JMLR: Workshop and Conference Proceedings vol 49:1–4, 2016

Open Problem: Best Arm Identification: Almost Instance-Wise Optimality and the Gap Entropy Conjecture

Lijie Chen* CHENLJ13@MAILS.TSINGHUA.EDU.CN
Institute for Interdisciplinary Information Sciences (IIIS), Tsinghua University, China.

Jian Li* LIJIAN83@MAIL.TSINGHUA.EDU.CN
Institute for Interdisciplinary Information Sciences (IIIS), Tsinghua University, China.

Abstract

The best arm identification problem (BEST-1-ARM) is the most basic pure exploration problem in stochastic multi-armed bandits. The problem has a long history and attracted significant attention for the last decade. However, we do not yet have a complete understanding of the optimal sample complexity of the problem: The state-of-the-art algorithms achieve a sample complexity of $O(\sum_{i=2}^{n}\Delta_i^{-2}(\ln\delta^{-1}+\ln\ln\Delta_i^{-1}))$ (Δ_i is the difference between the largest mean and the i^{th} mean), while the best known lower bound is $\Omega(\sum_{i=2}^{n}\Delta_i^{-2}\ln\delta^{-1})$ for general instances and $\Omega(\Delta^{-2}\ln\ln\Delta^{-1})$ for the two-arm instances. We propose to study the instance-wise optimality for the BEST-1-ARM problem. Previous work has proved that it is impossible to have an instance optimal algorithm for the 2-arm problem. However, we conjecture that modulo the additive term $\Omega(\Delta_2^{-2}\ln\ln\Delta_2^{-1})$ (which is an upper bound and worst case lower bound for the 2-arm problem), there is an instance optimal algorithm for BEST-1-ARM. Moreover, we introduce a new quantity, called the gap entropy for a best-arm problem instance, and conjecture that it is the instance-wise lower bound. Hence, resolving this conjecture would provide a final answer to the old and basic problem.

1. Introduction

In the BEST-1-ARM problem, we are given n stochastic arms $A_1,\dots,A_n$. The i^{th} arm A_i has a reward distribution $\mathcal{D}_i$ with an unknown mean $\mu_i\in[0,1]$. We assume that all reward distributions are Gaussian distributions with variance 1. Upon each play of A_i, we can get a reward value sampled i.i.d. from $\mathcal{D}_i$. Our goal is to identify the arm with largest mean using as few samples as possible. We assume here that the largest mean is strictly larger than the second largest (i.e., $\mu_{[1]}>\mu_{[2]}$) to ensure the uniqueness of the solution, where $\mu_{[i]}$ denotes the i^{th} largest mean. The problem is also called the *pure exploration problem* in the stochastic multi-armed bandit literature.

We say an algorithm $\mathbb{A}$ is δ-correct for BEST-1-ARM, if it outputs the correct answer on *any* instance with probability at $1-\delta$, and we use $T_{\mathbb{A}}(I)$ to denote the expected number of total samples taken by algorithm $\mathbb{A}$ on instance I. We also define the gap of i^{th} arm, $\Delta_{[i]}:=\mu_{[1]}-\mu_{[i]}$.

2. Background

During the last decade, the BEST-1-ARM problem and its optimal sample complexity have attracted significant attention. We only mention a small subset that are most relevant to us. The current best lower bound is due to Mannor and Tsitsiklis (2004), who showed that for any δ-correct algorithm for BEST-1-ARM, it requires $\Omega\left(\sum_{i=2}^{n}\Delta_{[i]}^{-2}\ln\delta^{-1}\right)$ (referred to as the MT lower bound from now

* Institute for Interdisciplinary Information Sciences (IIIS), Tsinghua University, Beijing, China. Research supported in part by the National Basic Research Program of China grants 2015CB358700, 2011CBA00300, 2011CBA00301, and the National NSFC grants 61033001, 61361136003.

on) samples in expectation for any instance. We note that the MT lower bound is an *instance-wise lower bound*, i.e., any BEST-1-ARM instance requires the stated number of samples. On the other hand, the current published best known upper bound is $O\left(\sum_{i=2}^{n}\Delta_{[i]}^{-2}\left(\ln\ln\Delta_{[i]}^{-1}+\ln\delta^{-1}\right)\right)$, due to Karnin et al. (2013). Jamieson et al. (2014) obtained a UCB-type algorithm (called lil'UCB), which achieves the same sample complexity. We refer the above bound as the KKS-JMNS bound. Back in 1964, Farrell (1964) provided an $\Omega(\Delta^{-2}\ln\ln\Delta_2^{-1})$ lower bound for the two-arm cases (which matches the KKS-JMNS bound for two arms).

Very recently, in an unpublished manuscript (Chen and Li (2015)), the authors obtained improved lower and upper bounds for BEST-1-ARM. The work lead the authors to make an intriguing conjecture which we detail in the next section. We will also state the improved bounds and their connection to the conjecture in more details.

3. Open Problem: Almost Instance Optimality and the Gap Entropy Conjecture

We propose to study BEST-1-ARM from the perspective of instance optimality, the ultimate notion of optimality (see e.g., Fagin et al. (2003); Afshani et al. (2009)).

