Large-Scale Optimistic Adaptive Submodularity

Victor Gabillon, Branislav Kveton, Zheng Wen, Brian Eriksson, S. Muthukrishnan

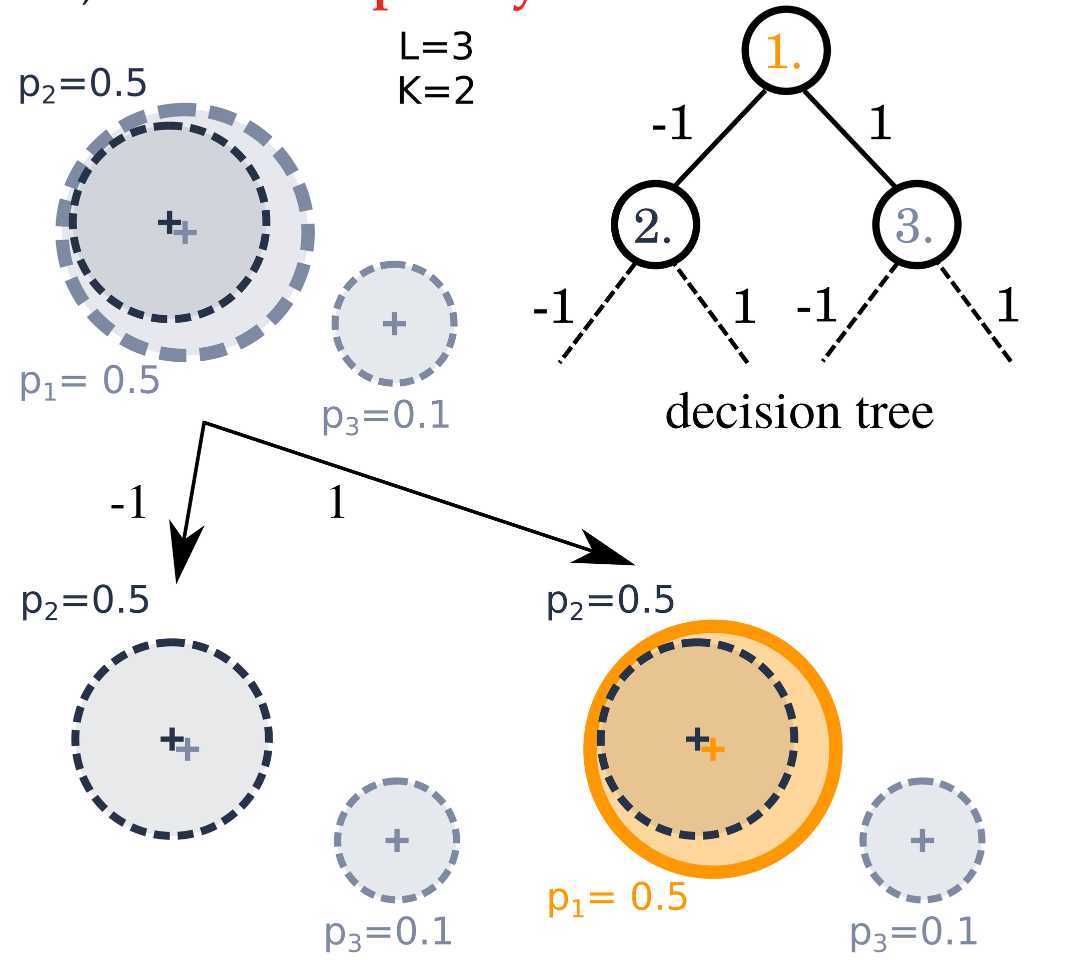
technicolor

Abstract

- Maximization of submodular functions (SM) has wide applications in machine learning.
- We propose a scalable learning algorithm for maximizing an adaptive submodular function
- Key structural assumption: the state of an item is distributed according to a generalized linear model conditioned on the item's feature vector
- Our work brings give a tractable solution to a kind of POMDP problem by bringing together the concepts of adaptive submodular maximization and linear bandits.

