Novelty Assessment Report

Paper: Instance-Dependent Fixed-Budget Pure Exploration in Reinforcement Learning
PDF URL: https://openreview.net/pdf?id=xIycolc5Xw
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2025-12-29

Abstract

We study the problem of fixed budget pure exploration in reinforcement learning. The goal is to identify a near-optimal policy, given a fixed budget on the number of interactions with the environment. Unlike the standard PAC setting, we do not require the target error level ϵ and failure rate δ as input. We propose novel algorithms and provide, to the best of our knowledge, the first instance-dependent ϵ-uniform guarantee, meaning that the probability that ϵ-correctness is ensured can be obtained simultaneously for all ϵ above a budget-dependent threshold. It characterizes the budget requirements in terms of the problem-specific hardness of exploration. As a core component of our analysis, we derive a ϵ-uniform guarantee for the multiple bandit problem—solving multiple multi-armed bandit instances simultaneously—which may be of independent interest. To enable our analysis, we also develop tools for reward-free exploration under the fixed-budget setting, which we believe will be useful for future work.

Core Task Landscape

This paper addresses: **fixed-budget pure exploration in reinforcement learning**

A total of **50 papers** were analyzed and organized into a taxonomy with **29 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Multi-Armed Bandit Pure Exploration**
- **Episodic and Full MDP Pure Exploration**
- **Meta-Learning and In-Context Exploration Strategies**
- **Theoretical Foundations and Lower Bounds**
- **Applications and Domain-Specific Exploration**

Complete Taxonomy Tree

- fixed-budget pure exploration in reinforcement learning Survey Taxonomy
- Multi-Armed Bandit Pure Exploration
  - Standard Multi-Armed Bandit Best-Arm Identification
  - Fixed-Budget Best-Arm Identification Algorithms and Theory (3 papers)
    - [21] UCB Exploration for Fixed-Budget Bayesian Best Arm Identification (Zhu, 2024) View paper
    - [32] Open Problem: Optimal Best Arm Identification with Fixed Budget (Qin Chao, 2023) View paper
    - [47] On Sequential Elimination Algorithms for Best-Arm Identification in Multi-Armed Bandits (Shahin Shahrampour, 2016) View paper
  - Fixed-Confidence Pure Exploration (1 papers)
    - [29] Pure Exploration in Infinitely-Armed Bandit Models with Fixed-Confidence (Aziz, 2022) View paper
  - Infinite-Armed Bandit Pure Exploration (2 papers)
    - [15] Asymptotically optimal quantile pure exploration for infinite-armed bandits (EXY Gong, 2023) View paper
    - [30] Asymptotically Optimal Pure Exploration for Infinite-Armed Bandits (Gong, 2023) View paper
  - Robust and Risk-Averse Best-Arm Identification (1 papers)
    - [40] Statistically Robust, Risk-Averse Best Arm Identification in Multi-Armed Bandits (Anmol Kagrecha, 2022) View paper
  - Combinatorial and Structured Bandit Pure Exploration
  - Combinatorial Pure Exploration with General Reward Functions (4 papers)
    - [12] Fixed-Budget Real-Valued Combinatorial Pure Exploration of Multi-Armed Bandit (Nakamura Shintaro, 2024) View paper
    - [31] Improved learning complexity in combinatorial pure exploration bandits (Gabillon, 2016) View paper
    - [37] Pure Exploration of Multi-armed Bandit Under Matroid Constraints (Chen Li-jie, 2022) View paper
    - [44] Disagreement-Based Combinatorial Pure Exploration: Sample Complexity Bounds and an Efficient Algorithm (Cao, 2017) View paper
  - Combinatorial Pure Exploration with Bottleneck Rewards (1 papers)
    - [7] Combinatorial pure exploration with bottleneck reward function (Du Yi-han, 2021) View paper
  - Thresholding and Constrained Pure Exploration (5 papers)
    - [3] Constrained pure exploration multi-armed bandits with a fixed budget (Faizal, 2022) View paper
    - [8] Learning to explore with Lagrangians for bandits under unknown constraints (U Das, 2024) View paper
    - [23] Thompson sampling-based recursive block elimination for dynamic assignment under limited budget in pure-exploration (S. P. Parambath, 2025) View paper
    - [26] Learning to Explore with Lagrangians for Bandits under Unknown Linear Constraints (U. C. Das, 2024) View paper
    - [49] An optimal algorithm for the Thresholding Bandit Problem (Locatelli, 2016) View paper

