衝突検知情報なしマルチプレイヤー多腕バンディット問題の並列通信アルゴリズム

Multi-Channel Communication Algorithm for Multi-Player Multi-Armed Bandits without Collision Sensing

泉 知成 / 北海道大学大学院 情報科学院 情報科学専攻 情報理工学コース アルゴリズム研究室

研究の背景・動機

- 近年，IoT や **認知無線** の発展に伴いリソース割当問題の重要性が増す中，中央制御を持たない 分散型 **マルチプレイヤー多腕バンディット** (**MP-MAB**) が注目されている [Boursier 24].
- 「**衝突検知情報なし**」という他プレイヤーとの衝突を直接観測できない最も困難な設定に対し，期待値の高い1つの有望なアーム（Good Arm）を通信チャネルとする手法が提案されている [Huang 22].
- しかし，単一チャネルでの逐次通信は他プレイヤーの待機時間がボトルネックとなり，プレイヤー数の増加に伴い通信リグレットが悪化する問題があった.

研究の目的

- 既存手法を拡張し，**n 本の Good Arm を通信チャネル**として活用する並列分散アルゴリズムを提案する.
- 同期・通信プロセスの並列化により，待機時間のボトルネックを解消し，特定の環境下において提案手法が優れたリグレット上界を達成することを示す.

問題設定

- 本研究では，アームが K 本，プレイヤーが M 人 $(M < K)$からなる 分散型 MP-MAB を扱う (図1).

アーム k	1,	⋯	K	時刻 $t = 1, \dots, T$
k **の本来の報酬** $X_k(t)$	$X_1(t)$,	⋯	$X_K(t)$	• **アーム** k を選択した人数：$N_k(t) = \sum_{m=1}^{M} \mathbb{I}\{\pi_m(t) = k\}$
i.i.d. Bernoulli with mean μ_k $X_k(t) \in \{0,1\}$		⋯		
選択したアーム $\pi_m(t)$	$\pi_1(t)$,	$\pi_2(t)$, ⋯	$\pi_M(t)$	• **衝突指示子**：$\eta_k(t) = \mathbb{I}\{N_k(t) \geq 2\}$ (※ $\mathbb{I}\{\cdot\}$ は指示関数)
$\pi_m(t) \in \{1, \dots, K\} = [K]$				
プレイヤー m $(M < K)$	1,	2, ⋯	M	• **報酬**：$r_m(t) = X_{\pi_m(t)}(t)[1 - \eta_{\pi_m(t)}(t)]$
m **が獲得する報酬** $r_m(t)$	$r_1(t)$,	$r_2(t)$ ⋯	$r_M(t)$	

図1. 分散型 MP-MAB の問題設定

- 衝突が発生していない時の各アームの期待報酬を降順に並べたものを $\mu_{(1)}, \dots, \mu_{(K)}$ と定義し，$\mu_{(1)} \geq \mu_{(2)} \geq \cdots \geq \mu_{(K)}$ が成り立つとする. (ただし，プレイヤーはこの順序や具体的な値を知らない.)
- プレイヤーは，各時刻で必ず 1つだけ アームを引く.
- **衝突情報検知なし**：各プレイヤーは，自身の報酬 $r_m(t)$ のみを観測し，それ以外の情報 (他者の行動，衝突指示子 $\eta_k(t)$，通信など) を一切受け取らない. (報酬が 0 だったとき，その原因が衝突によるものか，単にアームの報酬が外れただけなのかをプレイヤーは区別でいない.)
- **目的**：以下の累積リグレット $R(T, \mu)$ を最小化する:

$$R(T,\mu) = \sum_{t=1}^{T} \sum_{m=1}^{M} \mu_{(m)} - \mathbb{E}[r_m(t)]$$

提案するアルゴリズム

- 提案手法は，4つのフェーズによって構成される (図2).

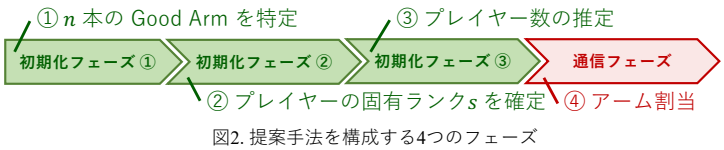
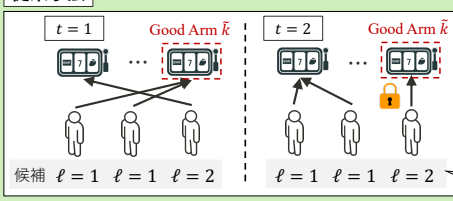


図2. 提案手法を構成する4つのフェーズ

- **初期化フェーズ ①** では，通信路として使用可能な n 本の Good Arm を特定する. (特定した Good Arm $\tilde{k}_i$ は番号の小さい順から i 番目.)

■ **初期化フェーズ ②**：椅子取りゲーム　(以降，従来手法は [Huang 22])

従来手法 正の報酬を得たら $s = \ell$ としてランクを確定し，以降は $\ell = s$ を固定する.



