

論文読み会：*“Identifying New Podcasts with  
High General Appeal Using a Pure Exploration  
Infinitely-Armed Bandit Strategy”*, Recsys 2022

宮本 隆志

ナビプラス株式会社

2023/06/14

自己紹介

紹介する論文

podcast 推薦の性質

提案手法

具体的イメージ

既存手法との比較

まとめ

予備

Empirical CBT

Successive Rejects

# 自己紹介

機械学習・レコメンドシステムの論文読み会に参加したくて、このたび参加させて頂きました。よろしくお願いします。

- ▶ 名前 (twitter): 宮本 隆志 (@tmiya\_)
- ▶ 所属: ナビプラス株式会社
  - ▶ 主に EC ショップ用の推薦システムの開発と、特許関係の業務をやっています。
- ▶ 職歴: 某外資系企業 (液晶製造 R&D → SW 開発 → SE) → 某ベンチャー (広告推薦の DSP) → 現職
- ▶ 興味: 機械学習・レコメンドシステム・関数型言語・形式手法
- ▶ 社外勉強会: 週 1 で機械学習や統計学に近いところの教科書の輪読してます。
  - ▶ 今は 5 人で佐古「ゲージ理論・一般相対論のための微分幾何入門」を読んでいます。(そのうち情報幾何の本を読みたい)

# 紹介する論文

*"Identifying New Podcasts with High General Appeal Using a Pure Exploration Infinitely-Armed Bandit Strategy" (Recsys'22)*

- ▶ Spotify の論文
- ▶ 新規 podcast の推薦: cold-start 問題 → 多腕バンディットを使いたい
- ▶ しかし、許される試行回数に対して腕の数が多い → 無限腕バンディットという定式化
- ▶ オフライン評価例: 既存の  $\text{lil'UCB}$  アルゴリズムでランダムに腕を  $X$  個選択したものと比較
  - ▶  $\text{lil'UCB-X}$  の場合、収束の速さと損失とトレードオフになる
    - ▶ 試行回数  $\approx$  腕の数  $\approx 2^{12}$  だと、100 腕以上は無理
  - ▶ 提案手法 (ISHA) は腕の数を制限せずに、性能低下も無い

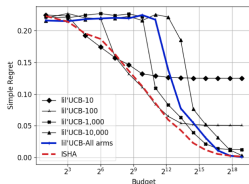


Figure 1: 3,795 arms. Difference of the recommended and optimal arm's mean as a function of total pulls.

# podcast 推薦の性質 (ここ重要では無い)

## ▶ P@K のオフライン評価:

**Consumption** 累積閲覧数・合計視聴時間・合計視聴者数・フォロワー数

**Content** トピック (one-hot enc.)・エピソード数・エピソード間の間隔・エピソード名の長さ・難しい単語が使われているか、など内容に (ある程度) 関係した指標

**SortByPopularity** 過去 30 日間の人気 podcast

## ▶ 30 日以内だと SortByPopularity が強い。通常の推薦方法は新規 podcast 推薦の役に立たない → bandit が有用

▶ (真面目に content-based な推薦してないからでは???)

	Podcast launch day								
	Consumption			Content			Consumption+content		
Model	k=50	k=100	k=150	k=50	k=100	k=150	k=50	k=100	k=150
LinearReg.	0.220	0.180	0.127	0.020	0.010	0.007	0.220	0.170	0.147
Lasso	0.280	0.200	0.167	0.020	0.010	0.007	0.260	0.160	0.120
AdaBoost	0.180	0.090	0.060	0.000	0.000	0.013	0.080	0.080	0.060
RandomForest	0.200	0.100	0.073	0.000	0.000	0.007	0.160	0.090	0.060
<b>SortByPopularity</b>	<b>0.360*</b>	<b>0.240*</b>	<b>0.180*</b>	—	—	—	—	—	—
30 days after podcast launch									
LinearReg.	0.860	0.720	0.607	0.080	0.050	0.033	0.640	0.530	0.393
Lasso	0.860	0.720	0.607	0.080	0.040	0.033	0.660	0.530	0.393
AdaBoost	0.400	0.400	0.427	0.000	0.020	0.013	0.800	0.660	0.573
RandomForest	0.640	0.580	0.500	0.020	0.020	0.027	0.800	0.640	0.513
<b>SortByPopularity</b>	<b>0.880*</b>	<b>0.740*</b>	<b>0.613*</b>	—	—	—	—	—	—
60 days after podcast launch									
LinearReg.	0.820	0.890	0.813	0.140	0.090	0.073	0.820	0.840	0.767
Lasso	0.820	0.890	0.813	0.140	0.090	0.073	0.820	0.840	0.787
AdaBoost	<b>0.960</b>	<b>0.900</b>	0.713	0.040	0.020	0.020	<b>0.960</b>	<b>0.900</b>	0.707
RandomForest	0.940	0.770	0.733	0.060	0.040	0.040	0.860	0.810	0.767
<b>SortByPopularity</b>	0.900	0.880	<b>0.813*</b>	—	—	—	—	—	—