    - Constrained Best Mixed Arm Identification (1 papers)
        - [19] Pure exploration for constrained best mixed arm identification with a fixed budget (Tang, 2024) View paper
    - Linear and Structured Bandit Pure Exploration
    - Linear Bandit Pure Exploration Algorithms (3 papers)
        - [4] Robust pure exploration in linear bandits with limited budget (Ayya Alieva, 2021) View paper
        - [6] Gamification of pure exploration for linear bandits (Degenne, 2020) View paper
        - [45] An Empirical Process Approach to the Union Bound: Practical Algorithms for Combinatorial and Linear Bandits (Katz-Samuels, 2020) View paper
    - Model Selection in Linear Bandit Pure Exploration (1 papers)
        - [41] Near Instance Optimal Model Selection for Pure Exploration Linear Bandits (Yinglun Zhu, 2021) View paper
    - Multi-Task Representation Learning for Pure Exploration (1 papers)
        - [17] Multi-task Representation Learning for Fixed Budget Pure-Exploration in Linear and Bilinear Bandits (S Mukherjee, 2025) View paper
    - Kernel Bandit Pure Exploration (1 papers)
        - [28] Collaborative Pure Exploration in Kernel Bandit (Du Yi-han, 2022) View paper
    - Specialized Bandit Structures and Applications
    - Multinomial Logit Bandit Pure Exploration (1 papers)
        - [1] Fixed-Budget Pure Exploration in Multinomial Logit Bandits. (Boli Fang, 2022) View paper
    - Unimodal Bandit Pure Exploration (1 papers)
        - [27] UB3: Fixed Budget Best Beam Identification in mmWave Massive MISO via Pure Exploration Unimodal Bandits (Debamita Ghosh, 2024) View paper
    - Pure Exploration with General Distribution-Dependent Rewards (1 papers)
        - [9] The pure exploration problem with general reward functions depending on full distributions (Siwei Wang, 2022) View paper
    - Federated and Distributed Pure Exploration (1 papers)
    - [11] Pure exploration in asynchronous federated bandits (Wang Zichen, 2023) View paper
- Episodic and Full MDP Pure Exploration
    - Episodic Fixed-Horizon MDP Pure Exploration ★ (2 papers)
    - [0] Instance-Dependent Fixed-Budget Pure Exploration in Reinforcement Learning (Anon et al., 2026) View paper
    - [24] Pure Exploration in Episodic Fixed-Horizon Markov Decision Processes. (Sudeep Raja Putta, 2017) View paper
    - Budgeted and Constrained MDP Exploration (1 papers)
    - [33] Budgeted reinforcement learning in continuous state space (Carrara Nicolas, 2019) View paper
    - Reward-Free and Initial Pure Exploration in RL (1 papers)
    - [43] Efficient Reinforcement Learning via Initial Pure Exploration (Putta, 2017) View paper
- Meta-Learning and In-Context Exploration Strategies (2 papers)
    - [10] Learning to Explore: An In-Context Learning Approach for Pure Exploration (A Russo, 2025) View paper
    - [18] Meta-Learning Exploration Strategies with Decision Transformers (Welch, 2025) View paper
- Theoretical Foundations and Lower Bounds (4 papers)
    - [13] Contributions to a Theory of Pure Exploration in Sequential Statistics (Barrier, 2023) View paper
    - [16] Pure exploration in multi-armed bandits problems (SÃ©bastien Bubeck, 2009) View paper
    - [20] Pure Exploration in Multi-Armed Bandits (Stephens, 2023) View paper
    - [36] Fundamental Limits in Stochastic Bandits (Wang, 2024) View paper
- Applications and Domain-Specific Exploration
    - Hyperparameter Optimization and Active Model Selection (3 papers)
    - [22] Hyperband: A novel bandit-based approach to hyperparameter optimization (Li LiSha, 2018) View paper
    - [38] Pre-Training Acquisition Functions by Deep Reinforcement Learning for Fixed Budget Active Learning (Yusuke Taguchi, 2021) View paper
    - [42] Active Model Selection (Madani, 2022) View paper
    - Resource Allocation and Scheduling Applications (3 papers)
    - [5] AEM-D3QN: A Graph-Based Deep Reinforcement Learning Framework for Dynamic Earth Observation Satellite Mission Planning (Shuo Li, 2025) View paper
    - [14] Efficient Multi-Agent Exploration in Area Coverage Under Spatial and Resource Constraints (Maram Hasan, 2025) View paper
    - [46] Making the Cut: A Bandit-based Approach to Tiered Interviewing (Candice Schumann, 2019) View paper
    - Communication and Network Applications (1 papers)
    - [35] Pure-exploration bandits for channel selection in mission-critical wireless communications (Yuan Xue, 2018) View paper
    - Clustering and Similarity-Based Exploration (1 papers)
    - [34] Query-Efficient Correlation Clustering with Noisy Oracle (Kuroki Yuko, 2024) View paper
    - Game-Theoretic and Simulation-Based Exploration (2 papers)
    - [39] Learning Probably Approximately Correct Maximin Strategies in Simulation-Based Games with Infinite Strategy Spaces (Marchesi, 2022) View paper
    - [50] CURIOSITY INCREASES EQUALITY IN COMPETITIVE RESOURCE ALLOCATION (B Bucher, n.d.) View paper
    - Healthcare and Domain-Specific Decision Making (1 papers)
    - [25] Applying reinforcement learning techniques to detect hepatocellular carcinoma under limited screening capacity (Elliot Lee, 2015) View paper
    - Large Language Model Exploration Budget Optimization (1 papers)
    - [2] Knapsack rl: Unlocking exploration of llms via optimizing budget allocation (Li, 2025) View paper
    - Cognitive and Uncertainty-Directed Exploration (1 papers)
    - [48] Cognitive Ranking of Information Patches by Equiprobable Pure Exploration and Uncertainty-directed Search (WIEBRINGHAUS, n.d.) View paper

