Evolutionary Submodular Competition: Problem Definitions

Aneta Neumann, Chao Qian, Frank Neumann, Viet Anh Do

March 31, 2022

We define the different submodular problems included in the competition. We denote by c(x) the cost of a solution x and $f(x)$ the value of the (submodular) function applied to x.

1 Monotone Submodular Problems

1.1 Maximum Coverage

Given an undirected weighted graph $G = (V, E, c)$ with costs $c\colon V \to \mathbb{R}_{\geq 0}$ on the vertices. We denote by $N(V') = \{v_i \mid \exists e \in E : e \cap V' \neq \emptyset \wedge e \cap v_i \neq \emptyset\}$ the set of all nodes of V' and their neighbors in G.

For a given search point $\boldsymbol{x} \in \{0,1\}^n$ where $n = |V|$, we have $V'(\boldsymbol{x}) = \{v_i \mid x_i = 1\}$ and $c(\boldsymbol{x}) = \sum_{v \in V'(\boldsymbol{x})} c(v)$.

1.1.1 Deterministic Setting

In the deterministic setting, the goal is to maximize

$$f(\boldsymbol{x}) = N(V'(\boldsymbol{x}))$$

under the constraint that $c(\boldsymbol{x}) \leq B$ holds.

The fitness of a search point $\boldsymbol{x}$ is given as the 2-dimensional vector $g(\boldsymbol{x}) = (f'(\boldsymbol{x}), c(\boldsymbol{x}))$ where

$$f'(\boldsymbol{x}) = \begin{cases} f(\boldsymbol{x}) & c(\boldsymbol{x}) \leq B \\ B - c(\boldsymbol{x}) & c(\boldsymbol{x}) > B \end{cases}$$

1.1.2 Chance constrained setting

For the chance constraint setting, we assume that costs are chosen independently and uniformly at random in $c(v) \in [a(v) - \delta, a(v) + \delta]$ where $a(v)$ is the expected cost of node v and δ is a parameter specifying the level of uncertainty which is the same for all nodes. We have $f(\boldsymbol{x}) = N(V'(\boldsymbol{x}))$ as defined in the deterministic setting. We take the cost of a node for the given benchmark instance as expected cost and consider the uncertainty parameterized by δ.

The goal is to maximize $f(\boldsymbol{x})$ under the chance constraint that $Prob(c(\boldsymbol{x}) > B) \leq \alpha$ holds. Here $\alpha \leq 1/2$ is a given parameter that limits the probability of a constraint violation. Let $a(\boldsymbol{x}) = \sum_{v \in V'(\boldsymbol{x})} a(v)$ be the expected cost of solution $\boldsymbol{x}$ and $v(\boldsymbol{x}) = |\boldsymbol{x}|_1 \cdot \delta^2/3$ be its variance.

Based on tail bounds, we consider the following cost functions to make sure that the chance constraint is met.

Cost function based on Chebyshev' inequality We use the cost function

$$c_{Cheby}(\boldsymbol{x}) = a(\boldsymbol{x}) + \sqrt{\frac{1-\alpha}{\alpha} \cdot v(\boldsymbol{x})} = a(\boldsymbol{x}) + \delta \cdot \sqrt{\frac{1-\alpha}{3\alpha} \cdot |\boldsymbol{x}|_1}$$

based on Chebyshev's inequality.

The fitness of a search point $\boldsymbol{x}$ using c_{Cheby} is given as the 2-dimensional vector $g_{Cheby}(\boldsymbol{x}) = (f'_{Cheby}(\boldsymbol{x}), c_{Cheby}(\boldsymbol{x}))$ where

$$f'_{Cheby}(\boldsymbol{x}) = \begin{cases} f(\boldsymbol{x}) & c_{Cheby}(\boldsymbol{x}) \leq B \\ B - c_{Cheby}(\boldsymbol{x}) & c_{Cheby}(\boldsymbol{x}) > B \end{cases}$$

Cost function based on Chernoff bounds We use the cost function

$$c_{Cher}(\boldsymbol{x}) = a(\boldsymbol{x}) + \delta \cdot \sqrt{\ln(1/\alpha) \cdot 2|\boldsymbol{x}|_1}$$

based on Chernoff bounds. The chance constraint is met if $c_{Cher}(\boldsymbol{x}) \leq B$ holds.

