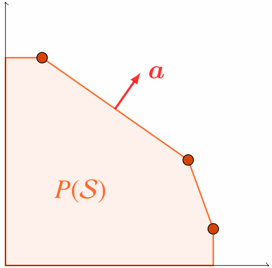
Regret Ratio Minimization in Multi-objective Submodular Function Maximization

(AAAI 2017)

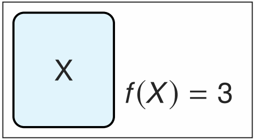


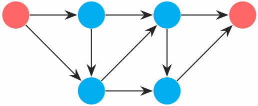
相馬 輔 (東京大学)

with 吉田 悠一 (NII & PFI)

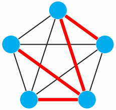
劣モジュラ最適化

劣モジュラ関数: “効率的” に解ける離散構造に現れる関数





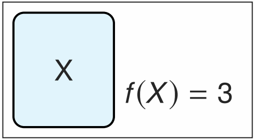
ネットワークフロー

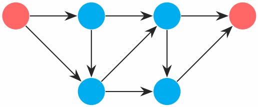


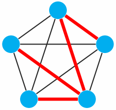
全域木

劣モジュラ最適化

劣モジュラ関数: “効率的” に解ける離散構造に現れる関数




ネットワークフロー


全域木

$f : 2^S \to \mathbb{R}$ が **劣モジュラ関数**:

$$f(X \cup s) - f(X) \geq f(Y \cup s) - f(Y) \quad (X \subseteq Y, s \notin Y)$$

“限界効用逓減 (diminishing return)”

劣モジュラ関数最大化

$$\max \quad f(X) \qquad \text{sub. to} \quad X \in \mathcal{C}$$

解の品質

制約 (コスト)

劣モジュラ関数最大化

$$\max \quad f(X) \qquad \text{sub. to} \quad X \in \mathcal{C}$$

解の品質

制約 (コスト)

既存研究

- 一般には NP 困難
- 多くの制約に対して効率的な**定数近似**アルゴリズム

劣モジュラ関数最大化

$$\max \quad f(X) \qquad \text{sub. to} \quad X \in \mathcal{C}$$

解の品質

制約 (コスト)

既存研究

- 一般には NP 困難
- 多くの制約に対して効率的な **定数近似**アルゴリズム

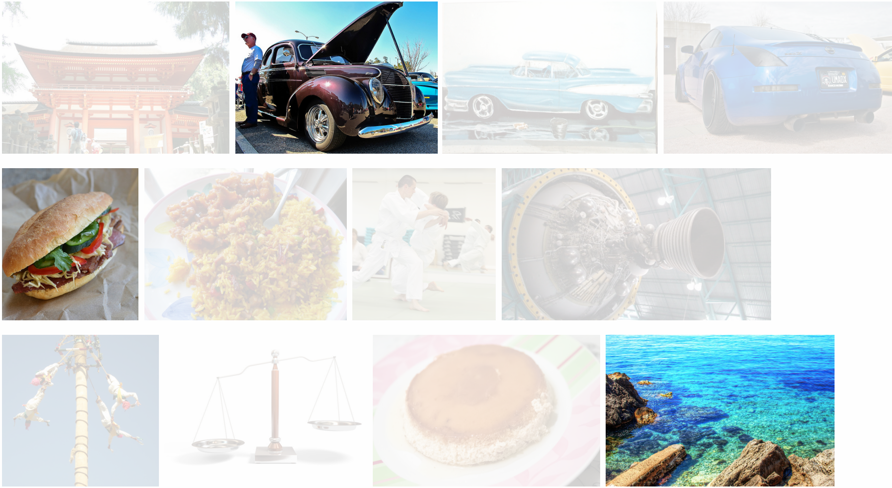
応用

- 影響力最大化
- データ要約, etc

Multiple Criteria

出典: Image Net Dataset (`www.image-net.org`)