Adaptive Submodular Maximization

- Three sensors ($L = 3$) with known placements.
- The area covered by sensors is known.
- Upon choosing, each sensor i covers its area with probability p_i and the state of the sensor (active (1) or inactive (-1)) is **observed**.
- **Goal:** Maximize the **expected** area covered by two sensors, **chosen adaptively**.



Maximize a **real** function of the form:

$f(A, \phi) \to \mathbb{R}$

Set of chosen items — State of all items

The **state** $\phi \in \{-1, 1\}^L$ is drawn i.i.d. from a probability distribution $P(\Phi)$. The i-th entry of the state ϕ, $\phi[i]$, is the state of item i.

An **observation** is a vector $y \in \{-1, 0, 1\}^L$.

$\phi = (-1, -1, 1, -1, 1)$ ← State

$y = (0, -1, 0, -1, 1)$ ← Observation

$y = (0, 0, 0, -1, 0)$ ← Observation

We say that $y \subseteq y$, $\phi \sim y$, and $\phi \sim y$.

The observed items in y are $\mathrm{dom}(y) = \{2, 4, 5\}$.

- $\pi_k(\phi)$ are the first k items chosen by π in state ϕ.
- The optimal K-step policy is defined as:

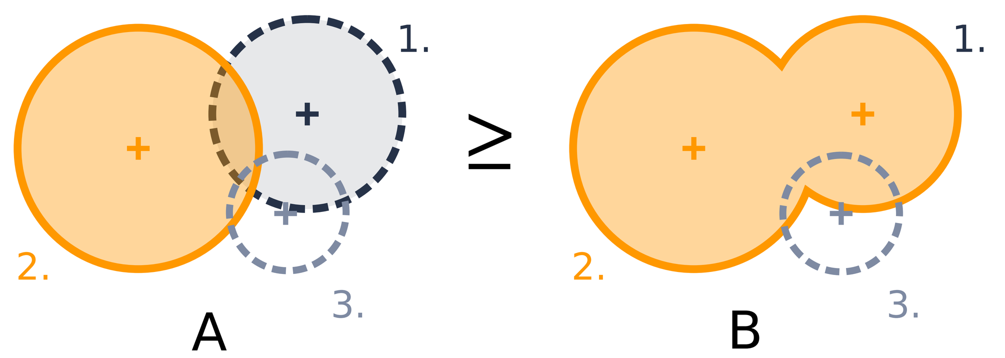
$$\pi^* = \arg\max_\pi \mathbb{E}_\phi[f(\pi_K(\phi), \phi)]$$

- Computing π^* is NP-hard **BUT** …

Function f is **adaptive submodular** if:

$$\mathbb{E}_\phi[f(A \cup \{i\}, \phi) - f(A, \phi) \mid \phi \sim y_A] \geq$$
$$\mathbb{E}_\phi[f(B \cup \{i\}, \phi) - f(B, \phi) \mid \phi \sim y_B]$$

for all i, $y_B \supseteq y_A$, $A = \mathrm{dom}(y_A)$, $B = \mathrm{dom}(y_B)$.



Function f is **adaptive monotonic** if

$$\mathbb{E}_\phi[f(A \cup \{i\}, \phi) - f(A, \phi) \mid \phi \sim y_A] \geq 0$$

for all i and $A = \mathrm{dom}(y_A)$.

- Let π^g be the **greedy policy** for maximizing f, the policy that chooses the item with the highest expected gain:

$$\pi^g(y) = \arg\max_{i \in I \setminus \mathrm{dom}(y)} g_i(y)$$

where $g_i(y) =$

$$\mathbb{E}_\phi[f(\mathrm{dom}(y) \cup \{i\}, \phi) - f(\mathrm{dom}(y), \phi) \mid \phi \sim y]$$

is the expected gain of item i after observing y.

- If f is adaptive submodular and monotonic, then **π^g is a $(1 - 1/e)$-approximation to π^*.**

The greedy policy cannot be computed when $P(\phi)$ is unknown!

Linear Optimistic Adaptive Submodular Maximiz (LinOASM)

Our approach: Mimic the greedy policy π^g **while learning** $P(\Phi)$.

