比較バンディット問題における最適アルゴリズム
(Regret Lower Bound and Optimal Algorithm in Dueling Bandit Problem)

小宮山純平[1]

本発表は以下の方々との共同研究の内容です。

本多淳也[1], 鹿島久嗣[2], 中川裕志[1]（敬称略）

1. 東京大学 2. 京都大学

概要

- 問題設定：比較バンディット問題
  - モチベーション：検索エンジンのランキング手法比較
- 理論解析とアルゴリズム提案
  - アルゴリズムの性能 (Regret) 限界
  - アルゴリズム：RMED
    - 性能限界を漸近的に達成する現在唯一のアルゴリズム
- 数値実験による性能比較

本研究の主貢献

概要

- 問題設定：比較バンディット問題
  - モチベーション：検索エンジンのランキング手法比較
- 理論解析とアルゴリズム提案
  - アルゴリズムの性能 (Regret) 限界
  - アルゴリズム：RMED
    - 性能限界を漸近的に達成する現在唯一のアルゴリズム
- 数値実験による性能比較

モチベーション：検索エンジンのランキング手法の比較

- 複数のランキング手法のどれが最も良いかを比較したい
- 従来法：専門のチーム（品質検証チーム）による評価
  - 高コスト、ユーザ評価との差異
- 代案：インターリービング比較[Joachims+ 02]
  - 次頁

インターリービング比較

Slide from
[Radlinski et al. 2008,
Yue et al. 2011]

Ranking A

1.Napa Valley – The authority for lodging...
www.napavalley.com
2.Napa Valley Wineries - Plan your wine...
www.napavalley.com/wineries
3.Napa Valley College
www.napavalley.edu/homex.asp
4. Been There | Tips | Napa Valley
www.ivebeenther
81
5. Napa Valley Wine
www.napavintne
6. Napa Country, Ca
Wikipedia
en.wikipedia.org/

Ranking B

1. Napa Country, California – Wikipedia
en.wikipedia.org/wiki/Napa_Valley
2. Napa Valley – The authority for lodging...
www.napavalley.com
he Story of an American
oogl uk/books?isbn=...
ls – Bed and
palinks.com
lley.org
apavalley.org
Marathon
marathon.org

Presented Ranking

1.Napa Valley – The authority for lodging...
www.napavalley.com
2. Napa Country, California – Wikipedia
en.wikipedia.org/wiki/Napa_Valley
3. Napa: The Story of an American Eden...
books.google.co.uk/books?isbn=...
4.Napa Valley Wineries – Plan your wine...
www.napavalley.com/wineries
5. Napa Valley Hotels – Bed and Breakfast...
www.napalinks.com
6.Napa Balley College

Click
Click
B wins!

例：検索エンジンのランキング

- インターリービング比較：
  - ユーザに２つのランキングを混ぜて提示
    →フィードバック（クリック）による評価
- インターリービング比較の一般化：比較バンディット問題 ← **本研究の対象**
  - 一対比較（pairwise feedback）から最も良い選択肢を動的に探す問題

比較バンディット問題
Dueling bandit problem [Yue+ COLT2009]

- 入力：アーム数K
- 各ラウンド $t = 1, \ldots, T$に、
  1. ２つのアーム$l(t), m(t)$ in [K]を選択
     - 無情報だが、$l(t) = m(t)$の比較も可能（後述）
  2. どちらのアームが良いかのフィードバックを得る（$l(t)$ or $m(t)$）

目的：Regretの最小化（後述）

比較バンディット問題
Dueling bandit problem [Yue+ COLT2009]

- 入力：アーム数K
- 各ラウンド $t = 1, \ldots, T$に、

アームを選択するアルゴリズム

1. 2つのアーム$l(t), m(t)$ in [K]を選択
   - 無情報だが、$l(t) = m(t)$の比較も可能（後述）
2. どちらのアームが良いかのフィードバックを得る（$l(t)$ or $m(t)$）

目的：Regretの最小化（後述）

インターリービング比較と比較バンディット問題の対応関係

比較バンディット問題	インターリービング
ラウンド	ユーザの来訪
アーム	ランキング手法
フィードバック	インターリービングで、どちらのランキング手法由来のページをクリックしたかどうか
Regretの最小化	ユーザの効用の最大化（後述）

$l(t) = m(t)$はインターリービング比較なし（ユーザに単一ランキングを提示）に対応

確率的仮定

- 確率的仮定：選好行列$M = \{\mu_{i,j}\} \in (0,1)^{K\times K}$が存在し、アームのペア$(i,j)$を比較したときに$i$が好まれる確率が$\mu_{i,j}$.
- Mは歪対称：$1 - \mu_{j,i} = \mu_{i,j}$.

