A General Framework for Clustering and Distribution Matching with Bandit Feedback

Recep Can Yavas

7 Oct. 2024

Group meeting

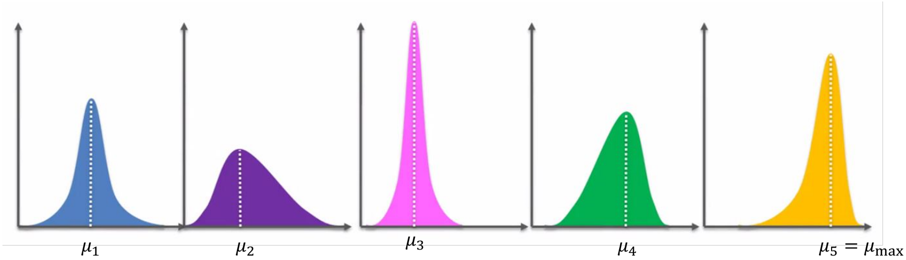
Joint work with Yuqi Huang, Vincent Tan, and Jonathan Scarlett

Outline

- Problem formulation
- Lower Bound
- Algorithm
- Theoretical and experimental results

Multi-armed bandits (MABs)

Motivation: A suitable mathematical model for applications such as drug design, online advertisement, and online recommendation systems



There are two main objectives:

1. **Regret minimization:** Maximize the total reward after T pulls
2. **Pure exploration:** Answer a specific question about K unknown distributions
   e.g., best arm identification, odd arm identification, ϵ-good arm identification

Clustering and Distribution Matching Problem

Any pure exploration problem can be viewed as a sequential multi-hypothesis testing with bandit feedback [Prabhu et al. 2022].

- We are given K arms.
- Arm distributions are on a finite alphabet $\mathcal{X}$.
- Each hypothesis σ is denoted by a partition of a subset of $[K]$

  $$\sigma = \{\mathcal{A}_1^\sigma, \dots, \mathcal{A}_M^\sigma\}$$

  where $\mathcal{A}_m^\sigma$'s are disjoint and $\cup_{m \in [M]} \mathcal{A}_m^\sigma \subseteq [K]$.
- For each $m \in [M]$, $\mathcal{A}_m^\sigma$ indicates a cluster with identical distributions $\implies$ arm $i, j \in \mathcal{A}_m^\sigma$, then $P_i = P_j$.
- $\mathcal{A}_{M+1}^\sigma \triangleq [K] \setminus \cup_{m \in [M]} \mathcal{A}_m^\sigma$ is called the unconstrained group, which doesn't restrict the arm distributions in it.
- We assume $|\mathcal{A}_m^\sigma| \geq 2$ for $m \leq M$. Hence, $K \geq 2M$.
- Arms from distinct subsets in $\{\mathcal{A}_1^\sigma, \dots, \mathcal{A}_{M+1}^\sigma\}$ follow distinct distributions.

Clustering and Distribution Matching Problem

Let $A_t \in [K]$ be the arm pulled at time t. Let $X_{t,A_t} \in \mathcal{X}$ be the reward at time t from arm A_t.

- We design an online algorithm: A_t may depend only on $(A_1, X_{1,A_1}, A_2, X_{2,A_2}, \ldots, A_{t-1}, X_{t-1,A_{t-1}})$.
- Let $P = (P_1, \ldots, P_K)$ be the underlying problem instance whose hypothesis is σ_P.
- **Fixed confidence:** Algorithm stops at a random time τ and outputs a hypothesis $\hat{\sigma}(\tau) \in \mathcal{C}$.