# 提案手法: Successive Halving

1. 初期状態: 腕の集合を  $S_0$ 、腕の総数を  $n = |S_0|$ 、予算を  $T = \lceil n \log_2 n \rceil$  とする
2. 各ラウンド  $t = 0, 1, 2, \dots, \lceil \log_2 n \rceil - 1$ 
  - ▶ 腕の集合  $S_t$  に対して、各腕を  $\frac{T}{|S_t| \lceil \log_2 n \rceil}$  回ずつ引く
  - ▶ 得られた経験平均  $\hat{\mu}_{i,t}$  を見て、下位半分を捨てて腕の数を半分にする
3. 2 を繰り返して腕を絞り込む

各腕の試行結果を全部使う改良版が Unforgetful ISHA (UISHA)。

---

**Successive Halving.** ISHA chooses  $n \in \mathbb{N}$  so that  $T = \lceil n \log_2(n) \rceil$ .  
For anytime, double  $n$  and repeat.

---

**Input:** Budget  $T$ , number of arms  $n$

Draw  $n$  arms and add them to  $S_0$

**for**  $t = 0, 1, 2, \dots, \lceil \log_2(n) \rceil - 1$  **do**

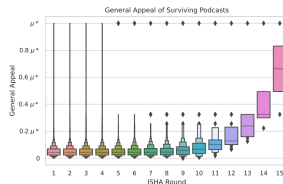
    Pull each arm  $i \in S_t$  for  $\left\lfloor \frac{T}{|S_t| \lceil \log_2(n) \rceil} \right\rfloor$  times and compute  
    empirical means  $\hat{\mu}_{i,t}$

    Set  $S_{t+1}$  to be  $\lceil |S_t|/2 \rceil$  arms with the highest empirical means

$\hat{\mu}_{i,t}$ , breaking ties at random

**return** Single arm in  $S_{\lceil \log_2(n) \rceil}$

---



(b) General Appeal of survivors in each round of ISHA.  
This gives a noisy partial ordering by general appeal.

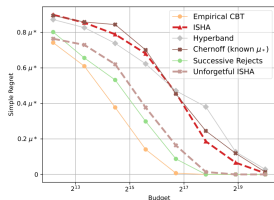
# 具体的イメージ

腕が 1,024 個ある場合。

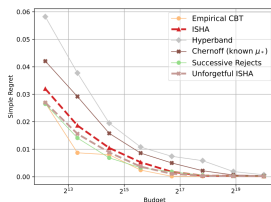
1. 1,024 個の腕を 1 回ずつ試して、成績の悪い腕を半分捨てる。
  2. 残った 512 個の腕を 2 回ずつ試して、成績の悪い腕を半分捨てる。
  3. 残った 256 個の腕を 4 回ずつ試して、成績の悪い腕を半分捨てる。
  4. 残った 128 個の腕を 8 回ずつ試して、成績の悪い腕を半分捨てる。
  5. 残った 64 個の腕を 16 回ずつ試して、成績の悪い腕を半分捨てる。
  6. 残った 32 個の腕を 32 回ずつ試して、成績の悪い腕を半分捨てる。
  7. 残った 16 個の腕を 64 回ずつ試して、成績の悪い腕を半分捨てる。
  8. 残った 8 個の腕を 128 回ずつ試して、成績の悪い腕を半分捨てる。
  9. 残った 4 個の腕を 256 回ずつ試して、成績の悪い腕を半分捨てる。
  10. 残った 2 個の腕を 512 回ずつ試して、成績の悪い腕を半分捨てる。
- ▶  $n \log_2 n = 10,240$  回で最適腕が決まる。
  - ▶ round 1 の割り切り方がすごい。真の最適腕を捨てるリスクを恐れてない。
  - ▶ explicit rating みたいに採点されるならまだ良いが、0/1 評価の implicit feedback で CTR=1%とかだとかなり辛そう。

