Clustering with bandit feedback: breaking down the computation/information gap

Victor Thuot[1], Alexandra Carpentier[2], Christophe Giraud[3], Nicolas Verzelen[1]

[1]INRAE, Mistea, Institut Agro, Univ Montpellier, France. [2]Institut für Mathematik, Universität Potsdam, Germany. [3]Université Paris-Saclay, LMO, France.

Clustering with Bandit feedback Problem (CBP)

Bandit learning protocol

Consider a multi-armed bandit with N arms. Each arm $a \in \{1,\dots,N\}$ is associated with a multidimensional mean-vector $\mu_a \in \mathbb{R}^d$ (with d possibly large).

For each time step $t = 1,\dots,T$,

- chooses an arm $A_t \in \{1,\dots,N\}$ (*based on the passed observations*)
- receives X_t with mean μ_{A_t} and σ-subGaussian noise (e.g., $X_t \sim \mathcal{N}(\mu_a, \sigma^2 I_d)$)

Hidden partition assumption

We assume that there exists a hidden partition G^* of $[N]$ into exactly K non-empty groups, such that **all arms in the group G_k^* share the same mean-vector** $\Lambda(k)$.

Objective: clustering in the PAC-setting

Given a prescribed probability $\delta \in (0,1)$, the objective of the learner is to **recover exactly the unknown partition of the arms**. She collects observation until some time T, at which she is confidence enough to construct a partition $\hat{G}$ equal to G^* with high probability (*up to permutation of the groups*).

An algorithm $\mathcal{A}$ is δ-PAC if for any environment ν, $\mathbb{P}_{\mathcal{A},\nu}(\hat{G} \sim G^*$ *up to permutation* $) \geqslant 1-\delta$.

Objective: minimizing the budget spent

The performance of a δ-PAC algorithm is mesaured by its budget T (by $\mathbb{E}[T]$ or $\|T\|_\infty$) – as the number of samples collected to construct $\hat{G}$.

For an environment ν, we define two quantities, the minimal gap $\Delta_*(\nu) = \min_{k\neq k'} \|\Lambda(k) - \Lambda(k')\|$, and the balancedness $\theta_*(\nu) = \min_k \frac{|G_k^*|}{N}$. We denote as $\mathcal{E}(\Delta,\theta)$ as the family of environment such that $\Delta_* \geqslant \Delta$ and $\theta_* \geqslant \theta$.

Our main contribution is in showing that the complexity of the problem is characterizing by the following quantity:

$$T^* = N + \frac{\sigma^2}{\Delta^2} N \log\left(\frac{N}{\delta}\right) + \frac{\sigma^2}{\Delta^2}\sqrt{dKN\log\left(\frac{N}{\delta}\right)} .$$

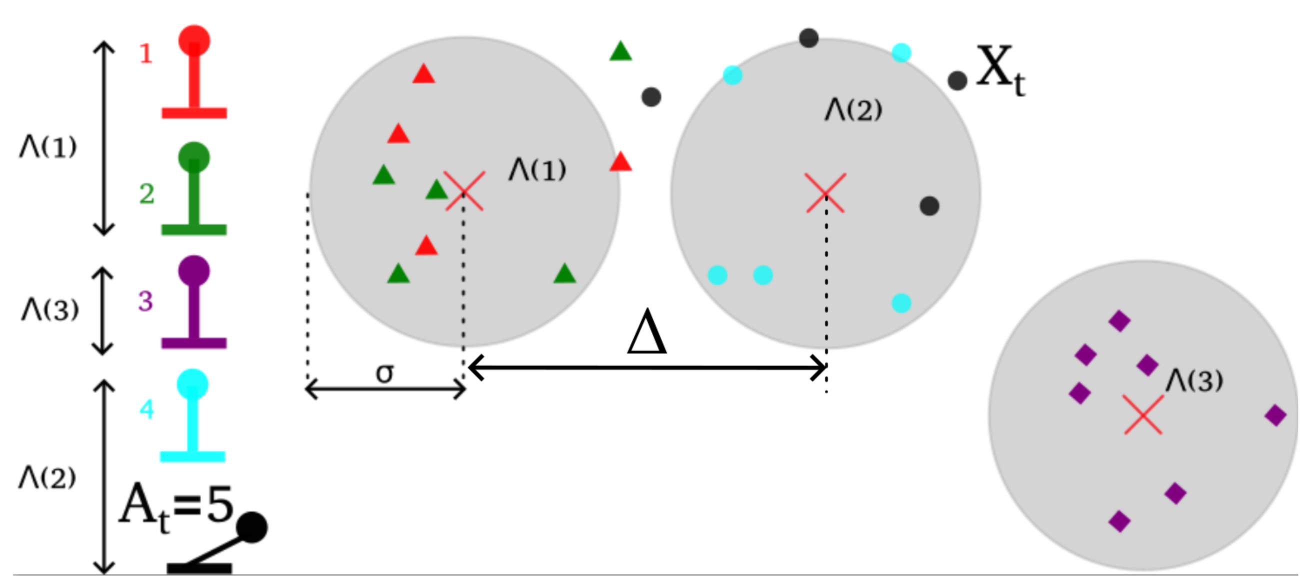


Figure 1. In this illustration, $N=5$, $K=3$, $d=2$, $\Delta_* = \|\Lambda(1) - \Lambda(3)\|$ and $\theta_* = 1/5$. Based on $X_1,\dots,X_{t-1}$, the algorithm chooses $A_t = 5$ and observes X_t centred on $\mu_5 = \Lambda(2)$.