The fitness of a search point $\boldsymbol{x}$ using c_{Cher} is given as the 2-dimensional vector $g_{Cher}(\boldsymbol{x}) = (f'_{Cher}(\boldsymbol{x}), c_{Cher}(\boldsymbol{x}))$ where

$$f'_{Cher}(\boldsymbol{x}) = \begin{cases} f(\boldsymbol{x}) & c_{Cher}(\boldsymbol{x}) \leq B \\ B - c_{Cher}(\boldsymbol{x}) & c_{Cher}(\boldsymbol{x}) > B \end{cases}$$

1.2 Maximum Influence

Let a directed graph $G(V, E)$ represent a social network, where each node is a user and each edge $(u, v) \in E$ has a probability $p_{u,v}$ representing the strength of influence from user u to v.

A fundamental propagation model is independence cascade. Starting from a seed set X, it uses a set A_t to record the nodes activated at time t, and at time $t+1$, each inactive neighbor v of $u \in A_t$ becomes active with probability $p_{u,v}$. This process is repeated until no nodes get activated at some time. The set of nodes activated by propagating from X is denoted as $IC(X)$, which is a random variable.

For a given search point $\boldsymbol{x} \in \{0,1\}^n$ where $n = |V|$, we have $V'(\boldsymbol{x}) = \{v_i \mid x_i = 1\}$ and $c(\boldsymbol{x}) = \sum_{v \in V'(\boldsymbol{x})} c(v)$.

1.2.1 Deterministic Setting

In the deterministic setting, the goal is to maximize the expected number of nodes activated by propagating from $V'(\boldsymbol{x})$, i.e.,

$$f(\boldsymbol{x}) = E[|IC(V'(\boldsymbol{x}))|]$$

under the constraint that $c(\boldsymbol{x}) \leq B$ holds.

The fitness of a search point $\boldsymbol{x}$ is given as the 2-dimensional vector $g(\boldsymbol{x}) = (f'(\boldsymbol{x}), c(\boldsymbol{x}))$ where

$$f'(\boldsymbol{x}) = \begin{cases} f(\boldsymbol{x}) & c(\boldsymbol{x}) \leq B \\ B - c(\boldsymbol{x}) & c(\boldsymbol{x}) > B \end{cases}$$

1.2.2 Chance constrained setting

For the chance constraint setting, we assume that costs are chosen independently and uniformly at random in $c(v) \in [a(v) - \delta, a(v) + \delta]$ where $a(v)$ is the expected cost of node v and δ is a parameter specifying the level of uncertainty which is the same for all nodes. We have $f(\boldsymbol{x}) = E[|IC(V'(\boldsymbol{x}))|]$ as defined in the deterministic setting. We take the cost of a node for the given benchmark instance as expected cost and consider the uncertainty parameterized by δ.

The goal is to maximize $f(\boldsymbol{x})$ under the chance constraint that $Prob(c(\boldsymbol{x}) > B) \leq \alpha$ holds. Here $\alpha \leq 1/2$ is a given parameter that limits the probability of a constraint violation. Let $a(\boldsymbol{x}) = \sum_{v \in V'(\boldsymbol{x})} a(v)$ be the expected cost of solution $\boldsymbol{x}$ and $v(\boldsymbol{x}) = |\boldsymbol{x}|_1 \cdot \delta^2/3$ be its variance.

Based on tail bounds, we consider the following cost functions to make sure that the chance constraint is met.

Cost function based on Chebyshev' inequality We use the cost function

$$c_{Cheby}(\boldsymbol{x}) = a(\boldsymbol{x}) + \sqrt{\frac{1-\alpha}{\alpha}} \cdot v(\boldsymbol{x}) = a(\boldsymbol{x}) + \delta \cdot \sqrt{\frac{1-\alpha}{3\alpha}} \cdot |\boldsymbol{x}|_1$$

based on Chebyshev's inequality.

The fitness of a search point $\boldsymbol{x}$ using c_{Cheby} is given as the 2-dimensional vector $g_{Cheby}(\boldsymbol{x}) = (f'_{Cheby}(\boldsymbol{x}), c_{Cheby}(\boldsymbol{x}))$ where

$$f'_{Cheby}(\boldsymbol{x}) = \begin{cases} f(\boldsymbol{x}) & c_{Cheby}(\boldsymbol{x}) \leq B \\ B - c_{Cheby}(\boldsymbol{x}) & c_{Cheby}(\boldsymbol{x}) > B \end{cases}$$

Cost function based on Chernoff bounds We use the cost function

$$c_{Cher}(\boldsymbol{x}) = a(\boldsymbol{x}) + \delta \cdot \sqrt{\ln(1/\alpha) \cdot 2|\boldsymbol{x}|_1}$$

based on Chernoff bounds. The chance constraint is met if $c_{Cher}(\boldsymbol{x}) \leq B$ holds.