Narrative

Core task: fixed-budget pure exploration in reinforcement learning. This field addresses the challenge of identifying optimal policies or states when an agent has a strict, predetermined budget of interactions with the environment. The taxonomy reveals a rich structure organized around problem formulations and methodological approaches. Multi-Armed Bandit Pure Exploration forms a foundational branch, encompassing classical best-arm identification and extensions to combinatorial, constrained, and structured settings, with works like Pure Exploration Multi-Armed[16] and Combinatorial Pure Exploration[31] establishing early frameworks. Episodic and Full MDP Pure Exploration extends these ideas to sequential decision problems, where agents must explore state-action spaces under episodic constraints, as exemplified by Episodic Fixed-Horizon MDP[24]. Meta-Learning and In-Context Exploration Strategies represent a newer direction, leveraging prior task experience to accelerate exploration, with In-Context Pure Exploration[10] and Meta-Learning Exploration[18] demonstrating how learned policies can adapt quickly. Theoretical Foundations and Lower Bounds provide rigorous characterizations of sample complexity, while Applications and Domain-Specific Exploration translate these methods to real-world problems ranging from hyperparameter tuning to resource allocation.

Several active lines of work reveal key trade-offs between generality and efficiency. Constrained and budgeted settings, explored in Constrained Budget Bandits[3] and Knapsack RL[2], introduce resource limitations beyond sample counts, adding practical realism but complicating algorithm design. Instance-dependent approaches seek to exploit problem structure for tighter guarantees, contrasting with worst-case analyses. Within this landscape, Instance Dependent Budget Exploration[0] sits naturally in the Episodic Fixed-Horizon MDP branch alongside Episodic Fixed-Horizon MDP[24], focusing on how instance-specific properties can be leveraged to improve fixed-budget exploration in episodic settings. While Episodic Fixed-Horizon MDP[24] provides foundational algorithms for this problem class, Instance Dependent Budget Exploration[0] emphasizes adaptive strategies that tailor exploration to the particular MDP instance at hand, bridging classical episodic methods with the growing interest in instance-optimality seen in works like Instance Optimal Linear[41].

Related Works in Same Category

The following **1 sibling papers** share the same taxonomy leaf node with the original paper:

1. Pure Exploration in Episodic Fixed-Horizon Markov Decision Processes.

Authors: Sudeep Raja Putta, Theja Tulabandhula | **Year/Venue**: 2017 | **URL**: View paper

Abstract

Multi-Armed Bandit (MAB) problems can be naturally extended to Markov Decision Processes (MDP). We extend the Best Arm Identification problem to episodic fixed-horizon MDPs. Here, the goal of an agent interacting with the MDP is to reach a high confidence on the optimal policy in as few episodes as possible. We propose Posterior Sampling for Pure Exploration (PSPE), a Bayesian algorithm for pure exploration in MDPs. We empirically show that PSPE achieves deep exploration and the number of episod...