Goal: Design an online algorithm such that

1. δ-correct: $\mathbb{P}\left[\hat{\sigma}(\tau) \neq \sigma_P\right] \leq \delta$ and $\mathbb{P}\left[\tau < \infty\right] = 1$
2. $\mathbb{E}\left[\tau\right]$ as small as possible

Distinguishability of Hypotheses

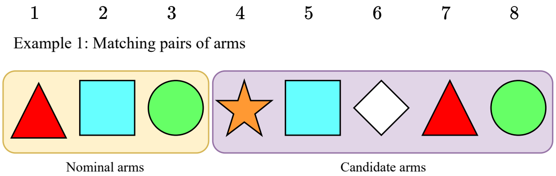
A hypothesis σ is said to *dominate* another hypothesis σ' if every subset in the partitioning of σ, $\mathcal{A}^{\sigma}_{[M]}$, is a subset of some subset in the partitioning of σ', $\mathcal{A}^{\sigma'}_{[M]}$.

Consider two hypotheses $\sigma_1 = \{\{1,2\},\{4,5\}\}$ and $\sigma_2 = \{\{1,2,3\},\{4,5\}\}$ with $M = 2$ and $K = 5$. The equality relations implied by σ_1 ($P_1 = P_2$ and $P_4 = P_5$) are contained in the equality relations implied by σ_2 ($P_1 = P_2$, $P_1 = P_3$, and $P_4 = P_5$). Hence, σ_1 dominates σ_2 (σ_1 is less stringent than σ_2).

Assumption: For a given clustering problem $\mathcal{C}$,

1. there exists no hypothesis pair $(\sigma, \sigma') \in \mathcal{C}^2$ for which $\sigma \neq \sigma'$ and σ dominates σ',
2. for each problem instance $P \in \Lambda$, there exists a unique hypothesis $\sigma \in \mathcal{C}$ such that $P \in \Lambda_\sigma$.

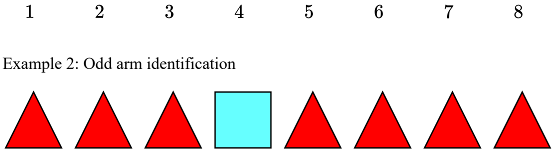
Λ_σ denotes all instances assoc. with σ and Λ denotes all instances included in a given problem.



$$\sigma_P = \{\{1,7\},\{2,5\},\{3,8\}\}$$

- Each nominal arm has exactly 1 match in the set of candidate arms.
- The designer knows which arms are nominal arms (here, $\{1,2,3\}$).
- Offline version of this problem is studied by [Zhou et al. 2024]. The offline is referred to as sequence matching.

Example 2: Odd arm identification



Example 2: Odd arm identification

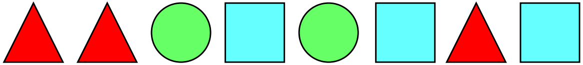
odd arm $= 4$ $\quad \sigma_P = \{\{1, 2, 3, 5, 6, 7, 8\}\}$

- Odd arm follows a different distribution than the rest.
- Studied by [Vaidhiyan and Sundaresan, 2018] and [Karthik and Sundaresan, 2020].

Example 3: M-ary clustering of K arms

1 2 3 4 5 6 7 8

Example 3: M-ary clustering of K arms



$$\sigma_P = \{\{1,2,7\},\{3,5\},\{4,6,8\}\}, K = 8, \text{ and } M = 3$$

- K arms are partitioned into M groups whose size can be as small as 1.
- Highest number of hypotheses for fixed K and M (appr. $M^K/M!$).
- Studied by [Yang et al., 2024] for d-dimensional Gaussian arms. Their algorithm utilizes K-means algorithm, which relies on the fact that the arms are Gaussian distributed.

Preliminary Definitions

- Generalization of (Generalized Jensen-Shannon (GJS) divergence)

$$G(P_{\mathcal{A}}, w_{\mathcal{A}}) \triangleq \begin{cases} 0 & \text{if } w_i = 0, \ \forall i \in \mathcal{A} \\ \sum_{i \in \mathcal{A}} w_i D(P_i \| W) & \text{otherwise} \end{cases}$$

where $W \triangleq \frac{\sum_{i \in \mathcal{A}} w_i P_i}{\sum_{i \in \mathcal{A}} w_i} \in \mathcal{P}(\mathcal{X})$.

