Beyond Adaptive Submodularity: Approximation Guarantees of Greedy Policy with Adaptive Submodularity Ratio

Kaito Fujii (University of Tokyo) & Shinsaku Sakaue (NTT)

The 36th International Conference on Machine Learning

Our Contributions

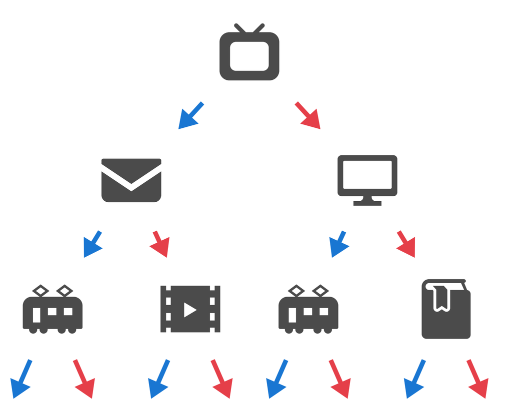
Adaptive Submodularity Ratio is applied to

- **Theorem 1** Bounds on approx. ratio of Adaptive Greedy
- **Theorem 2** Bounds on adaptivity gaps
- **Application 1** Influence maximization on bipartite graphs
- **Application 2** Adaptive feature selection

Adaptive Stochastic Optimization

A decision maker repeats selecting an element and observing its state alternately

Select element $v \in V$ ⇄ Observe its state $\phi(v) \in \mathcal{Y}$



An adaptive policy π determines the element to be selected next

Goal To find a near-optimal policy

$$\text{Maximize}_{\pi \in \Pi_k} \; \mathbb{E}_\Phi[f(E(\pi, \Phi), \Phi)]$$

Π_k: the set of all policies of height at most k

$E(\pi, \Phi)$: the subset selected by π under realization ϕ

V finite set $\quad \mathcal{Y}$ set of possible states

$\phi \colon V \to \mathcal{Y}$ states $\quad p \in \Delta_{\mathcal{Y}^V}$ distribution of ϕ

$f \colon 2^V \times \mathcal{Y}^V \to \mathbb{R}_{\geq 0}$ objective function

Adaptive Greedy works well in many applications even if the objective function lacks adaptive submodularity [Golovin–Krause'11]

Q Under what condition does Adaptive Greedy work well?

Definition: Adaptive Submodularity Ratio

Adaptive submodularity ratio $\gamma_{\ell,k} \in [0, 1]$ measures the distance to adaptive submodular functions

adaptive submodular functions	$(1 - \exp(-\gamma_{k,k}))$-approximation [this study]	arbitrary adaptive monotone functions
$\gamma_{\ell,k} = 1$	⟷	$\gamma_{\ell,k} = 0$
submodular functions	$(1 - \exp(-\gamma_{k,k}))$-approximation [Das–Kempe'11]	arbitrary monotone functions

$$\gamma_{\ell,k} \triangleq \min_{|\psi| \leq \ell,\; \pi \in \Pi_k} \frac{\sum_{v \in V} \Pr(v \in E(\pi, \Phi) \mid \Phi \sim \psi)\, \Delta(v \mid \psi)}{\Delta(\pi \mid \psi)}$$

- $\Pr(v \in E(\pi, \Phi) \mid \Phi \sim \psi)$: the probability that v is selected by π
- $\Delta(v \mid \psi)$: the expected marginal gain of v
- $\Delta(\pi \mid \psi)$: the expected marginal gain of π

$\psi = \{(v_1, \phi(v_1)), \cdots, (v_\ell, \phi(v_\ell))\}$ observations obtained so far

$\Delta(v \mid \psi) = \mathbb{E}_\Phi[f(\mathrm{dom}(\psi) \cup \{v\}, \Phi) - f(\mathrm{dom}(\psi), \Phi) \mid \Phi \sim \psi]$

$\Delta(\pi \mid \psi) = \mathbb{E}_\Phi[f(\mathrm{dom}(\psi) \cup E(\pi, \Phi), \Phi) - f(\mathrm{dom}(\psi), \Phi) \mid \Phi \sim \psi]$

Theorem (1): Approximation Ratio

Adaptive Greedy achieves a good approximation if adaptive submodularity ratio is large

Adaptive Greedy [Golovin–Krause'11]

$\psi \leftarrow \emptyset$ // Initialize
For $|\psi| < k$:
  $v^* \in \operatorname{argmax}\{\Delta(v \mid \psi) \mid v \in V\}$ // greedy selection
  Observe $\phi(v^*)$ // observation of the state
  $\psi \leftarrow \psi \cup \{(v^*, \phi(v^*))\}$

Theorem **Adaptive Greedy** is $(1 - \exp(-\gamma_{k,k}))$-approx.

