Disagreement-Based Combinatorial Pure Exploration:

Sample Complexity Bounds and an Efficient Algorithm

UMassAmherst

Tongyi Cao and Akshay Krishnamurthy

Microsoft Research

Combinatorial Pure Exploration

- Stochastic multi-armed bandits: Arms $a \in [K]$, each with sub-Gaussian distribution ν_a with unknown mean $\mu_a \in [-1, 1]$. (Vectorized representation $\mu \in [-1, 1]^K$.)
- Combinatorial decision set: $\mathcal{V} \subseteq \{0, 1\}^K$.
- Pure exploration problem: find

$$v^\star \triangleq \operatorname{argmax}_{v \in \mathcal{V}} \langle v, \mu \rangle,$$

while minimizing samples. Query individual arms, sequentially.
  - Fixed confidence: Given $\delta \in (0, 1)$ ensure $\mathbb{P}[\hat{v} \neq v^\star] \leq \delta$, minimize samples.
  - Fixed budget: Given $T \in \mathbb{N}$ use at most T samples, minimize $\mathbb{P}[\hat{v} \neq v^\star]$.
- Optimization oracle-based computational model

$$\mathrm{Oracle}(c) \triangleq \operatorname{argmax}_{v \in \mathcal{V}} \langle v, c \rangle.$$

Non-interactive Baseline

Algorithm: Query each arm T/K times and output $\hat{v} = \operatorname{argmax}_v \langle v, \hat{\mu} \rangle$ with empirical mean $\hat{\mu}$.
Combinatorial Parameters: with $d(v, v^\star) \triangleq |v \ominus v^\star|, \mathcal{B}(k, v) \triangleq \{u \in \mathcal{V} \mid d(v, u) = k\}$,

$$\Phi \triangleq \Phi(\mathcal{V}) \triangleq \max_{k \in \mathbb{N}, v \in \mathcal{V}} \frac{\log(|\mathcal{B}(k, v)|)}{k}, \qquad \Psi \triangleq \Psi(\mathcal{V}) \triangleq \min_{u, v \in \mathcal{V}} d(u, v).$$

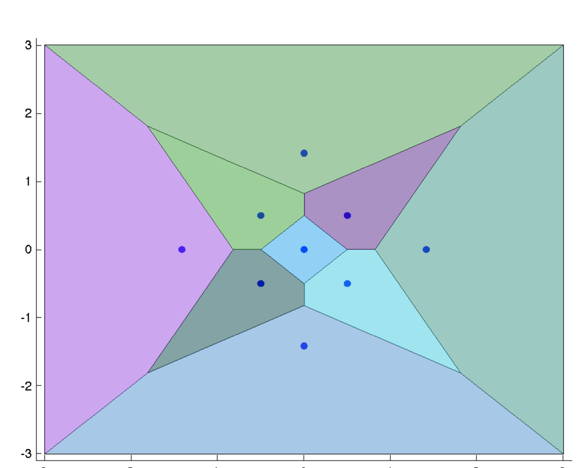
Instance-Specific Parameters (e.g., Gaps):

$$\Delta_v(\mu) \triangleq \frac{\langle v - v^\star(\mu), \mu \rangle}{d(v, v^\star)}, \qquad \Delta_a(\mu) \triangleq \min_{v : a \in v \ominus v^\star} \Delta_v(\mu).$$

Theorem 1. *Algorithm succeeds w.p.* $1 - \delta$ *when* $T \geq O\left(\frac{K}{\min_v \Delta_v^2}\left(\Phi + \frac{\log(K/\delta)}{\Psi}\right)\right)$.
This is minimax optimal among non-interactive methods.