- Score functions

$$g_P^{\sigma}(w) \triangleq \sum_{m=1}^{M} G(P_{\mathcal{A}_m^{\sigma}}, w_{\mathcal{A}_m^{\sigma}}) = \sum_{m=1}^{M} \inf_{Q \in \mathcal{P}(\mathcal{X})} \sum_{i \in \mathcal{A}_m^{\sigma}} w_i D(P_i \| Q)$$

$$G_P^{\sigma}(w) \triangleq \min_{\sigma' \in \mathcal{C} \setminus \{\sigma\}} g_P^{\sigma'}(w)$$

$$T(P, \sigma_P) \triangleq \max_{w \in \Sigma_K} G_P^{\sigma_P}(w) = \sup_{w \in \Sigma_K} \inf_{P' \in \mathrm{Alt}(P)} \sum_{i \in [K]} w_i D(P_i \| P_i')$$

Intuition on G

Let $(X_i^{n_i})_{i\in[B]}$ be $B \geq 2$ collection of sequences of lengths $n_1, \ldots, n_B$ from a finite alphabet $\mathcal{X}$. Let $N = \sum_{i=1}^{B} n_i$, $X^N = (X_1^{n_1}, \ldots, X_M^{n_B})$, and

$$H_0 \colon X^N \sim P^N \text{ for some } P \in \mathcal{P}(\mathcal{X})$$
$$H_1 \colon X_i^{n_i} \sim P_i^{n_i}, i \in [B] \text{ for some } P_{[B]} \in \mathcal{P}^B(\mathcal{X})$$

Lemma

Consider X^N and the hypotheses H_0 and H_1. Let $w_i = \frac{n_i}{N}$ for $i \in [B]$. Denote $\hat{P}_{[B]} = (\hat{P}_{X_i^{n_i}})_{i\in[B]}$. Then,

$$G(\hat{P}_{[B]}, w_{[B]}) = \frac{1}{N} \log \frac{\max\limits_{P_{[B]}\in\mathcal{P}^B(\mathcal{X})} \prod_{i=1}^{B} P_i^{n_i}(X_i^{n_i})}{\max\limits_{P\in\mathcal{P}(\mathcal{X})} P^N(X^N)}. \tag{1}$$

Lower (Converse) Bound

Theorem

For any δ-correct algorithm π with $\delta \in (0,1)$ and any problem instance $P \in \Lambda$,

$$\mathbb{E}[\tau] \geq \frac{d(\delta \| 1-\delta)}{T^*(P)} \geq \frac{1}{T^*(P)} \log \frac{1}{2.4\delta} = \frac{1}{T^*(P)} \log \frac{1}{\delta} + \Theta(1) \tag{2}$$

where $T^*(P) = T(P, \sigma_P)$ *and* $d(p\|q) = p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}$.

Proof: Apply the standard technique from [Garivier and Kaufmann, 2016]. It uses change of measure and Wald's identity.

Algorithm (TaS-FW)

- We want to design a computationally-efficient Track-and-Stop (TaS) algorithm.
- Ideally, a TaS algorithm computes

$$\begin{aligned} w^*(t) &= \arg\max_{w \in \Sigma_K} G_{\hat{P}(t-1)}^{\hat{\sigma}(t-1)}(w) \\ &= \arg\max_{w \in \Sigma_K} \min_{\sigma' \in \mathcal{C} \setminus \{\hat{\sigma}(t-1)\}} g_{\hat{P}(t-1)}^{\sigma'}(w) \end{aligned}$$

  at each time t and matches its fraction of arm pulls to the oracle $w^*(t)$.
- But calculating this maximum can be very difficult in the general case (e.g., $|\mathcal{X}| \geq 3$).
- Hence, we linearize the objective function and utilize a modified version of the Frank–Wolfe algorithm (that is tailored to the non-smoothness coming from the minimum of functions).
- The algorithm is inspired by [Wang et al., 2022], which is for general pure exploration problems but does not apply to ours.