Relationship Analysis

Both papers belong to the Episodic Fixed-Horizon MDP Pure Exploration category, studying best policy identification in episodic MDPs with fixed horizons. The original paper focuses on fixed-budget pure exploration with instance-dependent ε-uniform guarantees, proposing the BREA algorithm that does not require target error ε or failure rate δ as input. The candidate paper addresses a two-phase exploration problem where pure exploration occurs in a training phase followed by an evaluation phase, proposing PSPE (a Bayesian posterior sampling approach) to reach high confidence quickly, which differs from the original paper's focus on budget-dependent error guarantees and backward reachability estimation.

Contributions Analysis

Overall novelty summary. The paper proposes the BREA algorithm for fixed-budget pure exploration in episodic MDPs, introducing the first instance-dependent ε-uniform guarantee that simultaneously ensures ε-correctness for all ε above a budget-dependent threshold. Within the taxonomy, it resides in the 'Episodic Fixed-Horizon MDP Pure Exploration' leaf, which contains only two papers total. This sparse population suggests the specific combination of fixed-budget constraints, episodic MDPs, and instance-dependent guarantees represents a relatively underexplored research direction compared to the more crowded multi-armed bandit branches.

The taxonomy reveals substantial activity in adjacent areas: the parent 'Episodic and Full MDP Pure Exploration' branch includes work on budgeted/constrained MDPs and reward-free exploration, while sibling branches address multi-armed bandits with various structural assumptions (combinatorial, linear, robust). The paper's positioning bridges classical episodic MDP exploration with instance-optimality concepts more commonly studied in bandit settings. Its scope explicitly targets episodic fixed-horizon problems, excluding infinite-horizon or continuous-state formulations, and distinguishes itself from the reward-free exploration work by maintaining a pure exploration objective rather than a preparatory phase.

Among the three identified contributions, the literature search examined nine candidates total with no clear refutations found. The BREA algorithm contribution examined three candidates with none providing overlapping prior work; similarly, the multiple bandit problem guarantee and fixed-budget reward-free tools each examined three candidates without refutation. This limited search scope (nine papers, not hundreds) means the analysis captures nearby semantic matches but cannot claim exhaustive coverage. The absence of refutations among these candidates suggests the specific technical contributions—particularly the ε-uniform guarantee formulation—may represent novel angles within the examined neighborhood.

Based on the top-nine semantic matches examined, the work appears to occupy a distinctive position combining fixed-budget constraints with instance-dependent analysis in episodic MDPs. The sparse taxonomy leaf and lack of refutations among examined candidates suggest novelty, though the limited search scope means potentially relevant work in broader RL theory or alternative formulations may exist outside this analysis. The multiple bandit subproblem and reward-free exploration tools appear to serve as technical enablers rather than standalone contributions with extensive prior literature.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: BREA algorithm with instance-dependent ε-uniform guarantee

Description: The authors introduce BREA, a fixed-budget pure exploration algorithm for episodic MDPs that provides instance-dependent guarantees. The algorithm characterizes budget requirements in terms of problem-specific exploration hardness and ensures ε-correctness simultaneously for all ε above a budget-dependent threshold, without requiring ε or δ as input.

This contribution was assessed against **3 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. A Theory of Active Learning in Dynamic Environments

URL: View paper

Brief Assessment

Active Learning Dynamic[56] focuses on continuous control and reinforcement learning settings but does not address fixed-budget pure exploration in episodic MDPs with ε-uniform guarantees as proposed in the original paper.

---

2. Rate-Optimal Strategy for Best Policy Identification in Reinforcement Learning

URL: View paper

Brief Assessment

Rate-Optimal Policy Identification[54] focuses on rate optimality within fixed-budget ranking and selection (R&S) framework for tabular MDPs, which is a different theoretical framework than the ε-uniform guarantee approach presented in the original paper's BREA algorithm.

