Novelty Assessment Report

Paper: An Improved Model-free Decision-estimation Coefficient with Applications in Adversarial MDPs

PDF URL: https://openreview.net/pdf?id=lbLAgGF8OO

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2025-12-27

Abstract

We study decision making with structured observation (DMSO). The complexity for DMSO has been characterized by a series of work [FKQR21 , CMB22 , FGH23]. Still, there is a gap between known regret upper and lower bounds: current upper bounds incur a model estimation error that scales with the size of the model class. The work of [FGQ+23] made an initial attempt to reduce the estimation error to only scale with the size of the value function set, resulting in the complexity called optimistic decision-estimation coefficient (optimistic DEC). Yet, their approach relies on the optimism principle to drive exploration, which deviates from the general idea of DEC that drives exploration only through information gain.

In this work, we introduce an improved model-free DEC, called Dig-DEC, that removes the optimism mechanism in [FGQ+23], making it more aligned with existing model-based DEC. Dig-DEC is always upper bounded by optimistic DEC, and could be significantly smaller in special cases. Importantly, the removal of optimism allows it to seamlessly handle adversarial environments, while it was unclear how to achieve it within the optimistic DEC framework. By applying Dig-DEC to hybrid MDPs where the transition is stochastic but the reward is adversarial, we provide the first model-free regret bounds in hybrid MDPs with bandit feedback in multiple settings: bilinear classes, Bellman-complete MDPs with bounded Bellman-eluder dimension or coverability, resolving the main open problem left by [LWZ25].

We also improve online function-estimation procedure used in model-free learning: For average estimation error minimization, we improve the estimator to achieve better concentration. This improves the $T^{\frac{3}{4}}$ and $T^{\frac{5}{6}}$ regret of [FGQ+23] to $T^{\frac{2}{3}}$and $T^{\frac{7}{9}}$ in the cases with on-policy and off-policy exploration. For squared estimation error minimization in Bellman-complete MDPs, we redesign the two-timescale procedure in [AZ22 , FGQ+23], achieving $\sqrt{T}$ regret that improves over the $T^{\frac{2}{3}}$ regret by [FGQ+23]. This is the first time the performance of a DEC-based approach for Bellman-complete MDPs matches that of optimism-based approaches [JLM21, XFB+23].

Disclaimer

This report is **AI-GENERATED** using Large Language Models and WisPaper (a scholar search engine). It analyzes academic papers' tasks and contributions against retrieved prior work. While this system identifies **POTENTIAL** overlaps and novel directions, **ITS COVERAGE IS NOT EXHAUSTIVE AND JUDGMENTS ARE APPROXIMATE**. These results are intended to assist human reviewers and **SHOULD NOT** be relied upon as a definitive verdict on novelty.

Note that some papers exist in multiple, slightly different versions (e.g., with different titles or URLs). The system may retrieve several versions of the same underlying work. The current automated pipeline does not reliably align or distinguish these cases, so human reviewers will need to disambiguate them manually.

If you have any questions, please contact: mingzhang23@m.fudan.edu.cn

Core Task Landscape

This paper addresses: **Decision Making with Structured Observation in Reinforcement Learning**

A total of **50 papers** were analyzed and organized into a taxonomy with **15 categories**.

Taxonomy Overview

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

- **Hierarchical Reinforcement Learning Approaches**
- **Representation Learning for Structured State Spaces**
- **Structured Observation and Partial Observability**
- **Multi-Agent and Cooperative Reinforcement Learning**
- **Compositional and Structured Task Learning**
- **Sample-Efficient and Representation-Driven RL**
- **Structured Exploration and Decision Estimation**
- **Domain-Specific Applications with Structured Control**

Complete Taxonomy Tree

- Decision Making with Structured Observation in Reinforcement Learning Survey Taxonomy
- Hierarchical Reinforcement Learning Approaches
  - Hierarchical Policy Learning with Explicit Decomposition (5 papers)
  - [2] Reinforcement learning from hierarchical critics (Z Cao, 2021) View paper
  - [3] Reinforcement Learning with Anticipation: A Hierarchical Approach for Long-Horizon Tasks (Yang, 2025) View paper
  - [7] Hierarchical reinforcement learning: A comprehensive survey (Pateria Shubham, 2021) View paper
  - [8] Intelligent problem-solving as integrated hierarchical reinforcement learning (Eppe, 2022) View paper
  - [15] Hierarchical reinforcement learning: A survey and open research challenges (Matthias Hutsebaut-Buysse, 2022) View paper
  - Latent-Based Hierarchical Methods (3 papers)
  - [13] Latent space policies for hierarchical reinforcement learning (Tuomas Haarnoja, 2018) View paper
  - [26] Stochastic neural networks for hierarchical reinforcement learning (Carlos Florensa, 2017) View paper
  - [34] Hierarchical reinforcement learning by discovering intrinsic options (Zhang, 2021) View paper
  - Data-Efficient and Transfer-Oriented Hierarchical RL (3 papers)
  - [4] Learning to reinforcement learn (Jane X. Wang, 2016) View paper
  - [18] Data-efficient hierarchical reinforcement learning (Ofir Nachum, 2018) View paper
  - [46] Multi-task reinforcement learning: a hierarchical bayesian approach (Aaron Wilson, 2007) View paper
- Representation Learning for Structured State Spaces
  - Disentangled and Factored Representations (3 papers)