Frank–Wolfe Update

Instead of

$$w^*(t) = \arg\max_{w \in \Sigma_K} \min_{\sigma' \in \mathcal{C} \setminus \{\hat{\sigma}(t-1)\}} g_{\hat{P}(t-1)}^{\sigma'}(w)$$

we solve

$$z(t) = \arg\max_{z \in \Sigma_K} \min_{h \in H_{G_{\hat{P}(t-1)}^{\hat{\sigma}(t-1)}}(x(t-1), r_t)} \langle z - x(t-1), h \rangle \tag{3}$$

$$x(t) = \left(1 - \frac{1}{t}\right) x(t-1) + \frac{1}{t} z(t) = \frac{1}{t} \sum_{s=1}^{t} z(s) \tag{4}$$

where the r-subdifferential subspace $H_{G_P^\sigma}(w, r)$

$$H_{G_P^\sigma}(w, r) \triangleq \mathrm{co}(\nabla g_P^{\sigma'}(w) \colon \sigma' \neq \sigma, g_P^{\sigma'}(w) < G_P^\sigma(w) + r)$$

accounts for the non-smoothness in the objective function.
The maximin in (3) is an LP.

Algorithm 1 TaS-FW

Input: Target error probability $\delta \in (0,1)$, the collection of hypotheses $\mathcal{C}$

Initialization: Sample each arm $i \in [K]$ once, initialize $\tilde{x}(K) = \frac{1}{K}\mathbf{1}$ and $N(K) = (1, \ldots, 1)$.

For $t \in \mathbb{N}$, set $r_t = t^{-4/5}$.

$t \leftarrow K$

while $Z(t) \triangleq t\, G_{\hat{P}(t)}^{\hat{\sigma}(t)}(w(t)) < \beta(t,\delta) \triangleq \beta(t,\delta) = \log\frac{1}{\delta} + (M|\mathcal{X}| + \tilde{K} + 2)\log(t+1) + \log\left(\frac{\pi^2}{6} - 1\right)$ **do**

  $t \leftarrow t + 1$

  if $t \in \mathcal{I}_{\mathrm{f}} \triangleq \{t \in \mathbb{N} \colon \lceil\sqrt{t}\log t\rceil = \lceil\sqrt{t+1}\log(t+1)\rceil - 1\}$ **then**

    $\tilde{z}(t) \leftarrow \frac{1}{K}\mathbf{1}$ (Forced Exploration)

  else if $t \notin \mathcal{I}_{\mathrm{f}}$ **then**

    $\tilde{z}(t) \leftarrow \underset{z \in \Sigma_K}{\arg\max} \min_{h \in H_{G^{\hat{\sigma}(t-1)}_{\hat{P}(t-1)}}(\tilde{x}(t-1), r_t)} \langle z - \tilde{x}(t-1), h\rangle$ (FW Update)

  end if

  $\tilde{x}(t) \leftarrow \left(1 - \frac{1}{t}\right)\tilde{x}(t-1) + \frac{1}{t}\tilde{z}(t)$

  Sample the arm $A_t \leftarrow \underset{i \in [K]}{\arg\max}\,(t\tilde{x}_i(t) - N_i(t-1))$ (C-tracking rule)

  Update $N(t) \leftarrow N(t-1) + e_{A_t}$ and the empirical problem instance $\hat{P}(t)$ in (42)

  $\hat{\sigma}(t) \leftarrow \underset{\sigma \in \mathcal{C}}{\arg\min}\, g^{\sigma}_{\hat{P}(t)}(w(t))$

end while

Output: $\hat{\sigma}(t)$

A Second-order Achievability Bound

Theorem

For any problem instance $P \in \Lambda$, as $\delta \to 0^+$, our algorithm TaS-FW is δ-correct and achieves

$$\mathbb{E}[\tau] \leq \frac{\log \frac{1}{\delta}}{T^*(P)} \left(1 + O\left(\left(\log \frac{1}{\delta}\right)^{-1/4} \sqrt{\log \log \frac{1}{\delta}}\right)\right). \tag{5}$$

- The algorithm is first-order optimal as $\delta \to 0+$.
- We characterize an upper bound on the rate of convergence.
- The tightness of the second-order term is an interesting open problem.

