Highlights

I. Submodular maximization in the streaming setting with robustness to removals.
II. Constant-factor guarantee for any number of removals.
III. Uses only at most polylog more space compared to the optimum.

Setup

Submodularity of a set function $f : 2^V \rightarrow \mathbb{R}$:

$$f(X \cup \{e\}) - f(X) \geq f(Y \cup \{e\}) - f(Y)$$

whenever $X \subseteq Y$ and $e \notin Y$.

Submodular maximization under cardinality constraint $k$:

$$\text{OPT}(X, k) := \arg \max_{Y \subseteq X : |Y| = k} f(Y)$$

The elements of $V$ arrive in the streaming fashion in arbitrary order.

An arbitrary set $E \subseteq V$ is removed from $V$ after the stream ends.

Robust monotone submodular maximization: find $S \subseteq V$ such that

1. $|S| \leq m$, and
2. $f(\text{OPT}(S \setminus E, k)) \geq c \cdot f(\text{OPT}(V, k))$

for a parameter $c$ and any $E$ not known a priori.

Main result

- [MKK17]: Algorithm with $c=0.499$ and $m=\tilde{O}(k |E|)$.
- [BMKK14]: If $E=\emptyset$, algorithm with $c=0.499$ and $m=\tilde{O}(k)$.

Our result: Algorithm with $c=0.148$ and $m=\tilde{O}(k + |E|)$.

Applications

Robust dominating set

Find a set of nodes $S$ of size at most $t$ that maximizes

$$|S| = \max \{ |A| : A \subseteq \text{OPT}(S \setminus E, k) \cup \text{OPT}(S \cap E, k) \}$$

$|A|$: denotes all the neighborhood nodes of nodes in $A$.

Interactive personalized movie recommendation:

Given a feature vector $u$ of a user and feature vectors $v_j$ for a set of movies $M$, find a set of movies $S \subseteq M$ that maximizes

$$\sum_{z \in S} u_z \cdot v_j + \alpha \sum_{m \in M} \max_{z \in S} v_m$$

The set of movies $S$ is then used to provide, on the user’s request, movies with specific properties (genre, year, not-seen, etc.). The movies NOT falling in the category of the user’s request are treated as the set $E$.

Setting

First phase: select $S$ from the stream

$$S = \{ \text{HOLLYWOOD}, \text{DOWNTOWN}, \text{CITY}, \text{HOLLYWOOD} \}$$

Second phase: given $E$, run Greedy to approximate $\text{OPT}(S \setminus E, k)$

$$E = \{ \text{HOLLYWOOD}, \text{DOWNTOWN}, \text{CITY} \}$$

$$\text{OPT}(\{ \text{HOLLYWOOD}, \text{DOWNTOWN}, \text{CITY} \}, k)$$

STAR-T Algorithm

Structure of the robust set $S$:
- # of partitions: $\log k$
- partition $i$ contains: $(k/2^i)$ buckets with at most $2^i$ items

Streaming Algorithm to create $S$:
- add the item $e$ to the first non-full bucket $B$ if

$$f(e|B) \geq \frac{r}{2^i}$$

$i$: partition of the bucket $B$

$r$: threshold value that depends on $f(\text{OPT}(V \setminus E, k))$

$f(\text{OPT}(V \setminus E, k))$ is approximated by extending the techniques of [BMKK14]

- if such bucket does not exist discard the item.

Query Stage:
- for a given $E$ return $k$ elements greedily selected from $S \setminus E$

Numerical results

Robust dominating set:

Interactive personalized movie recommendation:

References