Multiple Criteria

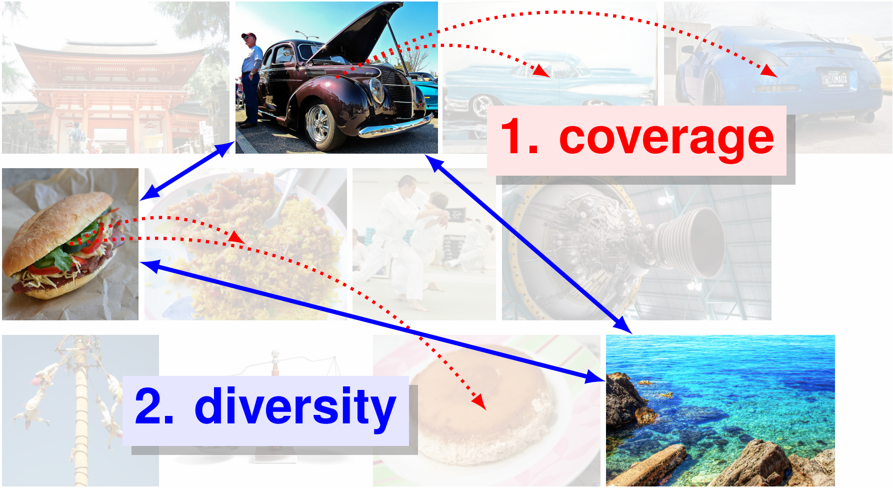


出典: Image Net Dataset (`www.image-net.org`)

Multiple Criteria

出典: Image Net Dataset (www.image-net.org)

Multiple Criteria



出典: Image Net Dataset (www.image-net.org)

多目的最適化の出番?

✕ パレート最適解は指数個ありうる!

多目的最適化の出番?

✕ パレート最適解は指数個ありうる!

パレート解の "良い" 代表集合を選ぶ既存手法

- *k*-representative skyline queries [Lin et al. 07, Tao et al. 09]
- top-*k* dominating queries [Yiu and Mamoulis 09]
- regret minimizing database [Nanongkai et al. 10]

多目的最適化の出番?

✕ パレート最適解は指数個ありうる!

パレート解の "良い" 代表集合を選ぶ既存手法

- *k*-representative skyline queries [Lin et al. 07, Tao et al. 09]
- top-*k* dominating queries [Yiu and Mamoulis 09]
- regret minimizing database [Nanongkai et al. 10]

既存手法の問題点

すべての実行可能解が**陽に与えられている**必要...

概要

Regret minimizing database を劣モジュラ最大化に拡張

概要

Regret minimizing database を劣モジュラ最大化に拡張

アルゴリズム

単目的最大化に対して α 近似アルゴリズムがあるとき

- リグレット比 $1 - \alpha/d$ (d:任意)
- リグレット比 $1 - \alpha + O(1/k)$ (k:任意, $d = 2$)

d = 目的関数の数, k = 実行可能解の個数

概要

Regret minimizing database を劣モジュラ最大化に拡張

アルゴリズム

単目的最大化に対して α 近似アルゴリズムがあるとき

- リグレット比 $1 - \alpha/d$ (d:任意)
- リグレット比 $1 - \alpha + O(1/k)$ (k:任意, $d = 2$)

d = 目的関数の数, k = 実行可能解の個数

情報理論的下界

- $\alpha = 1$ (厳密アルゴリズム), $d = 2$ であっても,
  リグレット比 $o(1/k^2)$ は達成不可能.

リグレット比

単目的最適化

$\mathcal{S} \subseteq \mathcal{C}, f$ に対する **リグレット比**

$$\mathrm{rr}(\mathcal{S}) = \frac{\max_{X \in \mathcal{C}} f(X) - \max_{X \in \mathcal{S}} f(X)}{\max_{X \in \mathcal{C}} f(X)}.$$

リグレット比

単目的最適化

$\mathcal{S} \subseteq \mathcal{C}, f$ に対する **リグレット比**

$$\mathrm{rr}(\mathcal{S}) = \frac{\max_{X \in \mathcal{C}} f(X) - \max_{X \in \mathcal{S}} f(X)}{\max_{X \in \mathcal{C}} f(X)}.$$