Assumption: Each item (movie) is associated with a d-dimensional feature vector x_i.
The state of an item is distributed independently of the others:

$$P(\Phi = \phi) = \prod_{i=1}^{L} p_i(x_i)^{1\{\phi[i]=1\}} (1 - p_i(x_i))^{1\{\phi[i]=-1\}},$$

$p_i(x_i)$ is the probability that item i is in state 1.

We model using GLM, here, logistic regression:

$$p_i(x) = P(\phi[i] = 1 \mid x) = \theta(x^\top \theta^*),$$

where $\theta(u) = 1/(1 + e^{-u})$ is a logistic function, $\theta^* \in \Theta$ is a column vector of parameters to be learned.

- Learning d parameters instead of L or 2^K.

The expected gain can be written as $g_i(y) = p_i \bar{g}_i(y)$, where $\bar{g}_i(y)$ is the expected gain of choosing item i in state 1 (area covered when sensor i is active). $\bar{g}_i(y)$ can be computed without knowing $P(\Phi)$.

LinOASM

Play n episodes of a K-step game.

for $t = 1, 2, \ldots, n$ **do**

Maximize f in K steps using the policy:

$$\pi^t(y) = \arg\max_i \left[\theta(x_i^\top \tilde{\theta}_t) + \rho_{k,t} \|x_i\|_{M_t^{-1}}\right] \bar{g}_i(y)$$

Update all statistics of the model

Regression to estimate $\tilde{\theta}_t$

Theoretical Results

Goal: Minimize the **expected cumulative regret**:

$$R(n) = \mathbb{E}_{\phi_1,\ldots,\phi_n}\left[\sum_{t=1}^{n} f(\pi_K^g(\phi_t), \phi_t) - f(\pi_K^t(\phi_t), \phi_t)\right]$$

Regret bound:

$$R(n) \leq \sum_{k=1}^{K} G_k(\ell_k + O(1)),$$

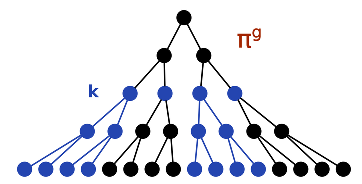
$O(\log^3 n)$

where ℓ_i is the number of pulls after which arm i is unlikely to be pulled suboptimally, G_k is an upper bound on the expected gain of the greedy policy π^g from step k forward.

$$\ell_k \approx d^2 \times \max_{i \,\&\, y \in Y_{k,i}} \frac{\bar{g}_i^2(y)}{\Delta_i^2(y)} \log^3 \frac{c^2 n K}{\lambda}$$

$$G_k = (K - k + 1) \max_{y \in Y_k} \max_i g_i(y)$$

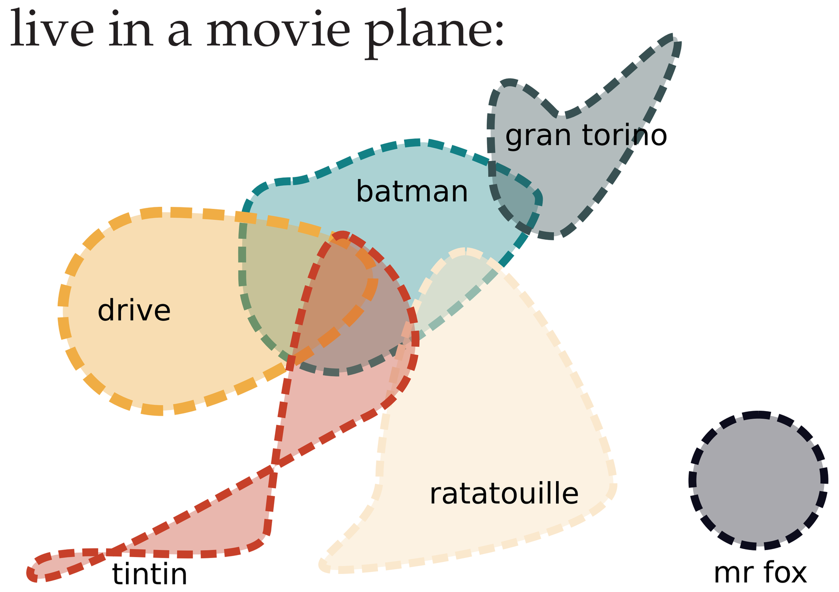
- Polynomial in both d and K **and not L !!**.
- Captures the shape of the optimized function f.