    - [10] DEAR: Disentangled Environment and Agent Representations for Reinforcement Learning without Reconstruction (Ameya Pore, 2024) View paper
    - [16] Structured world belief for reinforcement learning in pomdp (Singh Gautam, 2021) View paper
    - [49] Pves: Position-velocity encoders for unsupervised learning of structured state representations (Jonschkowski, 2017) View paper
    - Contrastive and Predictive Representation Learning (4 papers)
    - [19] iQRL - Implicitly Quantized Representations for Sample-efficient Reinforcement Learning (KujanpÃ¤Ã¤, 2024) View paper
    - [23] Local-Guided Global: Paired Similarity Representation for Visual Reinforcement Learning (Hyesong Choi, 2023) View paper
    - [27] MOOSS: Mask-Enhanced Temporal Contrastive Learning for Smooth State Evolution in Visual Reinforcement Learning (Jiarui Sun, 2025) View paper
    - [45] Decoupling representation learning from reinforcement learning (Stooke, 2021) View paper
    - Structured and Quantized Representations (3 papers)
    - [12] Action-sufficient state representation learning for control with structural constraints (Huang Bi-wei, 2022) View paper
    - [35] Learning structures through reinforcement (Collins, 2018) View paper
    - [36] Dictionary Learning-Structured Reinforcement Learning With Adaptive-Sparsity Regularizer (Zhenni Li, 2024) View paper
    - Action-Sufficient and Task-Relevant Representations (4 papers)
    - [17] For sale: State-action representation learning for deep reinforcement learning (Fujimoto, 2023) View paper
    - [44] Which mutual-information representation learning objectives are sufficient for control? (Rakelly, 2021) View paper
    - [48] Learning telic-controllable state representations (Amir, 2024) View paper
    - [50] State Chrono Representation for Enhancing Generalization in Reinforcement Learning (Jianda Chen, 2024) View paper
    - Sequence Modeling and Transformer-Based Representations (4 papers)
    - [1] Structured State Space Models for In-Context Reinforcement Learning (Lu, 2023) View paper
    - [29] Enhancing Reinforcement Learning via Transformer-Based State Predictive Representations (Minsong Liu, 2024) View paper
    - [30] StARformer: Transformer With State-Action-Reward Representations for Robot Learning (Jinghuan Shang, 2022) View paper
    - [42] StARformer: Transformer with State-Action-Reward Representations for Visual Reinforcement Learning (Shang, 2021) View paper
- Structured Observation and Partial Observability
    - Entity-Based and Graph-Structured Observations (3 papers)
    - [5] Efficient entity-based reinforcement learning (Jankovics, 2022) View paper
    - [11] Efficient Graph Bandit Learning with Side-Observations and Switching Constraints (GONG Xueping, 2025) View paper
    - [24] Survey on Graph-Based Reinforcement Learning for Networked Coordination and Control (Yifan Liu, 2025) View paper
    - Belief State and World Model Learning for POMDPs (3 papers)
    - [14] On the role of information structure in reinforcement learning for partially-observable sequential teams and games (Awni Altabaa, 2024) View paper
    - [33] Kinematic state abstraction and provably efficient rich-observation reinforcement learning (Dipendra Misra, 2020) View paper
    - [41] A deep hierarchical reinforcement learning algorithm in partially observable Markov decision processes (Tuyen P. Le, 2018) View paper
- Multi-Agent and Cooperative Reinforcement Learning (4 papers)
    - [9] SigmaRL: A sample-efficient and generalizable multi-agent reinforcement learning framework for motion planning (XU Jianye, 2024) View paper
    - [37] A structured prediction approach for generalization in cooperative multi-agent reinforcement learning (Nicolas Carion, 2019) View paper
    - [39] Adaptive Reinforcement Learning for Fault-Tolerant Optimal Consensus Control of Nonlinear Canonical Multiagent Systems With Actuator Loss of Effectiveness (Boyan Zhu, 2024) View paper
    - [43] Multi-agent hierarchical reinforcement learning for energy management (Imen Jendoubi, 2023) View paper
- Compositional and Structured Task Learning (4 papers)
    - [6] Blocks assemble! learning to assemble with large-scale structured reinforcement learning (Ghasemipour, 2022) View paper
    - [20] Compositional reinforcement learning from logical specifications (Jothimurugan, 2021) View paper
    - [31] Learn2Assemble with Structured Representations and Search for Robotic Architectural Construction (Niklas Funk, 2021) View paper
    - [40] Composable deep reinforcement learning for robotic manipulation (Tuomas Haarnoja, 2018) View paper
- Sample-Efficient and Representation-Driven RL (3 papers)
    - [21] SURRL: Structural unsupervised representations for robot learning (Fengyi Zhang, 2022) View paper
    - [25] DynSyn: Dynamical Synergistic Representation for Efficient Learning and Control in Overactuated Embodied Systems (He, 2024) View paper
    - [47] Sample efficient reinforcement learning for building control: Leveraging physics informed latent representations (Gargya Gokhale, 2024) View paper
- Structured Exploration and Decision Estimation ★ (2 papers)
    - [0] An Improved Model-free Decision-estimation Coefficient with Applications in Adversarial MDPs (Anon et al., 2026) View paper
    - [38] Exploiting Exogenous Structure for Sample-Efficient Reinforcement Learning (Wan Jia, 2024) View paper
- Domain-Specific Applications with Structured Control (3 papers)
    - [22] An Optimal Obstacle Avoidance Method Using Reinforcement Learning-Based Decision Parameterization for Autonomous Vehicles (Qin Zhaobo, 2025) View paper
    - [28] Reinforcement Learning for Fuzzy Structured Adaptive Optimal Control of Discrete-Time Nonlinear Complex Networks (Tao Wu, 2024) View paper
    - [32] ORAN-GUIDE: RAG-Driven Prompt Learning for LLM-Augmented Reinforcement Learning in O-RAN Network Slicing (Lotfi, 2025) View paper