Proof uses a simple concentration argument:

$$\begin{aligned}
\mathbb{P}[\hat{v} \neq v^\star] &= \mathbb{P}\left[\exists v \in \mathcal{V} : \frac{|\langle v^\star - v, \hat{\mu} - \mu \rangle|}{d(v^\star, v)} \geq \epsilon\right] \\
&\leq 2 \sum_{v \in \mathcal{V}} \exp\left(\frac{-T d(v^\star, v) \epsilon^2}{2K}\right) \\
&\leq 2 \sum_{k=\Psi}^{K} |\mathcal{B}(k, v^\star)| \exp\left(\frac{-T k \epsilon^2}{2K}\right) \\
&\leq 2K \exp\left(\max_{\Psi \leq k \leq K} \log |\mathcal{B}(k, v^\star)| - \frac{T k \epsilon^2}{2K}\right).
\end{aligned}$$



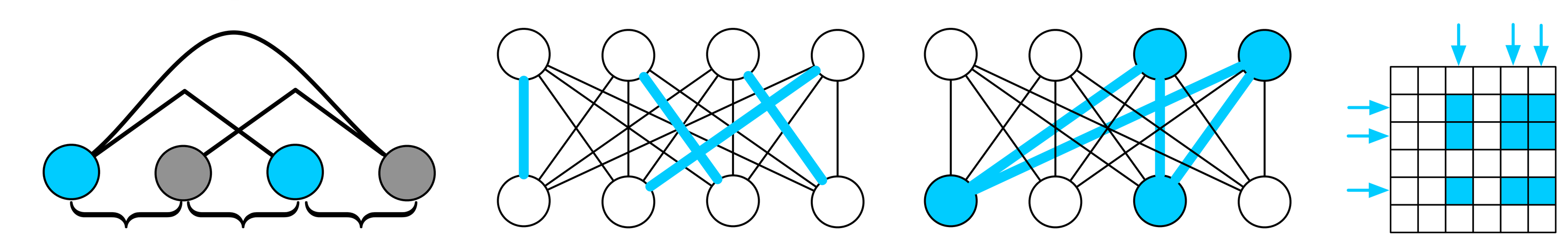
Follows since $\langle v^\star - v, \hat{\mu} - \mu \rangle$ is the average of $\frac{T d(v^\star, v)}{K}$ centered sub-Gaussian random variables.
Method succeeds with $\epsilon = \min_{v \neq v^\star} \Delta_v(\mu)$. Choosing T such that RHS is at most δ yields theorem.

Normalized Regret Inequality: With n samples per arm, for any δ we have

$$\mathbb{P}\left(\exists v \in \mathcal{V} : \frac{|\langle v^\star - v, \hat{\mu} - \mu \rangle|}{d(v^\star, v)} \geq \sqrt{\frac{2}{n}\left(\Phi + \frac{\log(2K/\delta)}{\Psi}\right)}\right) \leq \delta.$$

Examples and Motivation

Top-K | Matching | Biclique | Biclustering



$|\mathcal{V}| = \binom{K}{s}$ | $|\mathcal{V}| = \left(\sqrt{K}\right)!$ | $|\mathcal{V}| = \binom{\sqrt{K}}{\sqrt{s}}^2$ | $|\mathcal{V}| = \binom{\sqrt{K}}{\sqrt{s}}^2$

- Also: Disjoint Sets, partition $[K]$ into K/s blocks, choose one element per block.
- Many well-studied examples (Top-K, Matroids, Biclustering). Sharp guarantees known for matroids.

Comparisons: Compare leading terms in *homogeneous* setting: $\mu = \Delta(2v^\star - \mathbf{1})$.

Sample complexity	Top-k	DisjSet	Matching	Biclique
[CLKLC14] / Baseline	$\Theta(1)$	$\Theta(s)$	$\Omega(K)$	$\Omega(\sqrt{s})$
[CGLQW17] / Baseline	$\Theta(1)$	$\Theta(1)$	$\Omega(K^{1/2})$	$\Omega(1)$
[GLGOB16] / Baseline	$\Theta(1)$	$\Theta(s)$	$\Omega(1)$	$\Omega(\sqrt{s})$

Interactive algorithms can be polynomially *worse* than non-interactive baseline!

Why? Normalized regret inequality much sharper than other natural concentration arguments (e.g., uniform convergence on all arms, all sets, or all pairs of sets).
But regret inequality hard to use algorithmically!