For the 2-arm cases, the KKS-JMNS bound $O(\Delta^{-2}\ln\ln\Delta_2^{-1})$ is an upper bound for every instance, and the Farrell lower bound $\Omega(\Delta^{-2}\ln\ln\Delta_2^{-1})$ is a lower bound for the worst case instances. As we observed in (Chen and Li (2015)), it is impossible to obtain an instance optimal algorithm even for the 2-arm cases. While the observation has ruled out any hope of an instance optimal algorithm for BEST-1-ARM, however, as we will see, it is still possible to obtain very satisfiable answer in terms of instance optimality.

Now, we formally define what is an instance-wise lower bound. Clearly, two arm instances differ only by a permutation of arms should be considered as the same instance. Inspired by Afshani et al. (2009), we give the following natural definition.

Definition 3.1 *(Order-Oblivious Instance-wise Lower Bound)*
Given a BEST-1-ARM *instance I and a confidence level δ, we define*

$$\mathcal{L}(I,\delta) := \inf_{\mathbb{A}:\mathbb{A} \text{ is } \delta\text{-correct for BEST-1-ARM}} \frac{1}{n!}\cdot\sum\nolimits_{\pi\in\mathrm{Sym}(n)} T_{\mathbb{A}}(\pi\circ I),$$

where the summation is over all $n!$ permutations of $\{1,\ldots,n\}$.
The MT lower bound immediately implies that $\mathcal{L}(I,\delta)=\Omega(\sum_{i=2}^{n}\Delta_{[i]}^{-2}\ln\delta^{-1})$.

We conjecture that the two-arm instance is the *only* obstruction toward an instance-wise optimal algorithm. More precisely, we have the following conjecture.

Conjecture 3.2 *There is an algorithm for* BEST-1-ARM *with sample complexity*

$$O(\mathcal{L}(I,\delta)+\Delta_2^{-2}\ln\ln\Delta_2^{-1}),$$

for any instance I and $\delta<0.1$. And we say such an algorithm is almost instance-wise optimal for BEST-1-ARM.

In the light of the discussion for the 2-arm cases, there must be a gap between the sample complexity of a δ-correct algorithm and $\mathcal{L}(I,\delta)$, and Conjecture 3.2 states that the gap can be as small as an *additive factor* $\Delta_2^{-2}\ln\ln\Delta_2^{-1}$, which is all we need to find out the best arm from the top-2 arms, and is an inevitable gap even for the 2-arm instances.

Moreover, we provide an explicit formula for $\mathcal{L}(I,\delta)$. Interestingly, the formula involves an entropy term (similar entropy terms also appear in Afshani et al. (2009) for completely different problems). We define the entropy term first.

Definition 3.3 *Given a* BEST-1-ARM *instance I, let*

$$G_k = \{i \in [2,n] \mid 2^{-k} \le \Delta_{[u]} < 2^{-k+1}\}, \quad H_k = \sum\nolimits_{i \in G_k} \Delta_{[i]}^{-2}, \quad \text{and} \quad p_k = H_k / \sum\nolimits_j H_j.$$

We can view $\{p_k\}$ as a discrete probability distribution. We define the following quantity as the gap entropy *for the instance I*

$$\mathsf{Ent}(I) = \sum\nolimits_{G_k \neq \emptyset} p_k \log p_k^{-1}.$$[1]

Remark 3.4 *We choose to partition the arms based on the powers of 2. There is nothing special about 2 and replacing it by any other constant only changes $\mathsf{Ent}(I)$ by a constant factor.*

Then we formally state our conjecture.

Conjecture 3.5 *For any* BEST-1-ARM *instance I and $\delta < 0.1$, we have*

$$\mathcal{L}(I, \delta) = \Theta\left(\sum\nolimits_{i=2}^{n} \Delta_{[i]}^{-2} \cdot \left(\ln \delta^{-1} + \mathsf{Ent}(I)\right)\right).$$

In the next section, we will try to motivate the term $\mathsf{Ent}(I)$ and explain the reasons that lead us to make the above conjecture.

4. Motivation and Current Progress

In our recent work (Chen and Li (2015)), we provide an algorithm with the following sample complexity:

$$O\Big(\Delta_{[2]}^{-2} \ln\ln \Delta_{[2]}^{-1} + \sum\nolimits_{i=2}^{n} \Delta_{[i]}^{-2} \ln \delta^{-1} + \sum\nolimits_{i=2}^{n} \Delta_{[i]}^{-2} \ln\ln \min(n, \Delta_{[i]}^{-1})\Big). \tag{1}$$

Furthermore, the algorithm achieves a sample complexity of

$$O\left(\Delta_{[2]}^{-2} \ln\ln \Delta_{[2]}^{-1} + \sum\nolimits_{i=2}^{n} \Delta_{[i]}^{-2} \ln \delta^{-1}\right), \tag{2}$$

for clustered instances (We say an instance is clustered if the number of nonempty G_ks is bounded by a constant).

