31005 Machine learning
Assignment 3: Examination

Name: Liangzhu Jiang
Student ID: 13064593

1. Introduction

In this assignment, I choose question five which is about gambling machine. The task of me is to design a strategy to make more awards through as few pulls as possible. However, there is a most pivotal problem should be considered and solved which is the results in the gambling machine is random (RNG). In other words, the results generated by each gambling machine is different. In addition, these results (numbers) are created at a rate of several billion hundreds of times per second even if nobody plays these gambling machines. Therefore, it seems prediction is impossible, however, many people don't know the hit frequency and the payout ratio are set and can't be changed, no matter what. Therefore, I will discuss how can people find the arm with the best expected return as early as possible.

2. Strategy to earn reward as fast as possible

In this assignment, our gambling machine which has multiple arms so we call them multi-armed bandit. In the past years, many gamblers think of many ways to cheat on gambling machines however, physical slot machines are computerized nowadays, so these illegal means are useless. By contrast, if people want to find law, it is also impossible because the corn algorithm is Random Number Generator (RNG). I have briefly told the feature of this random mechanism above. Therefore, what can we do is to try to find the arm with the best expected return as early as possible, and then to keep gambling using that arm. (Robbins 1952) In this way, we can probably use minimum number of pulls to get as much return as possible. My strategy is to use algorithms which based on sequential elimination of arms in fixed confidence and fixed budget settings. This strategy refers two algorithms: Exponential-Gap Elimination algorithm and

Sequential Halving algorithm. (Fig 1,2) Before the detail description, I will define two settings first:

Fixed confidence: Give the player a target confidence, and his goal is to pull the arms as little as possible in order to find the best arm with a probability of at least 1.

input confidence $\delta > 0$

initialize $S_1 \leftarrow [n]$, $r \leftarrow 1$

while $|S_r| > 1$ **do**

  let $\varepsilon_r = 2^{-r}/4$ and $\delta_r = \delta/(50r^3)$

  sample each arm $i \in S_r$ for $t_r = (2/\varepsilon_r^2)\ln(2/\delta_r)$ times, and let $\hat{p}_i^r$ be the average reward

  invoke $i_r \leftarrow \textsc{MedianElimination}(S_r, \varepsilon_r/2, \delta_r)$ and let $\hat{p}_\star^r = \hat{p}_{i_r}^r$

  set $S_{r+1} \leftarrow S_r \setminus \{i \in S_r : \hat{p}_i^r < \hat{p}_\star^r - \varepsilon_r\}$

  update $r \leftarrow r + 1$

end while

output arm in S_r

Fig 1: Exponential-Gap Elimination algorithm

Fixed budget: Given that the total budget is pulled by the T arm, the player's goal is to maximize the correct identification of the best arm without pulling the arm more than T times.

input total budget T

initialize $S_0 \leftarrow [n]$

for $r = 0$ to $\lceil \log_2 n \rceil - 1$ **do**

  sample each arm $i \in S_r$ for

$$t_r = \left\lfloor \frac{T}{|S_r| \lceil \log_2 n \rceil} \right\rfloor$$

  times, and let $\hat{p}_i^r$ be the average reward

  let S_{r+1} be the set of $\lceil |S_r|/2 \rceil$ arms in S_r with the largest empirical average

end for

output arm in $S_{\lceil \log_2 n \rceil}$

Fig 2: Sequential Halving algorithm

This strategy was firstly considered and presented by Even-Dar et al. (2002) For detailed, the algorithm is performed in several rounds, in which the remaining arms are uniformly sampled in one round. At the end of each round, some of the arms can be excluded according to specific criteria. The whole process continues until only one arm remains. Although there are many scientists have researched for MAB problem so far, this strategy is a better algorithm and improved analysis for this problem. This strategy seems easy to understand but it also has some challenges should be face.

3. The main challenge in designing such strategies

There are two main challenges in designing this strategy. First is every time we exclude one arm should be highly trusted because the eliminated arm will never be resurrected. Therefore, before eliminating every arm, we should highly trust my estimation of award of corresponding arm so that I can estimated next arm. Therefore, in the analysis, it is necessary to combine the constraints of the constraint and generate a logarithmic factor in the result boundary. The best way to avoid this problem is to try to reduce the times of exclusion rounds. What's more, we can design estimators for each round so that try to ensure our rest of the arms are sampled have high trust. Through this way, we can avoid most union-bound arguments which maybe be used in previous works. Second challenge happens in the fixed confidence setting. In order to exclude a suboptimal arm, we need to accurately estimate the gap from the best arm with the greatest return. At this point, we don't take the maximum of the empirical rewards of the surviving arms but we use a sub-procedure which based on the Median Elimination algorithm which can help us to find the suboptimality of each arm so that the accuracy can reach our requirement. (Karnin et al. 2013) Both two Exponential-Gap Elimination algorithm and Sequential Halving algorithm belong to Median Elimination algorithm.

4. The up bound of the performance of the optimal

policy

In this strategy, the tightness of the upper bounds over the times of arm pulls required to reach an expected award through those two settings have been researched by previous people. Therefore, what we do now is just to improve previous algorithm. These two algorithms are optimal under the influence of doubly- logarithmic factors but what the tasks we do is just limited to logarithmic factors. Therefore, how to solve the disadvantage of logarithmic factor decides the up bound of the performance of this strategy. Fortunately, this advantage can be improved through our sequential Halving algorithm. That’s the reason why this strategy is better than most of the methods.

5. Conclusion

Through this assignment, I have explored the complexity of machine learning. Although it seems easy to analyze the best arm(s), you should consider a lot of questions and carry out a lot of experiments and proofs. Even finally get a “best” arm, it not means you can earn more money because the whole process is prediction. Therefore, it is also an interesting problem which deserved to research in the future.

6. Reference

1) Karnin, Z., Koren, T. & Somekh, O. 2013, ‘Almost Optimal Exploration in Multi-Armed Bandits’, *International Conference on Machine Learning,* 13 December, pp. 1238-1246.

2) Auer, P., Cesa-Bianchi, N., Freund, Y. & Schapire, R.E. 1995, ‘Gambling in a rigged casino: The adversarial multi-armed bandit problem’, *Proceeding of IEEE 36th Annual Foundations of Computer Science,* 23-23 October, Milwaukee, WI, USA, pp. 322-331.

Link: https://github.com/Liangzhu-Jiang/UTS_ML2019_ID13064593.git