Experiments on a Preference Elicitation Problem

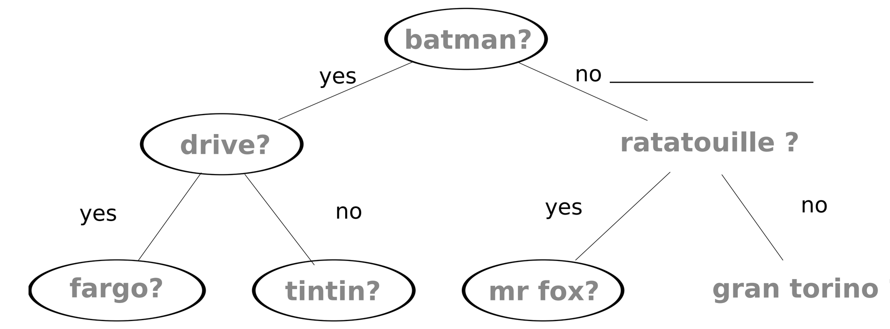
Goal: Identify the largest number of movies of interest in K questions.

Movies live in a movie plane:



- Movie preference vectors ϕ_j are estimated from users historical ratings using low-rank matrix factorization from 500 most rated movies.
- u_j describes the users, x_i the movies $d = 10$.
- User j rates movie i as $r_{j,i} = \langle u_j, x_i \rangle$, We set $\phi_j[i]$ to 1 if movie i is rated highly by user j, $r_{j,i} > 4$.

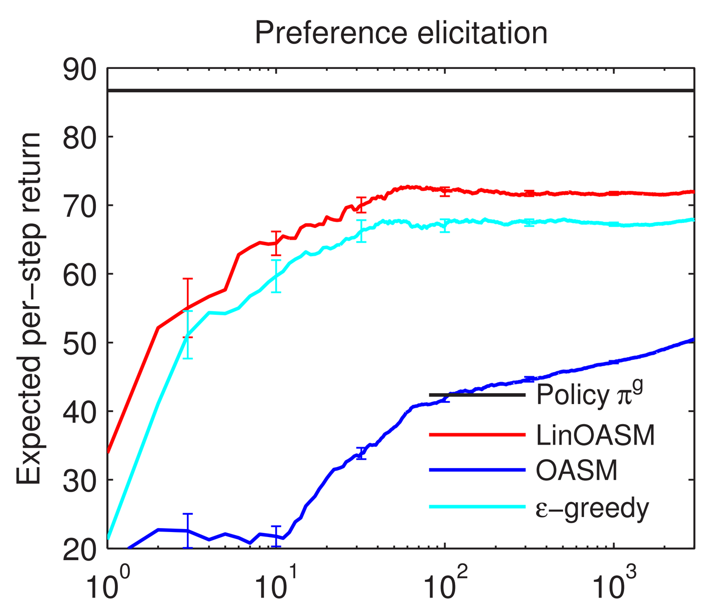
We want to cover this space as much as possible by asking adaptively K questions:



The return, $f(A, \phi) =$

$$\sum_{\ell=1}^{500} 0.1 \cdot 1\{\exists i \in A : \|x_\ell - x_i\|_2 \leq 0.2\} + 0.9 \cdot 1\{\exists i \in A : \|x_\ell - x_i\|_2 \leq 0.2, \phi[i] = 1\}$$

The function $f(A, \phi)$ is adaptive submodular and is maximized when A is a diverse set of liked movies.



a. The expected per-step return up to episode $n = 3k$. **b.** Visualization of the OASM & LinOASM in episode $n = 100$. The black dots are liked movies, projected on the first two components of the latent space. The colored dots are liked movies that are similar to at least one queried movie $i \in A$ that is liked.