Theorem (2): Bounds on Adaptivity Gaps

An optimal non-adaptive policy is an approximation to an optimal adaptive policy

$$\mathrm{GAP}_k(f, p) \triangleq \frac{\max_{S^* \colon |S^*| \leq k} \mathbb{E}_\Phi[f(S^*, \Phi)]}{\max_{\pi^* \in \Pi_k} \mathbb{E}_\Phi[f(E(\pi^*, \Phi), \Phi)]} \in [0, 1]$$

- numerator: the value achieved by an optimal **non-adaptive** policy
- denominator: the value achieved by an optimal **adaptive** policy

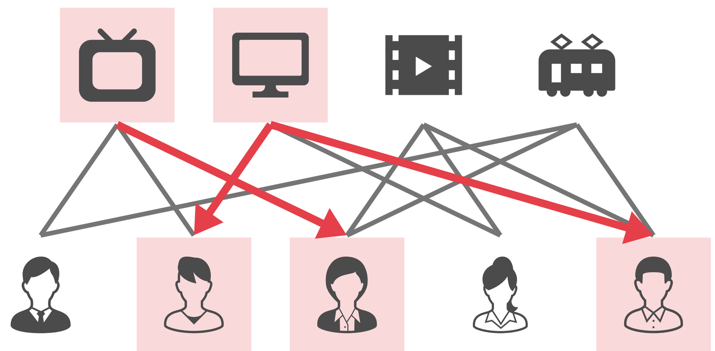
Theorem $\mathrm{GAP}_k(f, p) \geq \beta_{0,k} \gamma_{0,k}$

$$\beta_{0,k} \triangleq \min_{S \subseteq V \colon |S| \leq k} \frac{\mathbb{E}[f(S, \Phi)]}{\sum_{v \in S} \mathbb{E}[f(\{v\}, \Phi)]}$$

supermodularity ratio of $\mathbb{E}_\Phi[f(\cdot, \Phi)]$

Application (1): Influence Maximization

Select a subset of ads to influence many people



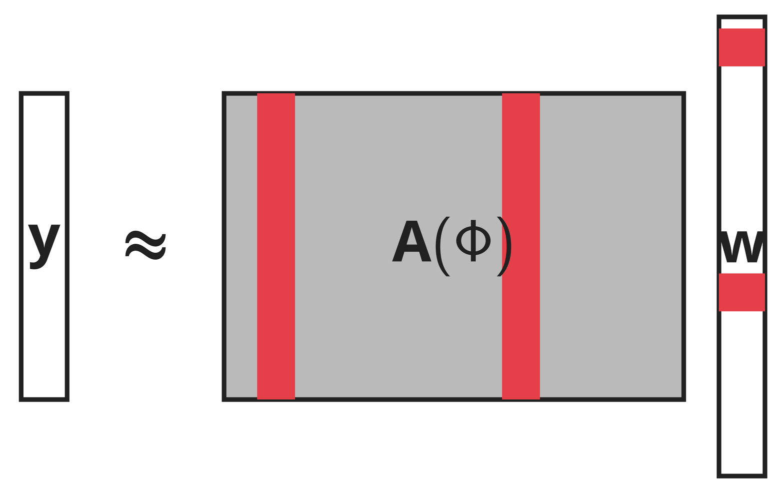
Non-adaptive setting Select a subset in advance

Adaptive setting Select ads one by one

Theorem $\gamma_{\ell,k} \geq \dfrac{k+1}{2k}$ under the triggering model

Application (2): Adaptive Feature Selection

Select a subset of features to be observed precisely



Non-adaptive setting Select a subset in advance

Adaptive setting Observe features one by one

Theorem $\gamma_{\ell,k} \geq \min_{\phi} \min_{S \subseteq V \colon |S| \leq \ell + k} \lambda_{\min}(\mathbf{A}(\phi)_S^\top \mathbf{A}(\phi)_S)$