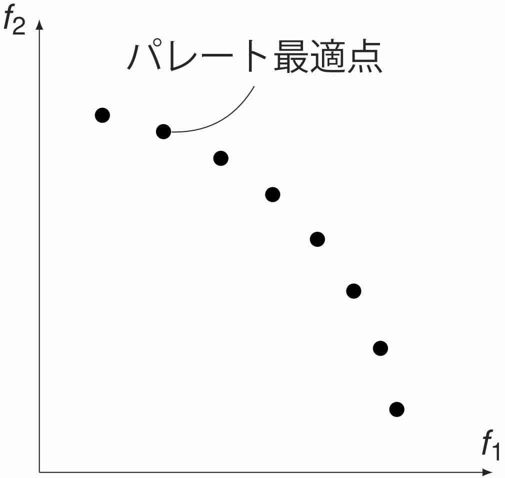
多目的最適化

$\mathcal{S} \subseteq \mathcal{C}$ and $f_1, \ldots, f_d$ に対する **リグレット比**

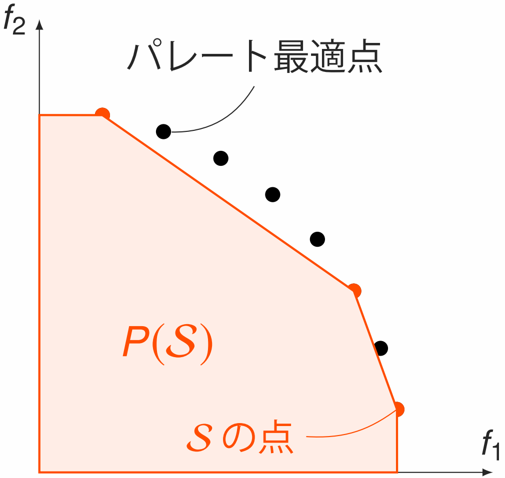
$$\mathrm{rr}_{f_1,\ldots,f_d,\mathcal{C}}(\mathcal{S}) = \max_{\mathbf{a} \in \mathbb{R}_+^d} \mathrm{rr}_{f_\mathbf{a},\mathcal{C}}(\mathcal{S}),$$

ただし $f_\mathbf{a} := a_1 f_1 + \cdots + a_d f_d$. (**非負結合**)

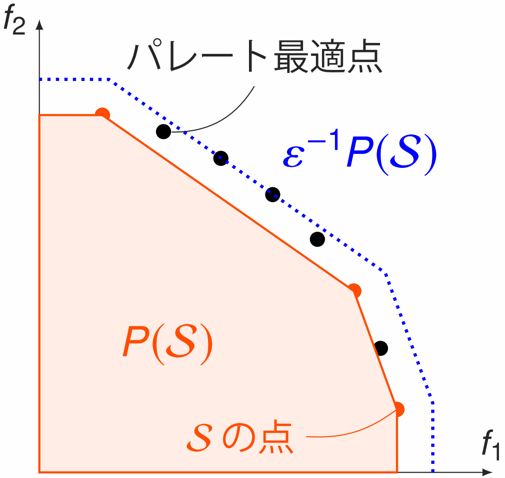
リグレット比の幾何的解釈



リグレット比の幾何的解釈



リグレット比の幾何的解釈



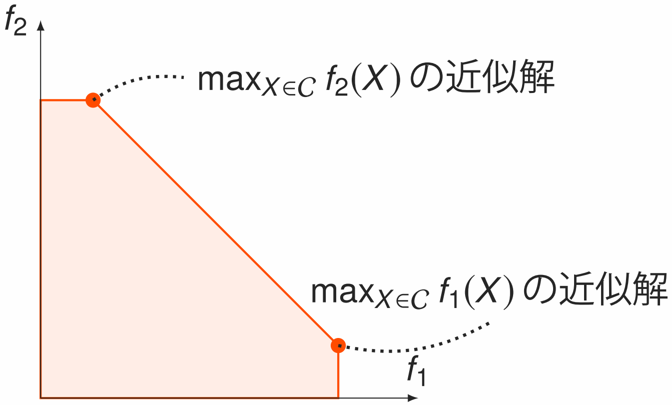
$$\mathrm{rr}(\mathcal{S}) \leq 1 - \varepsilon \iff \mathbf{f}(X) \in \varepsilon^{-1}P(\mathcal{S}) \quad (X \in \mathcal{C}).$$

リグレット比最小化

Given: $f_1, \ldots, f_d$: 劣モジュラ, $\mathcal{C} \subseteq 2^E$, $k > 0$

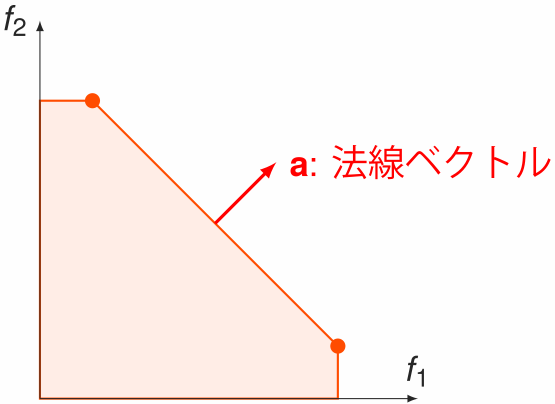
$$
\begin{array}{rl}
\text{minimize} & \mathrm{rr}(\mathcal{S}) \\
\text{subject to} & \mathcal{S} \subseteq \mathcal{C}, |\mathcal{S}| \le k.
\end{array}
$$