比較バンディット問題：選好行列の例

- arXivにおける６つの検索エンジンのランキングアルゴリズムの間での選好行列 [Yue+ ICML2011]

$\mu_{2,3}$

	1	2	3	4	5	6
1	0.50	0.55	0.55	0.54	0.61	0.61
2	0.45	0.50	0.55	0.55	0.58	0.60
3	0.45	0.45	0.50	0.54	0.51	0.56
4	0.46	0.45	0.46	0.50	0.54	0.50
5	0.39	0.42	0.49	0.46	0.50	0.51
6	0.39	0.40	0.44	0.50	0.49	0.50

比較バンディット問題における最も良いアームの定義

- 先ほどの場合は、アームに順序が存在
  - 順序: $1 > 2 > 3 > 4 > 5 > 6$であり、
  - $i > j$なら$\mu_{i,j} > 0.5$（iがjより好まれる）
- 順序の仮定は実データでは成立しないことが多い
  - MS検索エンジンのデータセット[MSR 2010, Zoghi+ WSDM2014]では、ランキング手法128個の間に完全な順序が存在しない

比較バンディット問題：選好行列の例

下の例（MS検索エンジンのランキング手法間の比較、部分行列）は、順序が存在しない

	1	2	3	4	5	6
1	0.50	0.36	0.36	0.34	0.36	0.36
2	0.64	0.50	0.49	0.46	0.48	0.47
3	0.64	0.51	0.50	0.48	0.49	0.51
4	0.66	0.54	0.52	0.50	0.53	0.52
5	0.64	0.52	0.51	0.47	0.50	0.49
6	0.64	0.53	0.49	0.48	0.51	0.50

コンドルセ勝者 (Condorcet winner)

- 最低限、一番良いアームが定義できるのはどのような仮定を置けばよいか？
- あるアームi^*（コンドルセ勝者[Urvoy+ ICML2013]）が存在し、$\forall_{j \neq i^*} \mu_{i^*,j} > 1/2$

比較バンディット問題：選好行列の例

- 順序のある例では、アーム１（順序１位）がコンドルセ勝者

	1	2	3	4	5	6
1	0.50	0.55	0.55	0.54	0.61	0.61
2	0.45	0.50	0.55	0.55	0.58	0.60
3	0.45	0.45	0.50	0.54	0.51	0.56
4	0.46	0.45	0.46	0.50	0.54	0.50
5	0.39	0.42	0.49	0.46	0.50	0.51
6	0.39	0.40	0.44	0.50	0.49	0.50

比較バンディット問題：選好行列の例

- MS検索エンジンのランキング手法間では、アーム4がコンドルセ勝者

	1	2	3	4	5	6
1	0.50	0.36	0.36	0.34	0.36	0.36
2	0.64	0.50	0.49	0.46	0.48	0.47
3	0.64	0.51	0.50	0.48	0.49	0.51
4	0.66	0.54	0.52	0.50	0.53	0.52
5	0.64	0.52	0.51	0.47	0.50	0.49
6	0.64	0.53	0.49	0.48	0.51	0.50

比較バンディット問題（再掲）
Dueling bandit problem [Yue+ COLT2009]

- 入力：アーム数K
- 各ラウンド $t = 1, \ldots, T$に、
  1. 2つのアーム$l(t), m(t)$ in [K]を選択
  2. どちらのアームが良いかのフィードバックを得る（$l(t)$ or $m(t)$）

アームを選択するアルゴリズム

目的：Regretの最小化（次頁で定義）

評価手法：Regret

- ユーザに最も良いランキング手法を提示したい
- コンドルセ勝者i^*をインターリービング比較せずに提示するのが最も良い（$(l(t), m(t)) = (i^*, i^*)$）
- $\mu_{i^*,j} - 1/2 \geq 0$: アームi^*とjの間の良さの差分
- $r_{i,j} = \frac{(\mu_{i^*,i} + \mu_{i^*,j}) - 1}{2} \geq 0$: アームのペア$(i, j)$をインターリービング比較したときのユーザの効用の損失
- Regret(T) $= \sum_{t=1}^{T} r_{l(t),m(t)}$
  - $(l(t), m(t)) = (i^*, i^*)$でない限りregretが増加
  - Regretの最小化：i^*をasapで発見し、(i^*, i^*)を残りのラウンドで選択したい