---

3. Policy Testing in Markov Decision Processes

URL: View paper

Brief Assessment

Policy Testing MDPs[55] addresses a different problem (policy testing - determining if a policy's value exceeds a threshold) rather than policy identification. The candidate focuses on fixed-confidence setting with static sampling rules, while the original paper addresses fixed-budget pure exploration with ε-uniform guarantees across all ε values simultaneously without requiring ε or δ as input.

Contribution 2: ε-uniform guarantee for multiple bandit problem

Description: The authors provide the first ε-uniform guarantee for the Successive Accepts and Rejects (SAR) algorithm applied to the multiple bandit problem, where multiple multi-armed bandit instances must be solved simultaneously. This result may be of independent interest beyond the main MDP setting.

This contribution was assessed against **3 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Quantile Bandits for Best Arms Identification

Brief Assessment

Quantile Best Arms[53] focuses on quantile-based best arm identification with fixed budget, while the original paper addresses ε-uniform guarantees for the SAR algorithm in multiple bandit problems with standard mean-based objectives. These are fundamentally different problem formulations.

2. Dart: Adaptive accept reject algorithm for non-linear combinatorial bandits

URL: View paper

Brief Assessment

Dart Adaptive Reject[51] addresses a different problem: non-linear combinatorial bandits where k arms are selected simultaneously with joint rewards. The original paper's contribution concerns ε-uniform guarantees for Successive Accepts and Rejects (SAR) in multiple bandit instances (solving multiple MAB problems simultaneously), which is a distinct setting from combinatorial selection.

3. Top Feasible-Arm Subset Identification in Constrained Multi-Armed Bandit with Limited Budget

URL: View paper

Brief Assessment

Top Feasible-Arm Subset[52] addresses a constrained multi-armed bandit problem with feasibility constraints (cost $\leq \tau$), not the multiple bandit problem (solving multiple MAB instances simultaneously) discussed in the original paper. The technical settings and objectives are fundamentally different.

Contribution 3: Fixed-budget reward-free exploration tools

Description: The authors develop new algorithmic and analytical tools for reward-free exploration under the fixed-budget setting by adapting the Learn2Explore (L2E) algorithm. They prove an ε-uniform guarantee for their fixed-budget reward-free algorithms, which they believe will be useful for future work.

This contribution was assessed against **3 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Adaptive Pure Exploration in Markov Decision Processes and Bandits

URL: View paper

Brief Assessment

Adaptive Pure Exploration[59] focuses on fixed-confidence settings with asymptotic optimality guarantees for best policy identification and reward-free exploration in episodic MDPs. The original paper develops fixed-budget tools with ε-uniform guarantees, which is a fundamentally different theoretical framework.

2. Multi-reward best policy identification

URL: View paper

Brief Assessment

Multi-Reward Policy Identification[57] focuses on best policy identification across multiple reward functions with fixed confidence guarantees, not fixed-budget reward-free exploration. The candidate addresses a different problem setting (multi-reward BPI) than the original's fixed-budget pure exploration framework.

3. Best policy identification in discounted linear mdps

URL: View paper

Brief Assessment

Discounted Linear MDPs[58] focuses on fixed-confidence PAC RL in discounted linear MDPs, not fixed-budget reward-free exploration. The paper explicitly states it operates in 'the fixed confidence setting' and aims to 'identify an ε-optimal policy with probability 1-δ', which is fundamentally different from the original paper's fixed-budget setting where no ε or δ are provided as input.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] Instance-Dependent Fixed-Budget Pure Exploration in Reinforcement Learning View paper
- [1] Fixed-Budget Pure Exploration in Multinomial Logit Bandits. View paper
- [2] Knapsack rl: Unlocking exploration of llms via optimizing budget allocation View paper
- [3] Constrained pure exploration multi-armed bandits with a fixed budget View paper
- [4] Robust pure exploration in linear bandits with limited budget View paper
- [5] AEM-D3QN: A Graph-Based Deep Reinforcement Learning Framework for Dynamic Earth Observation Satellite Mission Planning View paper
- [6] Gamification of pure exploration for linear bandits View paper
- [7] Combinatorial pure exploration with bottleneck reward function View paper
- [8] Learning to explore with Lagrangians for bandits under unknown constraints View paper