- How should we collect data to do unsupervised learning or structure discovery?
- Can we design an algorithm that is never worse than baseline and sometimes much better?
- Can we make the algorithm oracle efficient?

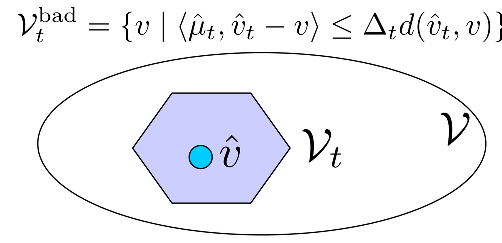
Disagreement-based Algorithm

Intuition: Use disagreement-based active learning

1. Maintain version space of "good" sets.
2. Query where version space disagrees.

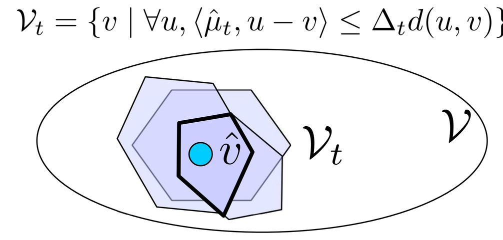
In active learning, standard version space is:

$$\mathcal{V}_t^{\text{bad}} = \{v \mid \langle \hat{\mu}_t, \hat{v}_t - v \rangle \leq \Delta_t d(\hat{v}_t, v)\}$$



For us, much better version space:

$$\mathcal{V}_t = \{v \mid \forall u, \langle \hat{\mu}_t, u - v \rangle \leq \Delta_t d(u, v)\}$$



Algorithm 1

1: **for** $t = 1, 2 \ldots,$ **do**
2: Compute $\hat{v}_t = \operatorname{argmax}_{v \in \mathcal{V}} \langle v, \hat{\mu}_t \rangle$
3: **for** $a \in [K]$ **do**
4: Query a if $\mathcal{V}_t$ disagrees

$$\exists v \in \mathcal{V}_t(\hat{\mu}_t, \Delta_t) \text{ s.t. } v(a) \neq \hat{v}_t(a)$$

5: Otherwise, hallucinate $y_t(a) = 2(\hat{v}_t(a) - 1)$
6: Update $\hat{\mu}_{t+1} \leftarrow \frac{1}{t+1} \sum_{i=0}^{t} y_i$
7: If no queries issued this round, output $\hat{v}_t$

Theorem 2. *For any* $\delta \in (0, 1)$, *Algorithm guarantees that* $\mathbb{P}[\hat{v} \neq v^\star] \leq \delta$, *with sample complexity*

$$\sum_{a \in K} \frac{144}{\Delta_a^2}\left(\Phi + \frac{2\log(144/(\Delta_a^2 \Psi)) + 2\log(K\pi^2/\delta)}{\Psi}\right).$$

- Modulo logarithmic factors, never worse than non-interactive algorithms. Better with heterogeneity.

Efficient Computation

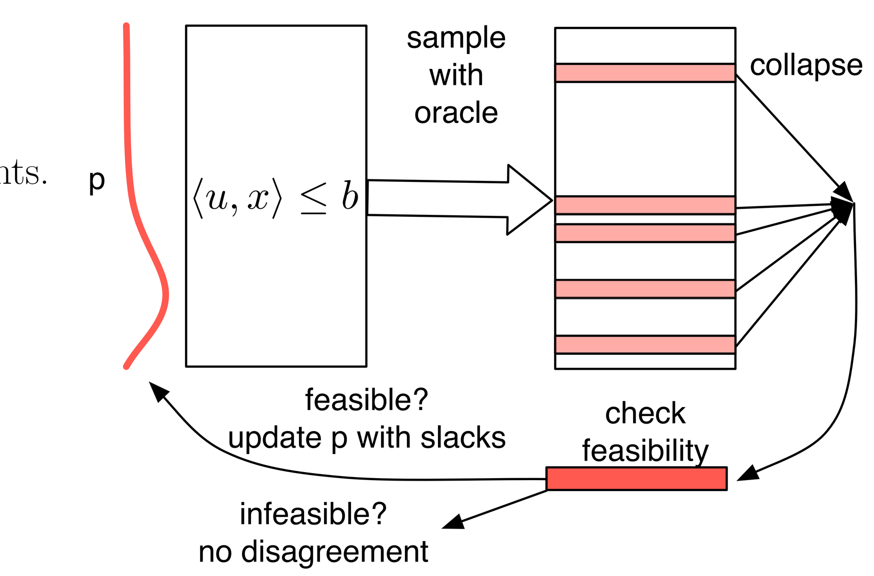
Bottleneck is computing disagreement:

$$\exists v \in \mathcal{V}_t \text{ s.t. } v(a) \neq \hat{v}_t(a).$$

- Linear feasibility, exponentially many constraints.
- Use online learner to collapse constraints

$$\max_v \sum_u p_t(u) \left(\Delta d(v, u) - \langle \hat{\mu}, u - v \rangle\right) \quad \text{s.t. } v(a) = b$$

- Update learner using slack of best-response.
- Use FTPL for implicit distribution.



Theorem 3. *FTPL runs in polynomial time with* $\tilde{O}(K^6/\Delta^4)$ *oracle calls. If it reports* FALSE *then there is no disagreement. Otherwise there exists* $v \in conv(\mathcal{V})$ *with* $v(a) = b$ *and* $\forall u \in \mathcal{V}$ $\langle \hat{\mu}, u - v \rangle \leq \Delta \|u - v\|_1 + \Delta$ *(There is approximate disagreement).*

- Approximate feasibility does not damage sample complexity.
- **Corollary**: Fixed confidence algorithm runs in polynomial time with optimization oracle.

Other results

Define symmetrized log-volume $D(v, v') \triangleq \max\{\log |\mathcal{B}(d(v, v'), v)|, \log |\mathcal{B}(d(v, v'), v')|\}$.

Theorem 4 (Refined fixed confidence)**.** *There exists a computationally inefficient fixed confidence algorithm with sample complexity*

$$O\left(\sum_{a \in [K]} H_a^{(1)}\left(\log(H_a^{(1)}) + \log(\pi^2 K/\delta)\right) + H_a^{(2)}\right),$$

where $H_a^{(1)} \triangleq \max_{v : a \in v \ominus v^\star} \frac{d(v, v^\star)}{\langle \mu, v^\star - v \rangle^2}$ *and* $H_a^{(2)} \triangleq \max_{v : a \in v \ominus v^\star} \frac{d(v, v^\star) D(v, v^\star)}{\langle \mu, v^\star - v \rangle^2}$.
Better dependence on combinatorial parameters Φ, Ψ, *since* $H_a^{(1)} \leq \frac{1}{\Delta_a^2 \Psi}$ *and* $H_a^{(2)} \leq \frac{\Phi}{\Delta_a^2}$.

Theorem 5 (Fixed budget)**.** *Given budget* $T \geq K$ *there exists an algorithm guaranteeing*

$$\mathbb{P}[\hat{v} \neq v^\star] \leq K^2 \exp\left(\psi\left(\Phi - \frac{T - K}{8 \log(K) \sum_a \Delta_a^{-2}}\right)\right).$$

Final Remark: In the high confidence regime ($\delta = \exp(-K)$), [CGLQW17] give tight instance-optimal results. But for $\delta = \mathrm{poly}(1/K)$ their algorithm can be significantly worse than non-interactive baseline and our algorithm. This "moderate confidence" regime is quite interesting and the instance optimal rates here remain unknown.

References

1. Chen, Gupta, Li, Qiao, and Wang. Nearly optimal sampling algorithms for combinatorial pure exploration. COLT 2017.
2. Chen, Lin, King, Lyu, and Chen. Combinatorial pure explo- ration of multi-armed bandits. NeurIPS 2014.
3. Gabillon, Lazaric, Ghavamzadeh, Ortner, and Bartlett. Improved learning complexity in combinatorial pure exploration bandits. AISTATS 2016.

Learn more at: `https://arxiv.org/abs/1711.08018`