概要（再掲）

- 問題設定：比較バンディット問題
  - モチベーション：検索エンジンのランキング手法比較
- 理論解析とアルゴリズム提案
  - アルゴリズムの性能 (Regret) 限界
  - アルゴリズム：RMED
    - 性能限界を漸近的に達成する現在唯一のアルゴリズム
- 数値実験による性能比較

本研究の主貢献

Regret漸近下限（アルゴリズムの性能限界）

- 強一致性：任意のコンドルセ勝者の存在する選好行列M, 実数$a > 0$に対し、Regretの期待値が$o(T^a)$.
- 主結果１：強一致なアルゴリズムに関して、以下が成立

$$\mathrm{E}[\mathrm{Regret}(T)] \geq \sum_{i \neq i^*} \min_{\mathrm{j}:\mu_{i,j}<\frac{1}{2}} \frac{r_{i,j}}{d_{KL}\left(\mu_{i,j}, \frac{1}{2}\right)} \log T - o(\log T).$$

  - 通常のバンディット問題における結果[Lai&Robbins 1985] の拡張

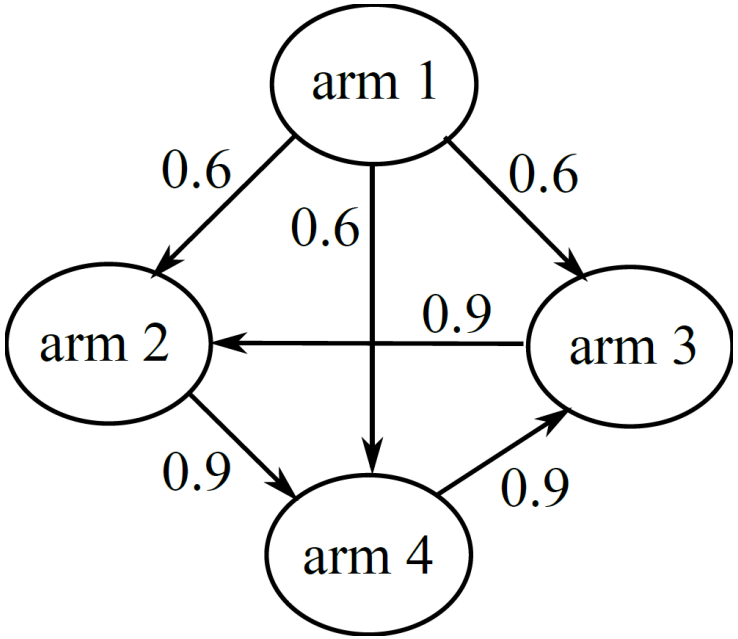
Regret漸近下限：導出

- 強一致性→各ラウンドtに有意水準$1/t$で、各アーム$i \neq i^*$がコンドルセ勝者でないことを示す必要性
- アームiがいずれか１つのアームjに対して$\mu_{i,j} < 1/2 \rightarrow i$はコンドルセ勝者（$\forall_{j \neq i^*}\ \mu_{i^*,j} > 1/2$）ではない
  - 比較回数$\frac{\log t}{d_{KL}\left(\mu_{i,j}, \frac{1}{2}\right)}$で有意水準$1/t$に（大偏差原理）
- $b^*(i) = \underset{\mathrm{j}:\mu_{i,j}<1/2}{\operatorname{argmin}} \frac{r_{i,j} \log t}{d_{KL}\left(\mu_{i,j}, \frac{1}{2}\right)}$と比較するのがRegretが最小
  - $d_{KL}\left(\mu_{i,j}, \frac{1}{2}\right)$が大きい（$i$が大きく負けている$j$）ものと比較したい

1比較あたりのRegret ×必要比較回数

極端な例: $b^*(i) \neq 1$

- コンドルセ勝者＝アーム1
- しかし、アーム2がコンドルセ勝者ではないことを示すためには、アーム1よりアーム3と比較するほうが早くできる（$b^*(2) = 3$)



提案：Relative Minimum Empirical Divergence (RMED) アルゴリズム

- 通常のバンディット問題のアルゴリズムDMED[Honda&Takemura 2010]を比較バンディット問題に拡張
- 定義： $I_i(t) = \sum_{j \in [K]: \hat{\mu}_{i,j} \leq \frac{1}{2}} N_{i,j}(t) d_{KL}\left(\hat{\mu}_{i,j}, \frac{1}{2}\right).$

iが負けている相手j

iとjの比較回数

Bernoulli(1/2)からのKL距離（どれだけ負けているか）

- $\exp(-I_i(t))$はアームiがコンドルセ勝者である”尤度”