Our new upper bounds (1) and (2) match our conjectured gap entropy lower bound in two extreme cases. On one extreme, the maximum value $\mathsf{Ent}(I)$ can get is $O(\ln\ln n)$. This can be achieved by instances in which there are $\log n$ nonempty groups G_i and they have almost the same weight H_i. Hence, (1) is optimal for such instances. On the other extreme where there is only a constant number of nonempty groups (i.e., the instance is clustered), $\mathsf{Ent}(I) = O(1)$, and our algorithm can achieve almost instance optimality (without relying on the Conjecture 3.5, due to the MT lower bound) in this case.

Besides the fact that our algorithm can achieve optimal results for both extreme cases, we have more reasons to believe why $\mathsf{Ent}(I)$ should enter the picture.

Upper Bounds:

First, we motivate the gap entropy Ent from the algorithmic side. Consider an elimination-based algorithm (such as Karnin et al. (2013) or our algorithm). We must ensure that the best arm is not eliminated in any round. Recall that in the r^{th} round, we want to eliminate arms with gap $\Delta_r = \Theta(2^{-r})$, which is done by obtaining an approximation of the best arm, then take $O(\Delta_r^{-2} \ln \delta_r^{-1})$ samples from each arm and eliminate the arms with smaller empirical means. Roughly speaking, we

1. Note that it is exactly the Shannon entropy for the distribution defined by $\{p_k\}$.

need to assign the failure probability δ_r carefully to each round (by union bound, we need $\sum_r \delta_r \leq \delta$). The algorithm in Karnin et al. (2013) used $\delta_r = O(\delta \cdot r^{-2})$, and we used a better way to assign δ_r. Indeed, if one can assign δ_r's optimally (i.e., minimize $\sum_r H_r \ln \delta_r^{-1}$ subject to $\sum_r \delta_r \leq \delta$), one could achieve the entropy bound $\sum_r H_r \cdot (\ln \delta^{-1} + \mathsf{Ent}(I))$ (by letting $\delta_r = \delta H_r / \sum_i H_i$). Of course, this does not lead to an algorithm directly, as we do not know H_is in advance.

Using our techniques, we can estimate the values H_r's when we enter the r^{th} elimination stage. The only obstacle for implementing the above idea of assigning δ_r's optimally is that we do not know $\sum_r H_r$ initially. We believe the difficulty can be overcome by additional new algorithmic ideas.

Lower Bounds:

In Chen and Li (2015), we prove the following lower bound, improving the MT lower bound.

Theorem 4.1 *(Theorem 1.6 in Chen and Li (2015)) There exist constants $c, c_1 > 0$ and $N \in \mathbb{N}$ such that, for any $\delta < 0.005$ and any δ-correct algorithm $\mathbb{A}$, and any $n \geq N$, there exists an n arms instance I such that $T_{\mathbb{A}}[I] \geq c \cdot \sum_{i=2}^{n} \Delta_{[i]}^{-1} \ln\ln n$. Furthermore, $\Delta_{[2]}^{-2} \ln\ln \Delta_{[2]}^{-1} < \frac{c_1}{\ln n} \cdot \sum_{i=2}^{n} \Delta_{[i]}^{-1} \ln\ln n$.*

In fact, in the lower bound instances, there are $\log n$ nonempty groups G_i and they have almost the same weight H_i (hence, $\mathsf{Ent}(I) = \Theta(\ln\ln n)$). Combining with the MT lower bound, we have covered the two extreme ends of Conjecture 3.5.

Moreover, it is possible to extend our current technique to construct many instances I_S such that any algorithm $\mathbb{A}$ requires at least $\Omega(H(I_S) \cdot \mathsf{Ent}(I_S))$ samples. This strongly suggests $\Omega(H(I) \cdot \mathsf{Ent}(I))$ is the right lower bound. However, a complete resolution of Conjecture 3.5 seems to require new techniques.

References

Peyman Afshani, Jérémy Barbay, and Timothy M Chan. Instance-optimal geometric algorithms. In *Foundations of Computer Science, 2009. FOCS'09. 50th Annual IEEE Symposium on*, pages 129–138. IEEE, 2009.

Lijie Chen and Jian Li. On the optimal sample complexity for best arm identification. *arXiv preprint arXiv:1511.03774*, 2015.

Ronald Fagin, Amnon Lotem, and Moni Naor. Optimal aggregation algorithms for middleware. *Journal of Computer and System Sciences*, 66(4):614–656, 2003.

RH Farrell. Asymptotic behavior of expected sample size in certain one sided tests. *The Annals of Mathematical Statistics*, pages 36–72, 1964.

Kevin Jamieson, Matthew Malloy, Robert Nowak, and Sébastien Bubeck. lil'ucb: An optimal exploration algorithm for multi-armed bandits. *COLT*, 2014.

Zohar Karnin, Tomer Koren, and Oren Somekh. Almost optimal exploration in multi-armed bandits. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, pages 1238–1246, 2013.

Shie Mannor and John N Tsitsiklis. The sample complexity of exploration in the multi-armed bandit problem. *The Journal of Machine Learning Research*, 5:623–648, 2004.