(**for** $t = 1, \dots, K\tau$ **do**)
- $\tilde{k}$ を引く $(t \bmod K = \ell)$
- $\tilde{k}$ 以外を引く (otherwise)

ただし，$\tau = \lceil \ln(\frac{1}{\delta}) / \tilde{\mu}_{\min} \rceil$

$t \bmod K = 1$ のとき候補を $\{1, \dots, K\}$ からランダムに設定

提案手法 例) 各プレイヤーが3つの候補 ℓ_1, ℓ_2, ℓ_3 を持つ場合. Good Arm $\mathcal{G} = \{\tilde{k}_1, \tilde{k}_2, \tilde{k}_3\}$

	t	1	2	3	4	5	⋯
($\tilde{k}_1$ を使う) $\ell_1 \rightarrow$	$t \bmod K$	1	2	3	4	5	⋯
($\tilde{k}_2$ を使う) $\ell_2 \rightarrow$	$t \bmod K + 1$	2	3	4	5	6	⋯
($\tilde{k}_3$ を使う) $\ell_3 \rightarrow$	$t \bmod K + 2$	3	4	5	6	7	⋯

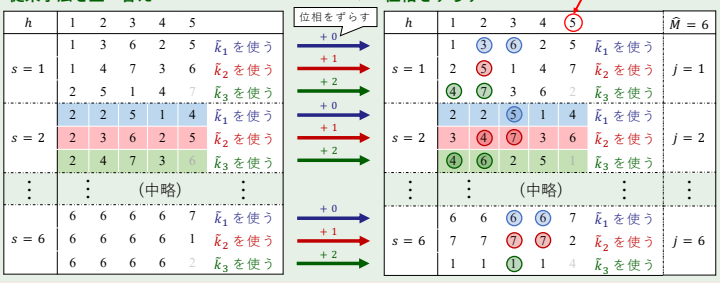
(**for** $t = 1, \dots, \lceil K\tau/n \rceil$ **do**)
- $\tilde{k}_i$ を引く $(\forall i \in [n]\ \text{s.t.}\ ((t + i - 2) \bmod K + 1) = \ell_i)$
- $\tilde{k}_i$ 以外を引く (otherwise)

■ **初期化フェーズ ③**：他者と必ず1度だけ衝突し，プレイヤー数を推定

従来手法：$K = 7, M = 6$ $(h > 2s$ でラウンドロビン)

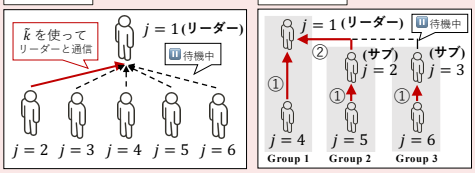
提案手法：$K = 7$, $M = 6$, Good Arms: $\mathcal{G} = \{\tilde{k}_1, \tilde{k}_2, \tilde{k}_3\}$

従来手法を並べ替え：　**位相をずらす**：　$\lceil 2K/n \rceil$ に短縮



■ **通信フェーズ**：フォロワーの観測結果をリーダーに集約・アーム割当

従来手法 リーダー ↔ フォロワー　**提案手法** リーダー ↔ サブリーダー (サブ) ↔ フォロワー



例) Good Arm $\mathcal{G} = \{\tilde{k}_1, \tilde{k}_2, \tilde{k}_3\}$ の時

①
- Group 1：$\tilde{k}_1$ を使って通信
- Group 2：$\tilde{k}_2$ を使って通信
- Group 3：$\tilde{k}_3$ を使って通信

② サブ→リーダー：$\tilde{k}_1$ を使って通信

※アーム割当は ②→① の逆順で通信

結論 (リグレット上界)

- 第2項 (初期化リグレット) が支配的な場合，提案手法は従来手法に劣る.
- 一方で，第3項 (通信リグレット) がシステム全体のボトルネックになる場合は優位性をもつ. 具体的には，アームの期待値の減衰が緩やかで $n\mu_{(n)} > \mu_{(1)}$ を満たす環境下では，学習効率を向上させる.

従来手法：$R_{\text{prev}}(T,\mu) \leq \underbrace{O\left(\sum_{k>M} \frac{\ln T}{\Delta_k}\right)}_{\text{理論限界}} + \underbrace{O(K^2 M \ln T)}_{\text{初期化リグレット}} + \underbrace{O\left(KM^2 \ln\left(\frac{1}{\Delta_M}\right)^2 \ln T\right)}_{\text{通信リグレット}}$

(ただし，$\Delta_k = \mu_{(M)} - \mu_{(k)}$, $\Delta_M = \mu_{(M)} - \mu_{(M+1)}$)

提案手法：$R_{\text{prop}}(T,\mu) \leq O\left(\sum_{k>M} \frac{\ln T}{\Delta_k}\right) + O\left(K^2 M \frac{\mu_{(1)}}{\mu_{(n)}} \ln(nT)\right)$

$$+ O\left(\frac{KM^2}{n} \frac{\mu_{(1)}}{\mu_{(n)}} \ln\left(\frac{1}{\Delta_M}\right)^2 \ln(nT)\right)$$

参考文献

[Boursier 24] Boursir, E. and Perchet, V.: A Survey on Multi-player Bandits, *JMLR*, Vol. 25, No.137, pp. 1−45, 2024.

[Huang 22] Huang et al.: Towards Optimal Algorithms for Multi-player Multi-armed Bandits without Collision Sensing Information, *COLT*, Vol. 178, pp. 1990−1012, 2022.