Algorithms

Algorithm 1: Sequential Representative Identification (SRI)

```
Input: δ, Δ, θ
Result: S a set of arms
Pick randomly a_0 ∈ [N] ;
Set S = {a_0}
μ̂_{a_0}, μ̂'_{a_0} ← empirical_mean(a_0, n_max) ;
  /* Estimate μ_{a_0} */
for u = 1,...,U do
    Sample uniformly at random a_u ∈ [N]
    for s = s_0,...,r do
        μ̂_{a_u}, μ̂'_{a_u} ← empirical_mean(a_0, n_0 2^s) ;
          /* Estimate μ_{a_u} */
        if min_{b∈S} ⟨μ̂_a − μ̂_b, μ̂'_a − μ̂'_b⟩ ≤ Δ²/2 then
            Break ;                        /* reject a_u */
        if s = r then          /* if a_u passed all tests */
            S ← S ∪ {a_u}     /* Add a_u to S */
            μ̂_{a_u}, μ̂'_{a_u} ← empirical_mean(a_u, n_max)
              /* Estimate μ_{a_u} */
    if |S| = K or budget > T_max then
        Break               /* Terminate u loop */
return S                   /* Return a set of representatives */
```

Algorithm 2: Active Distance-based Classifier (ADC)

```
Input: δ, Δ and S = {b_1,...,b_K}
Result: Ĝ a partition of the arms
Compute I = T*/N and J = T*/K
for j ∈ [K] do
    μ̂(j), μ̂'(j) ← empirical_mean(b_j, J) ;
      /* Estimate the centers */
    ĝ(b_j) ← j
for a ∈ [n] \ S do
    μ̂_a, μ̂'_a ← empirical_mean(a, I) ;
    Compute
    ĝ(a) ∈ argmin_{j=1,...,K} ⟨μ̂_a − μ̂(j), μ̂'_a − μ̂'(j)⟩
    ;                          /* Classify arm a */
return {ĝ(1),...,ĝ(n)}        /* Return a clustering */
```

Algorithm 3: Active Clustering Bandits (ACB)

```
Input: δ, Δ, θ
Ŝ ← SRI(δ/2, Δ, θ) ;                  /* Alg 1 */
return Ĝ = ADC(δ/2, Δ, Ŝ) ;           /* Alg 2 */
```

Lower bound

We derive a lower bound, combining methods from information theory and high-dimensionnal statistics:

Theorem 1

For any algorithm $\mathcal{A}$, any $\Delta > 0$, $\theta > 2/N$, there exists an environment $\nu \in \mathcal{E}(\Delta,\theta)$, such that

$$\mathbb{E}_{\mathcal{A},\nu}[T] \geqslant cT^* .$$

Upper bound

We introduce ACB, an algorithm which works as a two step procedure (describe in the above column in pseudocode):

1. (SRI): identifying S, a set of arms with exactly one arm from each cluster
2. (ADC): estimate the common means of the clusters and classify the arms with a distance-based classifier,

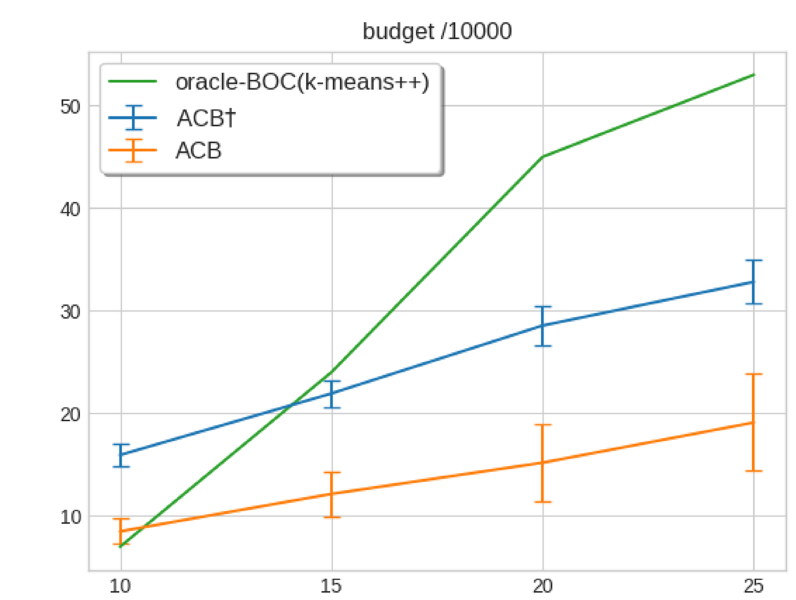
Contributions

We answer the following questions:

1. Can we improve the budget of a simple uniform sampling strategy ?
   Yes, we provide the ACB Algorithm, a polynomial-time algorithm which outperforms the uniform sampling strategy.
2. Can we achieve optimality ?
   Yes, ACB is δ-PAC, and we bound its budget, which matches the lower bound T^* in most regimes (for θ not too small, e.g., with balanced groups).
3. Is there an information-computation gap for ACP?
   No, there is no computational gap (contrary to the batch setting), ACB is optimal and computationnaly efficient.

Numerical experiments

Figure 2. Comparison of the necessary budget for ACB and oracle-BOC with varying number of clusters. In blue (resp. orange) the (empirical) budget of ACB† (resp. ACB) computed with 100 simulations. Algorithm ACB knows Δ, θ, while ACB does not know Δ (we use a doubling trick). In green, the smallest budget for which oracle-BOC (uniform sampling followed by kmeans++) makes less than 10% of error out of 100 experiments.



References

[1] V. Thuot, A. Carpentier, C. Giraud, and N. Verzelen. Clustering with bandit feedback: breaking down the computation/information gap. In G. Kamath and P.-L. Loh, editors, *Proceedings of The 36th International Conference on Algorithmic Learning Theory*, volume 272 of *Proceedings of Machine Learning Research*, pages 1221–1284. PMLR, 24–27 Feb 2025.

Contact: victor.thuot@inrae.fr