Comparison with Existing Work

Context	**Compared paper**	**Our algorithm**	**Compared algorithm**	**Reason**
Algorithm based on	Wang et al.	Frank–Wolfe	Frank–Wolfe	computing sup-inf efficiently
Forced exploration	Wang et al., GK, Prabhu et al.	$\sqrt{t} \log t$	$\sqrt{t}$	$\dim(\Lambda) < \dim(\mathcal{P}^K(\mathcal{X}))$
Efficiency	Prabhu et al.	Yes	No in general	Prabhu et al. do not provide an efficient algorithm to compute sup-inf
Approach	Yang et al.	FW + Seq. HT	K-means + simplification in the inner infimum	K-means doesn't work for finite alphabets + simplification isn't always possible
Second-order term	Wang et al., Prabhu et al., Yang et al., GK	Includes	Does not include	Refined analysis

Proof Steps

- *C-tracking lemma:*

$$\left| N_i(t) - \sum_{s=1}^{t} \tilde{z}_i(s) \right| \leq K - 1 \tag{6}$$

- Establish the Lipschitzness of $g_P^\sigma(w)$ and $G_P^\sigma(w)$ in w and P.
- *FW lemma:* Let $\tilde{\Delta}_t \triangleq G_P(w^*) - G_P(w(t))$ be the optimality gap. Under the event

$$\max_{z \in \Sigma_K} \min_{h \in H_{G_P}(\tilde{x}(t-1), r_t)} \langle z - \tilde{x}(t-1), h \rangle - \epsilon_t < \min_{h \in H_{G_P}(\tilde{x}(t-1), r_t)} \langle \tilde{z}(t) - \tilde{x}(t-1), h \rangle \tag{7}$$

for $t \in \{T_1, \ldots, T_2\} \cap \mathcal{I}_{\mathrm{f}}^{\mathrm{c}}$,

$$\tilde{\Delta}_{T_2} \leq \frac{T_1}{T_2} L + 2L T_2^{-1/2} \log T_2 + \frac{1}{T_2} \sum_{t=1}^{T_2} (r_t + \epsilon_t) + 32DKT_2^{-1/2} + \frac{L(K+3)}{T_2}. \tag{8}$$

Technical Novelties

- *Concentration of $G(\cdot)$:*

$$\mathbb{P}\left[N\sum_{m=1}^{M} G(\hat{P}_{\mathcal{A}_m}, w_{\mathcal{A}_m}) \geq \beta\right] \leq (N+1)^{M|\mathcal{X}|}\exp\{-\beta\} \tag{9}$$

  which is used to prove $\mathbb{P}\left[\hat{\sigma}(\tau) \neq \sigma_P\right] \leq \delta$.

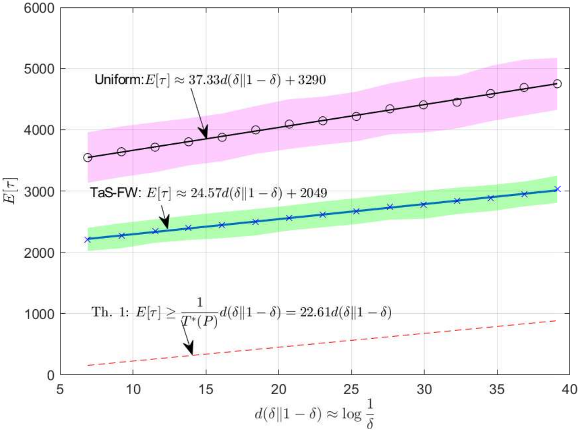
- Establishing a sufficient condition for ϵ_t such that

$$\|\hat{P}(t) - P\|_\infty \leq \epsilon_t$$
$$\hat{\sigma}(t) = \sigma_P \iff \min_{\sigma' \neq \sigma_P} g^{\sigma'}_{\hat{P}(t)}(w(t)) > g^{\sigma_P}_{\hat{P}(t)}(w(t))$$