- [9] The pure exploration problem with general reward functions depending on full distributions View paper
- [10] Learning to Explore: An In-Context Learning Approach for Pure Exploration View paper
- [11] Pure exploration in asynchronous federated bandits View paper
- [12] Fixed-Budget Real-Valued Combinatorial Pure Exploration of Multi-Armed Bandit View paper
- [13] Contributions to a Theory of Pure Exploration in Sequential Statistics View paper
- [14] Efficient Multi-Agent Exploration in Area Coverage Under Spatial and Resource Constraints View paper
- [15] Asymptotically optimal quantile pure exploration for infinite-armed bandits View paper
- [16] Pure exploration in multi-armed bandits problems View paper
- [17] Multi-task Representation Learning for Fixed Budget Pure-Exploration in Linear and Bilinear Bandits View paper
- [18] Meta-Learning Exploration Strategies with Decision Transformers View paper
- [19] Pure exploration for constrained best mixed arm identification with a fixed budget View paper
- [20] Pure Exploration in Multi-Armed Bandits View paper
- [21] UCB Exploration for Fixed-Budget Bayesian Best Arm Identification View paper
- [22] Hyperband: A novel bandit-based approach to hyperparameter optimization View paper
- [23] Thompson sampling-based recursive block elimination for dynamic assignment under limited budget in pure-exploration View paper
- [24] Pure Exploration in Episodic Fixed-Horizon Markov Decision Processes. View paper
- [25] Applying reinforcement learning techniques to detect hepatocellular carcinoma under limited screening capacity View paper
- [26] Learning to Explore with Lagrangians for Bandits under Unknown Linear Constraints View paper
- [27] UB3: Fixed Budget Best Beam Identification in mmWave Massive MISO via Pure Exploration Unimodal Bandits View paper
- [28] Collaborative Pure Exploration in Kernel Bandit View paper
- [29] Pure Exploration in Infinitely-Armed Bandit Models with Fixed-Confidence View paper
- [30] Asymptotically Optimal Pure Exploration for Infinite-Armed Bandits View paper
- [31] Improved learning complexity in combinatorial pure exploration bandits View paper
- [32] Open Problem: Optimal Best Arm Identification with Fixed Budget View paper
- [33] Budgeted reinforcement learning in continuous state space View paper
- [34] Query-Efficient Correlation Clustering with Noisy Oracle View paper
- [35] Pure-exploration bandits for channel selection in mission-critical wireless communications View paper
- [36] Fundamental Limits in Stochastic Bandits View paper
- [37] Pure Exploration of Multi-armed Bandit Under Matroid Constraints View paper
- [38] Pre-Training Acquisition Functions by Deep Reinforcement Learning for Fixed Budget Active Learning View paper
- [39] Learning Probably Approximately Correct Maximin Strategies in Simulation-Based Games with Infinite Strategy Spaces View paper
- [40] Statistically Robust, Risk-Averse Best Arm Identification in Multi-Armed Bandits View paper
- [41] Near Instance Optimal Model Selection for Pure Exploration Linear Bandits View paper
- [42] Active Model Selection View paper
- [43] Efficient Reinforcement Learning via Initial Pure Exploration View paper
- [44] Disagreement-Based Combinatorial Pure Exploration: Sample Complexity Bounds and an Efficient Algorithm View paper
- [45] An Empirical Process Approach to the Union Bound: Practical Algorithms for Combinatorial and Linear Bandits View paper
- [46] Making the Cut: A Bandit-based Approach to Tiered Interviewing View paper
- [47] On Sequential Elimination Algorithms for Best-Arm Identification in Multi-Armed Bandits View paper
- [48] Cognitive Ranking of Information Patches by Equiprobable Pure Exploration and Uncertainty-directed Search View paper
- [49] An optimal algorithm for the Thresholding Bandit Problem View paper
- [50] CURIOSITY INCREASES EQUALITY IN COMPETITIVE RESOURCE ALLOCATION View paper
- [51] Dart: Adaptive accept reject algorithm for non-linear combinatorial bandits View paper
- [52] Top Feasible-Arm Subset Identification in Constrained Multi-Armed Bandit with Limited Budget View paper
- [53] Quantile Bandits for Best Arms Identification
- [54] Rate-Optimal Strategy for Best Policy Identification in Reinforcement Learning View paper
- [55] Policy Testing in Markov Decision Processes View paper
- [56] A Theory of Active Learning in Dynamic Environments View paper
- [57] Multi-reward best policy identification View paper
- [58] Best policy identification in discounted linear mdps View paper
- [59] Adaptive Pure Exploration in Markov Decision Processes and Bandits View paper