The fitness of a search point $\boldsymbol{x}$ using c_{Cher} is given as the 2-dimensional vector $g_{Cher}(\boldsymbol{x}) = (f'_{Cher}(\boldsymbol{x}), c_{Cher}(\boldsymbol{x}))$ where

$$f'_{Cher}(\boldsymbol{x}) = \begin{cases} f(\boldsymbol{x}) & c_{Cher}(\boldsymbol{x}) \leq B \\ B - c_{Cher}(\boldsymbol{x}) & c_{Cher}(\boldsymbol{x}) > B \end{cases}$$

2 Non-monotone Submodular Problems

2.1 Maximum Cut

Given an undirected weighted graph $G = (V, E, w)$ with weights $w \colon E \to \mathbb{R}_{\geq 0}$ on the edges, the goal is to select a set $V_1 \subseteq V$ such that the sum of the weight of edges between V_1 and $V_2 = V \setminus V_1$ is maximal.

For a given search point $\boldsymbol{x} \in \{0,1\}^n$ where $n = |V|$, we have $V_1(\boldsymbol{x}) = \{v_i \mid x_i = 1\}$ and $V_2(\boldsymbol{x}) = \{v_i \mid x_i = 0\}$. Let $C(\boldsymbol{x}) = \{e \in E \mid e \cap V_1(\boldsymbol{x}) \neq \emptyset \wedge e \cap V_2(\boldsymbol{x}) \neq \emptyset\}$ be the cut of a given search point $\boldsymbol{x}$. The goal is to maximize

$$f'(\boldsymbol{x}) = \sum_{e \in C(\boldsymbol{x})} w(e).$$

Note that every search point in $\{0,1\}^n$ is feasible and there is therefore no penalty or second objective for treating potentially infeasible solutions.

2.2 Packing While Traveling

The Packing While Traveling (PWT) problem is a non-monotone submodular optimization problem which is obtained from the Traveling Thief problem (TTP) when the route is fixed. It can be formally defined as follows. Given $n+1$ cities, distances d_i, $1 \leq i \leq n$, from city i to city $i+1$, and a set of items M, $|M| = m$, distributed all over the first n cities. Each city i, $1 \leq i \leq n$, contains a set of items $M_i \subseteq M$, $|M_i| = m_i$. Each item $e_{ij} \in M_i$, $1 \leq j \leq m_i$, is characterised by its positive integer profit p_{ij} and weight w_{ij}.

In addition, a fixed route $N = (1, 2, ..., n+1)$ is given that is traveled by a vehicle with velocity $v \in [v_{min}, v_{max}]$. Let $x_{ij} \in \{0,1\}$ be a variable indicating whether or not item e_{ij} is chosen in a solution. Then a set $S \subseteq M$ of selected items can be represented by a decision vector

$$\boldsymbol{x} = (x_{11}, x_{12}, ..., x_{1m_1}, x_{21}, ..., x_{nm_n}) \in \{0,1\}^m.$$

The total benefit of selecting a subset of items selected by $\boldsymbol{x}$ is given as

$$PWT(\boldsymbol{x}) = P(\boldsymbol{x}) - R \cdot T(\boldsymbol{x}),$$

where $P(\boldsymbol{x})$ represents the total profit of selected items and $T(\boldsymbol{x})$ is the total travel time for the vehicle carrying these items. Formally, we have

$$P(\boldsymbol{x}) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} p_{ij} x_{ij}$$

and

$$T(\boldsymbol{x}) = \sum_{i=1}^{n} \frac{d_i}{v_{max} - \nu \sum\limits_{k=1}^{i} \sum\limits_{j=1}^{m_k} w_{kj} x_{kj}}$$

Here, $\nu = \frac{v_{max} - v_{min}}{B}$ is the constant defined by the input parameters, where B is the capacity of the vehicle.

The constraint is given as

$$c(\boldsymbol{x}) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} w_{ij} x_{ij} \leq B$$

Use $g(\boldsymbol{x}) = (f'(\boldsymbol{x}), c(\boldsymbol{x}))$ with

$$f'(\boldsymbol{x}) = \begin{cases} PWT(\boldsymbol{x}) & c(\boldsymbol{x}) \leq B \\ B - c(\boldsymbol{x}) - R \cdot T(v_{\min}) & c(\boldsymbol{x}) > B \end{cases}$$

where $T(v_{\min}) = \frac{1}{v_{\min}} \cdot \sum\limits_{i=1}^{n} d_i$ is the travel time at speed $v_{\min}$.