  Because the condition $\hat{\sigma}(t) = \sigma_P$ depends on $w(t)$, we choose a more aggressive forced exploration ($\sqrt{t}\log t$ times instead of $\sqrt{t}$ times)

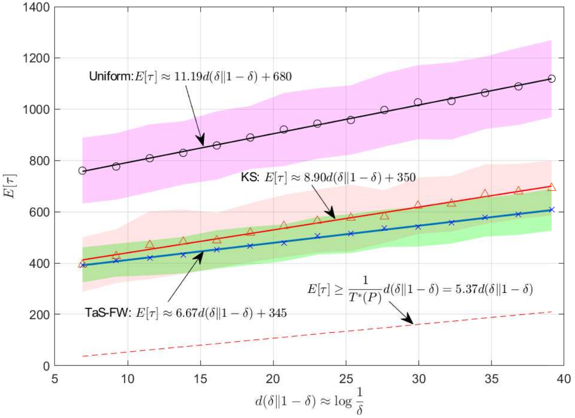
- Carefully choose $\epsilon_t = \Theta(t^{-1/4}\sqrt{\log t})$, T_1, and T_2 to optimize the second-order term in the upper bound.

Experiment 1



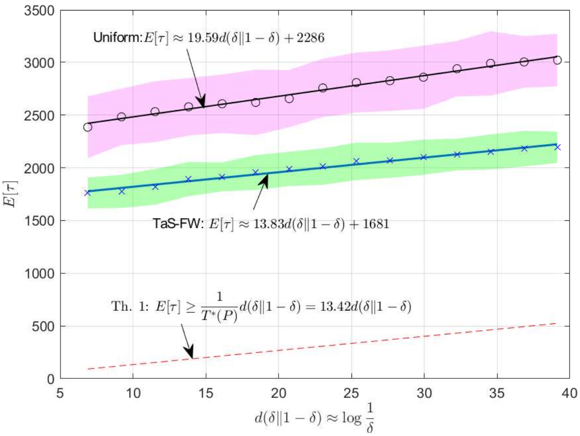
Matching pairs: $K = 6$, $M = 2$ with $P_1 = P_3 = (0.1, 0.1, 0.8)$, $P_2 = P_4 = (0.4, 0.4, 0.2)$, $P_5 = (0.5, 0.05, 0.45)$, and $P_6 = (0.1, 0.8, 0.1)$. True hypothesis is $\sigma_P = \{\{1, 3\}, \{2, 4\}\}$.

Experiment 2



Odd arm identification: $K = 7$, $M = 1$ with $P_i = (0.1, 0.1, 0.8)$ for $i \in [6]$, and $P_7 = (0.6, 0.2, 0.2)$. True hypothesis is $\sigma_P = \{\{1, \ldots, 6\}\}$.

Experiment 3



***M*-ary clustering of *K* arms:** $K = 6$, $M = 3$ with $P = (P_1, \dots, P_7)$, where $P_1 = P_2 = (0.6, 0.2, 0.2)$, $P_3 = P_4 = (0.25, 0.7, 0.05)$, and $P_5 = P_6 = (0.05, 0.05, 0.90)$.
True hypothesis is $\sigma_P = \{\{1, 2\}, \{3, 4\}, \{5, 6\}\}$.

Takeaways

1. We develop a generalized framework for PE problems that involve clustering. Our algorithm works for problems such as odd arm id. in [Karthik] and M-ary clustering in [Yang et al.] simultaneously.
2. Our refined algorithm is first-order optimal as δ approaches 0, and the achievability bound includes a second-order term, which is derived by optimizing the design parameters.
3. We consider distributions on a finite alphabet size, which are a special case of a vector exponential family. Using existing tools in the literature, the result can be extended to one-parameter exponential families.

Thank you for listening to me!