t. $f(S) \geq \tau$

Submodular Cover: Fairness Constraint

- (Submodular Maximization)
- $\max_{S \subseteq U} \; f(S)$
- s.t. $|S| \leq k$
- $l_c \leq |S \cap V_c| \leq u_c \,, \forall c \in [C]$

- (Submodular Cover)
- $\min_{S \subseteq U} \; |S|$
- s.t. $f(S) \geq \tau$

SC under fairness constraint? $|S \cap V_c|$ Proportional to $|S|$

Submodular Cost Submodular Cover with Fairness

- In this project, we consider submodular cover with fairness constraint
    - Objective: minimize the cardinality
    - Constraint: (1). $|S \cap V_c|$ should be proportional to $|S|$; (2). $f(S) \geq \tau$.

Submodular Cost Submodular Cover with Fairness

- The fairness constraint is:
- $\mathcal{F} = \{S \subseteq V : l_c|S| \leq |S \cap V_c| \leq u_c|S|, \forall c \in [C]\}$,
- l_c ,u_c are the lower bound and upper bound on the proportion of the number of elements in group c

Submodular Cost Submodular Cover with Fairness

- Therefore, the problem as following is considered:
  - (Submodular Cost Submodular Cover with Fairness)
  - $\min_{S \subseteq U} \quad |S|$
  - s.t. $\quad l_c |S| \leq |S \cap V_c| \leq u_c |S|, \forall c \in [C]$
  - $f(S) \geq \tau$

Proposed Algorithms and Main Results

Intuition of algs

- The problem of Submodular Cover is a dual problem of Submodular Maximization

- (Submodular Maximization)
- $\max_{S \subseteq U} \quad f(S)$
- s.t. $l_c \leq |S \cap V_c| \leq u_c, \forall c \in [C]$
- $|S| \leq k$

- (Submodular Cover)
- $\min_{S \subseteq U} \quad |S|$
- s.t. $l_c |S| \leq |S \cap V_c| \leq u_c |S|, \forall c \in [C]$
- $f(S) \geq \tau$

- Greedy for SM achieves 1/2 approximation ratio. If we know |OPT|, then we can use alg for SM as subroutine.

Algorithm 1

Algorithm 1 `greedy-fairness-known-OPT`

Input: ϵ
Output: $S \in V$
$S \leftarrow \emptyset$
$\mathcal{I}' \leftarrow \{S \subseteq V : |S \cap V_c| \leq u_c/\epsilon \cdot |O|, \forall c \in [C]$ and $\sum_{c=1}^{C} \max(|S \cap V_c|, l_c/\epsilon \cdot |O|) \leq |O|/\epsilon\}$, where O is the optimal solution
while $|S| < |O|/\epsilon$ **do**
    $U \leftarrow \{x \in U | S \cup \{x\} \in \mathcal{I}'\}$
    $u \leftarrow \text{argmax}_{x \in U} \Delta f(S, x)$ ← Greedy
    $S \leftarrow S \cup \{u\}$
return S

S is the current solution, O is the optimal solution

Theoretical Results for Algorithm 1

- ***Theorem 1.*** *Algorithm 1* produces a solution with $(1/\epsilon, 1-\epsilon)$-bicrateria approximation in at most $O(n|OPT|/\epsilon)$ queries of f.
- Comments on *Theorem 1*:
  - $(1/\epsilon, \epsilon)$-bicriteria approximation: $f(S) \geq (1-\epsilon)\tau,$ and $|S| \leq |OPT|/\epsilon$

Algorithm 2

Algorithm 2 `greedy-fairness-by-guesses`

Input: ϵ and α
Output: $S \in V$
$g \leftarrow 1 + \alpha$
while $f(S) \leq (1 - \epsilon)\tau$ **do**
    $S \leftarrow \emptyset$
    $\mathcal{I}'_g \leftarrow \{S \subseteq V : |S \cap V_c| \leq u_c/\epsilon \cdot g, \forall c \in [C]$ and $\sum_{c=1}^{C} \max(|S \cap V_c|, l_c/\epsilon \cdot g) \leq g/\epsilon\}$
    while $|S| < g/\epsilon$ **do**
        $U \leftarrow \{x \in U | S \cup \{x\} \in \mathcal{I}'_g\}$
        $u \leftarrow \text{argmax}_{x \in U} \Delta f(S, x)$
        $S \leftarrow S \cup \{u\}$
    $g \leftarrow (1 + \alpha)g$
return S

S is the current solution, V is the ground set, g is the guess of the size of the optimal solution and τ is the fixed threshold.

Theoretical Results for Algorithm 2

- ***Theorem 2*** *Algorithm 2* produces a solution with $\big((1+\alpha)/\epsilon, 1-\epsilon\big)$ -bicrateria approximation in at most $O(n|OPT|\log|OPT|/\epsilon)$ queries of f.
- Comments on *Theorem 2*:
  - $\big((1+\alpha)/\epsilon, 1-\epsilon\big)$-bicriteria approximation: $f(S) \geq (1-\epsilon)\tau$, and $|S| \leq (1+\alpha)|OPT|/\epsilon$
  - Needs at most $\log n$ rounds of guesses, where n is the size of the ground set V

Thank You！