アルゴリズム 1: Coordinate

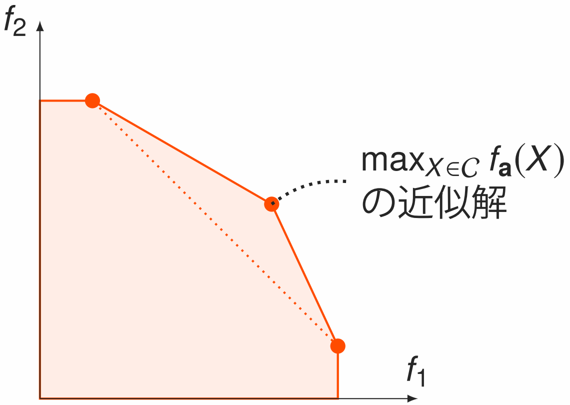


$\mathcal{S}_{\mathrm{coord}}$: $\max_{X \in \mathcal{C}} f_i(X)$ の α 近似解の集合 ($i = 1, \ldots, d$)
$\implies \mathrm{rr}(\mathcal{S}_{\mathrm{coord}}) \leq 1 - \alpha/d$.

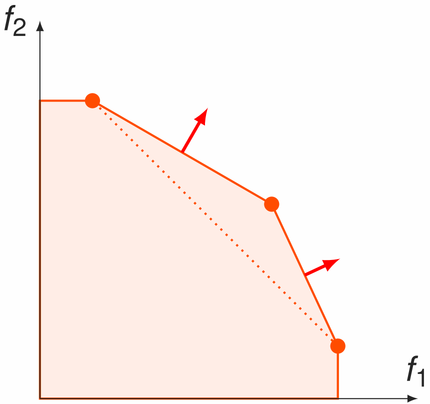
アルゴリズム2: Polytope



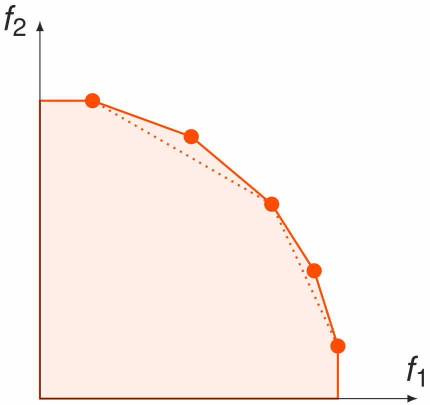
アルゴリズム 2: Polytope



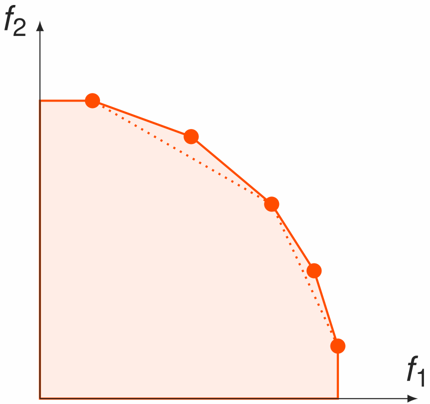
アルゴリズム 2: Polytope



アルゴリズム 2: Polytope



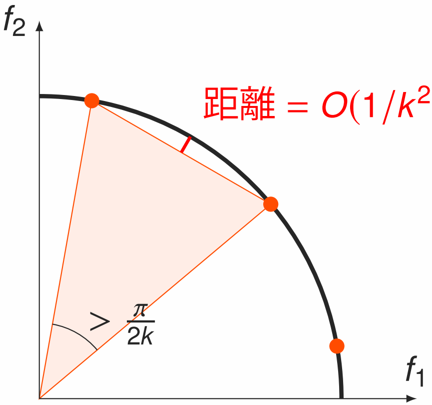
アルゴリズム2: Polytope



$\mathcal{S}$: `Polytope` の出力, $d = 2$
$\implies \mathrm{rr}(\mathcal{S}) \leq 1 - \alpha + O(1/k)$.