# 既存手法との比較

- ▶ 既存手法の Empirical CBT[10], Successive Rejects[1] に負けてる。
- ▶ podcast の場合、試行は即時に報酬が得られない（そもそもコンテンツを表示して、ユーザが評価する時間がかかる）
- ▶ 各ラウンドの試行回数が少なく、試行の報酬を待たないと次のラウンドに進めないアルゴリズムは実案件では使いにくい。
  - ▶ ECBT は腕の数だけラウンドが必要、SR もラウンド数が多い。
  - ▶ 提案手法は複数サーバで並列試行可能。



(a) Podcast Reservoir



(b) Movielens-1M



# まとめ

- ▶ 新規 podcast のレコメンドは content-based が上手くいかないので、cold-start 問題。
- ▶ 試行回数に比べてアイテムが多いので通常の bandit では困難。無限腕 bandit という定式化。
- ▶ Successive Halving アルゴリズムを提案。
- ▶ 既存手法 (ECBT, SR) には負けてるが、それらより実案件で使いやすいアルゴリズム。
- ▶ 感想
  - ▶ content-based はもっと頑張れるのでは？ (DL 使って素性作ったりとか)
  - ▶ 実案件で使いやすいアルゴリズムというのは良いが、0/1 な implicit feedback な商品推薦では使いにくそう。
  - ▶ 最初のラウンド ( $t = 0$ ) って、 $n$  個のアイテム全部を 1 回ずつ試して半分捨てる、みたいな感じになってる。全部のアイテムを 1 回ずつ試す、というのは通常の商品レコメンドではほぼ無いが、popularity bias とか考えると、とにかく全商品試すというのは本当は必要。

# Empirical Confidence Bound Target

Notation:  $m$  = total rewards,  $n_k(m) = \#$  of rewards from arm  $k$ ,  $K_m = \#$  of arms played,  $S'_m$  = sum of the losses  $X_{kt}$  over the first  $m$  trials,  $\zeta(m) = \frac{S'_m}{n}$ .

$$L_{kt} = \max \left( \frac{\bar{X}_{kt}}{b_n}, \bar{X}_{kt} - c_n \frac{\hat{\sigma}_{kt}}{\sqrt{t}} \right), \quad b_n = c_n = \log \log n.$$

Algorithms:

- ▶ For  $m = 0$ , play arm 1.  
( $\therefore K_1 = 1, n_1(1) = 1, n_k(1) = 0$  for  $k > 1$ .)
- ▶ For  $m = 1, \dots, n - 1$ :
  - ▶ If  $\min_{1 \leq k \leq K_m} L_{kn_k} \leq \zeta(m)$ , then play the arm  $k = \operatorname{argmin}_k L_{kn_k(m)}$ .
  - ▶ If  $\min_{1 \leq k \leq K_m} L_{kn_k} > \zeta(m)$ , play a new arm  $K_m + 1$ .

番号の小さい arm に最適腕があると、もっと良い腕があるのでは、と探索を増やしがち。

# Successive Rejects

Notation:  $n = \text{budget}$ ,  $A_1 = \{1, \dots, K\}$ ,  $\overline{\log}(K) = \frac{1}{2} + \sum_{i=2}^K \frac{1}{i}$ ,

$$n_0 = 0, n_k = \left\lceil \frac{1}{\overline{\log}(K)} \frac{n - K}{K + 1 - k} \right\rceil \text{ for } 1 \leq k \leq K - 1.$$

Algorithms:

- ▶ For each phase  $k = 1, 2, \dots, K - 1$ :
  - ▶ For each  $i \in A_k$ , select arm  $i$  for  $n_k - n_{k-1}$  rounds.
  - ▶  $A_{k+1} = A_k \setminus \operatorname{argmin}_{i \in A_k} \hat{X}_{i, n_k}$ .