Differentially Private Decomposable Submodular Maximization

Anamay Chaturvedi†, Huy Lê Nguyễn†, Lydia Zakynthinou†

†Northeastern University

Problem

We consider the problem of **differentially private decomposable submodular maximization**.

- Submodular functions $f : 2^V \to \mathbb{R}_+$ have **diminishing returns:**

$$S \subset T, u \notin T \Rightarrow f(S \cup \{u\}) - f(S) \geq f(T \cup \{u\}) - f(T).$$

- **Decomposable** submodular:

$$f_D(S) = \sum_{\text{agent} p \in D} f_p(S)$$

- Want to **privately maximize decomposable submodular functions** subject to a **matroid constraint**.
- **Central model of differential privacy**.
- **Motivation:** Posed by Papadimitriou, Schapira, and Singer 2008, derived from notion of **social welfare maximization**.
- **Applications: – exemplar-based clustering – image summarization – recommender systems – document and corpus summarization**

Approach

Continuous greedy methods of Vondrák 2008 and Feldman, Naor, and Schwartz 2011.

- Maximise multilinear relaxation of f_D

$$F_D(x) = \sum_{S \subset 2^V} f_D(S) \prod_{i \in S} x_i \prod_{i \notin S} (1 - x_i).$$

- T **rounds**. Iteratively pick feasible i maximising $F(x)$ on **increasing x_i** by a **$1/T$ step** (monotone f_D); **$(1 - x_i)/T$ step** (non-monotone f_D); .
- x in **convex hull** of **feasible sets**. **Swap-rounding** of Chekuri, Vondrak, and Zenklusen 2010 returns feasible solution with good utility.

Highlights

- Greedy picks via exponential mechanism - following Gupta et al. 2010 get **loss in privacy** independent of **number of rounds**.
- Estimate F_D by sampling and **sharing randomness between rounds** - this avoids additional utility loss in each round.
- Directly replacing each round of continuous greedy by the private greedy does not work.
- Additive error $\sim$ $O\left(\frac{r}{\epsilon} \log nr \cdot \log \frac{1}{\delta}\right)$ close to known lower bound of $O\left(\frac{r}{\epsilon} \log n/r\right)$.

Results

- **Monotone rank r matroid-constrained case** we are (ϵ, δ)**-private** using T rounds with expected utility

$$(1 - 1/e - O(1/T))f(\mathtt{OPT}) - O\left(\frac{rT}{\epsilon} \log nrT \cdot \log \frac{1}{\delta}\right)$$

- Analogous **non-monotone case**:

$$(1/e - O(1/T))f(\mathtt{OPT}) - O\left(\frac{rT}{\epsilon} \log nrT \cdot \log \frac{1}{\delta}\right)$$

Related work:

- Work by Gupta et al. 2010 and Mitrovic et al. 2017 used a discrete greedy algorithm. By adapting continuous methods improve multiplicative factor **from (1/2)** (Mitrovic et al. 2017) **to** $(1 - 1/e - O(1/T))$ in the **monotone case**.
- Rafiey and Yoshida 2020 also adapt continuous greedy methods but obtain significantly higher additive error of $nr^7 \log n/\epsilon^3$.

Experimental results

We replicate the **Uber location selection** experiment of Mitrovic et al. 2017.

- Given a set of pick-up locations in Manhattan, the goal is to **pick locations close to pick-ups** while **private with respect to pick-ups**.
- Scaled ℓ_1 distance between location l and pick-up p:

$$M(l, p) = \frac{|l_x - p_x| - |l_y - p_y|}{C} \leq 1.$$

- Utility of locations S evaluated on pick-ups D:

$$f_D(S) = \sum_{p \in D} \left(1 - \min_{l \in S} M(l, p)\right) = |D| - \sum_{p \in D} \min_{l \in S} M(l, p). \quad (1)$$

f_D is monotone decompasable submodular function. We conduct two experiments:

- a **rank constrained location selection** for 100 agents at a time. Comparison with more general algorithm of Mitrovic et al. 2017 that uses the composition laws of privacy instead of the Gupta privacy analysis.
- simple 3-element **partition matroid** instance measuring per-capita utility versus dataset size. Comparison with discrete method for matroids of Mitrovic et al. 2017