RMEDアルゴリズム：概要

初期化：すべてのペアを$O(\log\log T)$回比較

while $t < T$ do

$b^*(i)$の推定精度を確保

・リストの生成：すべての$\exp(-I_i(t)) > 1/t$を満たすアームをリストに入れる

・リストの中のそれぞれのアームiに関して、推定される$b^*(i)$と１回ずつ比較

RMEDの漸近最適性

- 主結果２：RMEDに関して以下が成立

$$\mathrm{E}[\mathrm{Regret}(T)] \leq \sum_{i \neq i^*} \min_{\mathrm{j}:\mu_{i,j}<\frac{1}{2}} \frac{r_{i,j}}{d_{KL}\left(\mu_{i,j}, \frac{1}{2}\right)} \log T + o(\log T)\,.$$

- 主結果１（再掲）：強一致なアルゴリズムに関して、以下が成立

定数係数が一致（漸近最適）！

$$\mathrm{E}[\mathrm{Regret}(T)] \geq \sum_{i \neq i^*} \min_{\mathrm{j}:\mu_{i,j}<\frac{1}{2}} \frac{r_{i,j}}{d_{KL}\left(\mu_{i,j}, \frac{1}{2}\right)} \log T - o(\log T)\,.$$

各アーム$i \neq i^*$を$b^*(i)$と比較し、コンドルセ勝者である仮説を有意水準$1/T$で棄却

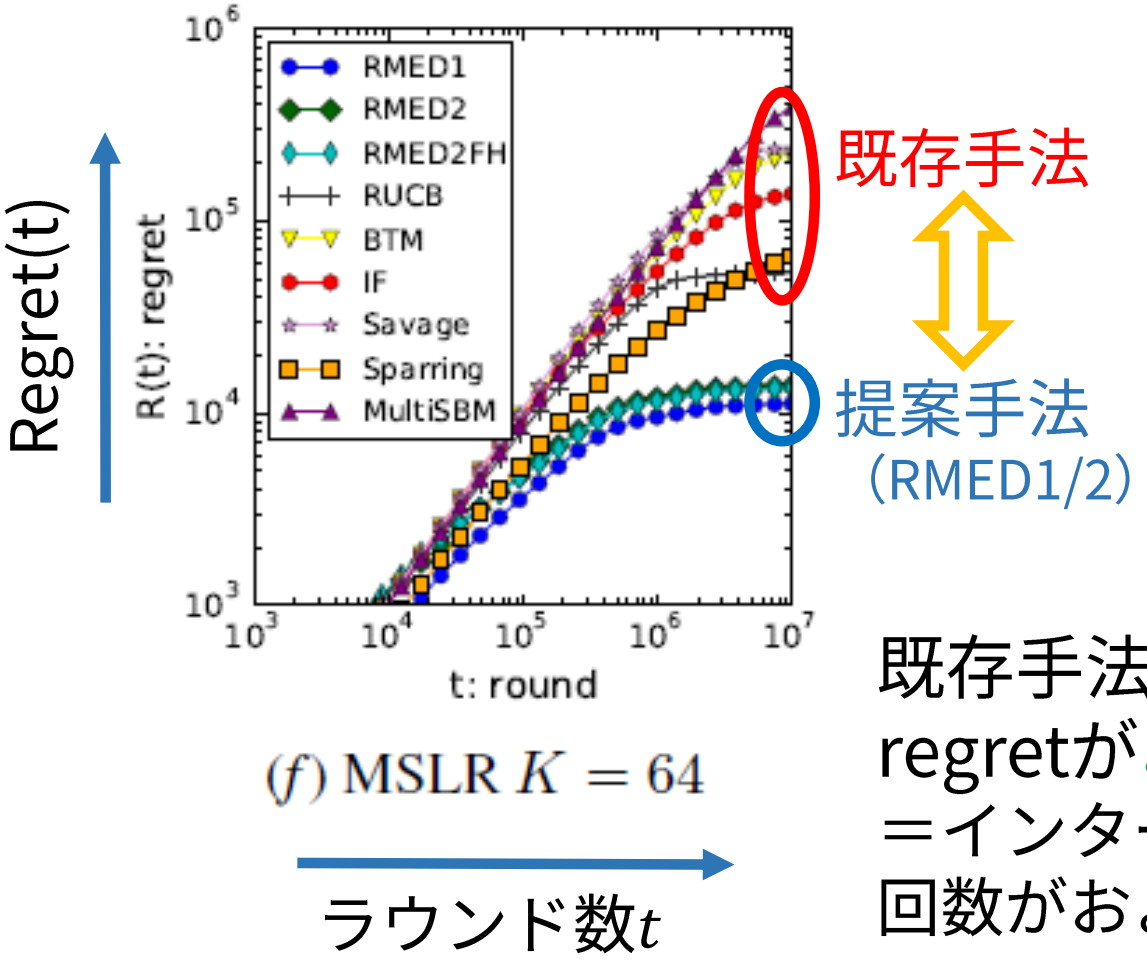
概要（再掲）

- 問題設定：比較バンディット問題
  - モチベーション：検索エンジンのランキング手法比較
- 理論解析とアルゴリズム提案
  - アルゴリズムの性能 (Regret) 限界
  - アルゴリズム：RMED
    - 性能限界を漸近的に達成する現在唯一のアルゴリズム
- 数値実験による性能比較

数値実験

