Accuracy Maximization

2022.1

Jianfeng Hou

houjf@shanghaitech.edu.cn

1 Problem Formulation

1.1 Action (Super Arm)

There are totally $N+1$ arms.

In each round, the agent chooses an action (*i.e.,* selects a super arm). Denote the set of all the super arms as $\mathcal{A}$.

1.2 Policy

A **policy** π is a distribution over actions. Formally, a policy π is defined as

$$\pi(a) = \mathbb{P}\left[A(t) = a\right], \forall a \in \mathcal{A}.$$

Not that $A(t)$ (*i.e.,* the action in round t) can be equivalently represented by a decision vector $\boldsymbol{s}(t) = [s_0(t), s_1(t), \ldots, s_N(t)]$, where

$$s_i(t) = \begin{cases} 1 & \text{if } i \in A(t), \\ 0 & \text{otherwise.} \end{cases}$$

1.3 Long-Term Time-Averaged Constraint

Under policy π, the energy consumption is *i.i.d.* over rounds with the following mean:

$$\begin{aligned} \mathbb{E}_\pi\left[\sum_{i\in A(t)} E_i(t)\right] &= \sum_{a\in\mathcal{A}} \mathbb{E}\left[\sum_{i\in A(t)} E_i(t)\middle| A(t) = a\right] \cdot \pi(a) \\ &= \sum_{a\in\mathcal{A}} \mathbb{E}\left[\sum_{i\in a} E_i(t)\right] \cdot \pi(a) \\ &= \sum_{a\in\mathcal{A}} \sum_{i\in a} \mathbb{E}\left[E_i(t)\right] \cdot \pi(a) \\ &= \sum_{a\in\mathcal{A}} \rho_a \cdot \pi(a), \end{aligned} \tag{1}$$

where $\rho_a = \sum_{i\in a} \mathbb{E}\left[E_i(t)\right]$ for all $a \in \mathcal{A}$.

There exists a long-term time-averaged constraint on the energy consumption of the smartphone:

$$\limsup_{T'\to+\infty} \frac{1}{T'} \sum_{t=1}^{T'} \mathbb{E}_\pi\left[\sum_{i\in\mathcal{N}} E_i(t) s_i(t)\right] \le b \tag{2}$$

where b is a positive constant representing the energy budget in each round.

According to (1), constraint (2) can be written as:

$$\sum_{a \in \mathcal{A}} \rho_a \cdot \pi(a) \leq b. \tag{3}$$

1.4 The Original Problem

$$K(t) = 1 - \prod_{i \in \mathcal{N}} (1 - C_i(t) s_i(t)) = 1 - \prod_{i \in A(t)} (1 - C_i(t)).$$

$$\begin{aligned} \underset{\pi}{\text{maximize}} \quad & \mathbb{E}_\pi \left[\sum_{t=1}^{T} K(t) \right] \\ \text{subject to} \quad & \sum_{a \in \mathcal{A}} \rho_a \cdot \pi(a) \leq b. \end{aligned} \tag{4}$$

For all $a \in \mathcal{A}$, the mean reward ϕ_a of pulling super arm a is given by

$$\begin{aligned} \phi_a &\triangleq \mathbb{E}\left[R_a(t)\right] \\ &= \mathbb{E}\Big[K(t) \Big| A(t) = a\Big] \\ &= \mathbb{P}\Big[K(t) = 1 \Big| A(t) = a\Big] \\ &= \mathbb{P}\left[\prod_{i \in A(t)} (1 - C_i(t)) = 0 \middle| A(t) = a\right] \\ &= \mathbb{P}\left[\prod_{i \in a} (1 - C_i(t)) = 0\right] \\ &= \mathbb{P}\left[\bigcup_{i \in a} C_i(t) = 1\right] \\ &= 1 - \mathbb{P}\left[\bigcap_{i \in a} C_i(t) = 0\right] \\ &= 1 - \prod_{i \in a} (1 - c_i). \end{aligned}$$

The mean reward obtained by the agent under policy π is given by

$$\mathbb{E}_\pi\left[R(t)\right] = \sum_{a \in \mathcal{A}} \phi_a \cdot \pi(a).$$

According to the INFOCOM 2019 Fairness paper (Jia Liu), assuming the mean reward vector $\boldsymbol{\phi} = \{\phi_a : a \in \mathcal{A}\}$ is known in advance, the reward maximization problem with a long-term time-averaged constraint can be formulated as the following linear program:

$$
\begin{aligned}
\underset{\pi}{\text{maximize}} \quad & \sum_{a \in \mathcal{A}} \phi_a \cdot \pi(a) \\
\text{subject to} \quad & \sum_{a \in \mathcal{A}} \rho_a \cdot \pi(a) \leq b, \\
& \pi(a) \in [0,1], \forall a \in \mathcal{A}, \\
& \sum_{a \in \mathcal{A}} \pi(a) = 1.
\end{aligned} \tag{5}
$$

1.5 The Transformed Problem

$$
\begin{aligned}
\underset{\pi}{\text{maximize}} \quad & \mathbb{E}_\pi \left[\sum_{t=1}^{T} \sum_{i \in A(t)} C_i(t) \right] \\
\text{subject to} \quad & \sum_{a \in \mathcal{A}} \rho_a \cdot \pi(a) \leq b.
\end{aligned}
$$

For all $a \in \mathcal{A}$, the mean reward ω_a of pulling super arm a is given by

$$
\begin{aligned}
\omega_a &\triangleq \mathbb{E}\left[R_a^{'}(t) \right] && (6) \\
&= \mathbb{E}\left[\sum_{i \in A(t)} C_i(t) \middle| A(t) = a \right] && (7) \\
&= \sum_{i \in a} c_i. && (8)
\end{aligned}
$$

The mean reward obtained by the agent under policy π is given by

$$
\begin{aligned}
\mathbb{E}_\pi \left[R'(t) \right] &= \sum_{a \in \mathcal{A}} \mathbb{E}\left[R_a^{'}(t) \right] \cdot \pi(a) \\
&= \sum_{a \in \mathcal{A}} \sum_{i \in a} c_i \cdot \pi(a) \\
&= \sum_{a \in \mathcal{A}} \omega_a \cdot \pi(a).
\end{aligned}
$$

$$
\begin{aligned}
\underset{\pi}{\text{maximize}} \quad & \sum_{a \in \mathcal{A}} \omega_a \cdot \pi(a) \\
\text{subject to} \quad & \sum_{a \in \mathcal{A}} \rho_a \cdot \pi(a) \leq b, \\
& \pi(a) \in [0,1], \forall a \in \mathcal{A}, \\
& \sum_{a \in \mathcal{A}} \pi(a) = 1.
\end{aligned} \tag{9}
$$

1.6 Proof of Equivalence

Now we prove that the optimal solution to Problem 9 is exactly the optimal solution to Problem 5.

$$\mathbb{E}_{\pi_1}\left[R^{'}(t)\right] \geq \mathbb{E}_{\pi_2}\left[R^{'}(t)\right] \Rightarrow \mathbb{E}_{\pi_1}\left[R(t)\right] \geq \mathbb{E}_{\pi_2}\left[R(t)\right], \forall \pi_1, \pi_2 \in \mathcal{F}.$$

Proof

As the constraint of Problem (5) is exactly the same as the constraint of Problem (9), we can denote the set of all the feasible policies as $\mathcal{F}$. Assume π^* is the optimal solution to Problem 9, then we have for any $\pi \in \mathcal{F}$, $\mathbb{E}_{\pi^*}\left[R^{'}(t)\right] \geq \mathbb{E}_{\pi}\left[R^{'}(t)\right]$.