リグレット比の下界

$$f_1(X) = \cos\frac{\pi|X|}{2n}, \quad f_2(X) = \sin\frac{\pi|X|}{2n}$$



数値実験

アルゴリズム

- `Coordinate`
- `Polytope`
- `Random`: k 個のランダムベクトル $\mathbf{a}_1, \ldots, \mathbf{a}_k$ を生成，$\{X_1, \ldots, X_k\}$ (X_i : $\max_{X \in \mathcal{C}} f_{\mathbf{a}_i}(X)$ の近似解) を出力

計算機環境

- Intel Xeon E5-2690 (2.90 GHz) CPU, 256 GB RAM
- C#で実装

データ要約

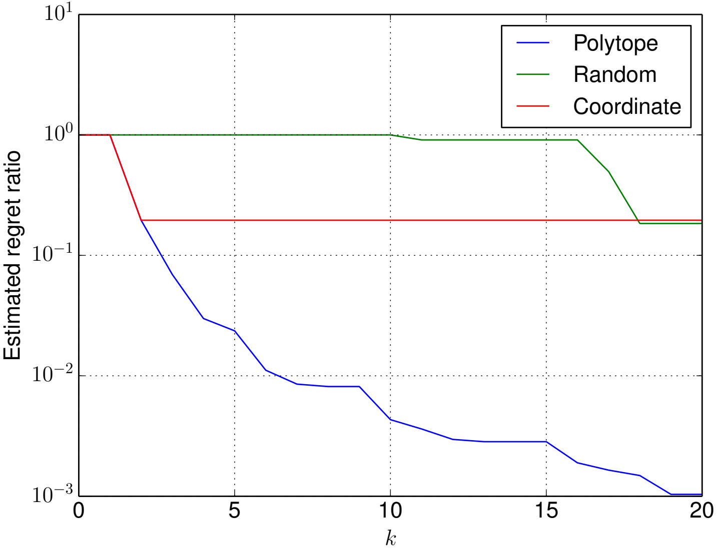
データセット: `MovieLens`
E: 映画の集合, $s_{i,j}$: 映画 i と j の類似度

$$f_1(X) = \sum_{i \in E} \sum_{j \in X} s_{i,j}, \quad \textbf{coverage}$$

$$f_2(X) = \lambda \sum_{i \in E} \sum_{j \in E} s_{i,j} - \lambda \sum_{i \in X} \sum_{j \in X} s_{i,j} \quad \textbf{diversity}$$

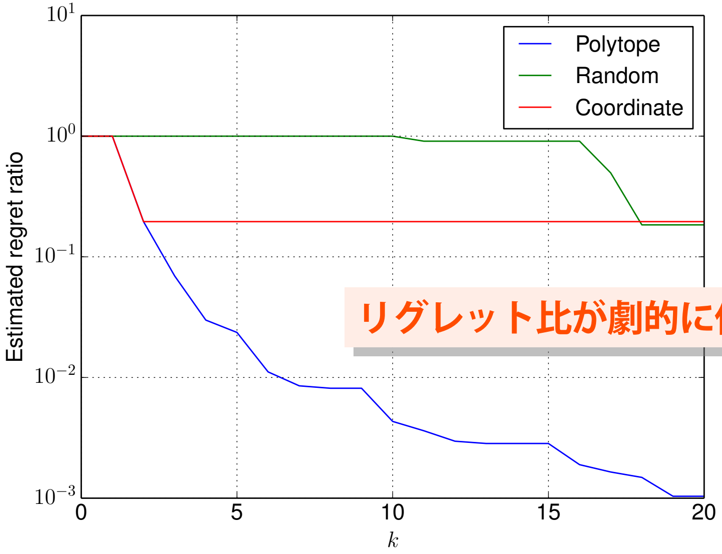
$\mathcal{C} = 2^E$ (無制約), $1 \le k \le 20$, $\lambda > 0$,
単目的アルゴリズム: double greedy (1/2 近似)
[Buchbinder et al. 12]

実験結果



Polytope
Random
Coordinate
Estimated regret ratio
10^{1}
10^{0}
10^{-1}
10^{-2}
10^{-3}
0
5
10
15
20
k

実験結果



リグレット比が劇的に低下

概要

Regret minimizing database を劣モジュラ最大化に拡張

アルゴリズム

単目的最大化に対して α 近似アルゴリズムがあるとき

- リグレット比 $1 - \alpha/d$ (d:任意)
- リグレット比 $1 - \alpha + O(1/k)$ (k:任意, $d = 2$)

d = 目的関数の数, k = 実行可能解の個数

情報理論的下界

- $\alpha = 1$ (厳密アルゴリズム), $d = 2$ であっても, リグレット比 $o(1/k^2)$ は達成不可能.