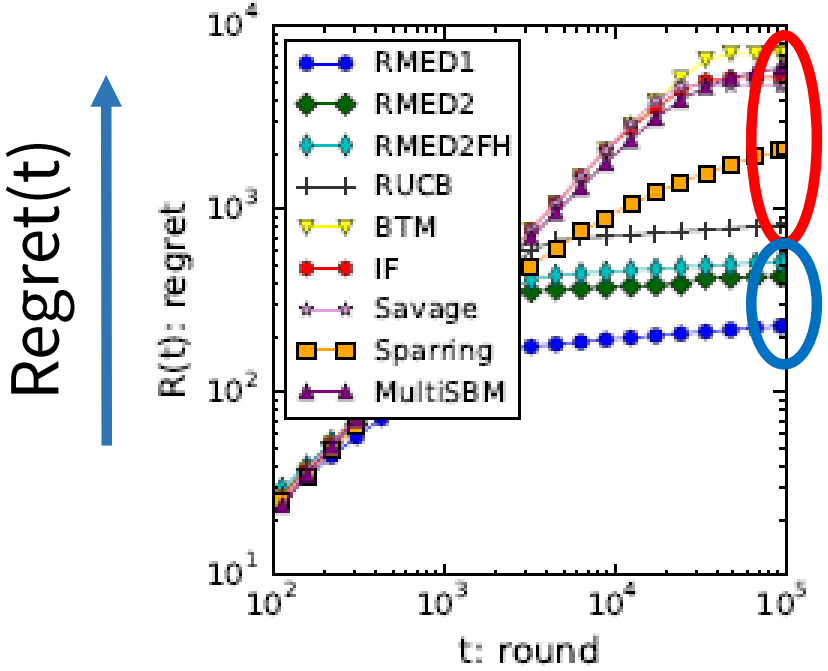
- 提案手法および既存の比較バンディット問題の手法をデータセット（実データから作成した選好行列）で比較
  - Microsoft Learning to Rank (MSLR) データセット [Microsoft 2010]：（アーム＝ランキング手法）
  - Sushiデータセット：Kamishima [KDD2003]による寿司の好みのデータセット（アーム＝寿司）
    - 寿司16種に絞る、コンドルセ勝者は中トロ

実験での性能比較：MSLR



既存手法と比べ
regretが**およそ1/5**
＝インターリービング比較
回数がおよそ1/5

実験での性能比較：Sushi



(*d*) Sushi

ラウンド数t

既存手法

提案手法（RMED1/2）

既存手法と比べ
regretが**およそ1/4**

まとめ

- 一対比較のフィードバックを通じて最も良い選択肢（アーム）を探す問題である比較バンディット問題を扱った
- 主結果１：理論性能限界（Regret漸近下限）の導出を行った
- 主結果２：最適な理論性能を漸近的に達成するアルゴリズムRMEDを提案した

今後の展望（future work）１

- 本研究ではコンドルセ勝者の存在を仮定
- コンドルセ勝者は必ずしも存在するとは限らない（例：トップの三すくみ構造）
- 一般の場合は？
  - 比較ベースで最も良いものを選ぶ場合、複数の選択基準が存在
  - コープランド勝者：コンドルセ勝者の一般化 [Urvoy+ ICML2013, Zoghi+ NIPS2015]

今後の展望（future work）２

- バンディット問題での有限時間性能：
  Thompson sampling > UCB > DMED
  - TSやUCBベースでRMEDと同様の漸近最適アルゴリズムの構築は可能か？
- パーソナライゼーションを考慮した拡張
  - [Dudik+ COLT2015]がある程度行っている