Narrative

Core task: decision making with structured observation in reinforcement learning. The field addresses how agents can exploit structure in their observations—whether hierarchical, compositional, or relational—to improve learning efficiency and generalization. The taxonomy reflects a diverse landscape organized around eight main branches. Hierarchical Reinforcement Learning Approaches (e.g., Hierarchical RL Survey[7], Hierarchical Critics[2]) focus on temporal abstraction and option discovery to decompose complex tasks. Representation Learning for Structured State Spaces emphasizes learning compact or disentangled encodings of high-dimensional observations, often leveraging entity-based or graph-based representations (Entity Based RL[5], Structured State Space[1]). Structured Observation and Partial Observability tackles settings where agents must infer hidden state from incomplete information (Information

Structure POMDP[14], Structured World Belief[16]). Multi-Agent and Cooperative Reinforcement Learning examines coordination and communication in multi-agent systems, while Compositional and Structured Task Learning explores modular skill composition (Blocks Assemble[6], Composable Deep RL[40]). Sample-Efficient and Representation-Driven RL targets data efficiency through better state abstractions, and Domain-Specific Applications with Structured Control applies these ideas to robotics, energy systems, and other real-world domains.

Within this landscape, a particularly active line of work centers on structured exploration and decision estimation, where agents must balance exploration with exploiting known structure to guide policy search. Decision Estimation Coefficient[0] sits squarely in this branch, proposing a novel complexity measure for decision-making that accounts for structured observations and sample efficiency. This contrasts with nearby efforts such as Exogenous Structure Sampling[38], which focuses on leveraging exogenous variables to improve sample complexity, and Anticipation Hierarchy[3], which uses hierarchical anticipation to guide exploration in temporally extended tasks. While hierarchical methods like Anticipation Hierarchy[3] emphasize temporal decomposition, Decision Estimation Coefficient[0] offers a more general framework for quantifying decision complexity across diverse structured environments. The interplay between exploration strategies, representation quality, and sample efficiency remains a central open question, with recent works exploring how to tightly integrate structural priors into both the learning objective and the exploration mechanism.

Related Works in Same Category

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

1. Exploiting Exogenous Structure for Sample-Efficient Reinforcement Learning

Authors: Wan Jia, Jia Wan, Sinclair, Sean R., Sean R. Sinclair, et al. (11 authors total) | **Year/Venue**: 2024 • arXiv.org | **URL**: View paper

Abstract

We study Exo-MDPs, a structured class of Markov Decision Processes (MDPs) where the state space is partitioned into exogenous and endogenous components. Exogenous states evolve stochastically, independent of the agent's actions, while endogenous states evolve deterministically based on both state components and actions. Exo-MDPs are useful for applications including inventory control, portfolio management, and ride-sharing. Our first result is structural, establishing a representational equivale...

Relationship Analysis

Both papers belong to the Structured Exploration and Decision Estimation category, focusing on efficient learning through structured approaches in reinforcement learning. The original paper introduces Dig-DEC, a model-free decision-estimation coefficient for handling both stochastic and adversarial MDPs with general function approximation, while the candidate paper focuses specifically on Exo-MDPs where state spaces are partitioned into exogenous (action-independent) and endogenous (action-dependent) components. The key difference is that the original paper develops a general complexity measure and algorithmic framework applicable across various MDP structures, whereas the candidate paper exploits a specific structural assumption about state decomposition to achieve sample efficiency that decouples from action and endogenous state sizes.

Contributions Analysis

Overall novelty summary. The paper introduces Dig-DEC, a model-free decision-estimation coefficient that removes the optimism mechanism from prior work while maintaining alignment with model-based DEC frameworks. It sits within the 'Structured Exploration and Decision Estimation' leaf of the taxonomy, which contains only two papers total. This is a notably sparse research direction compared to more crowded areas like hierarchical policy learning or representation learning, suggesting the paper addresses a relatively specialized problem within the broader field of decision-making with structured observations.

The taxonomy reveals that neighboring research directions emphasize different structural aspects: hierarchical methods focus on temporal abstraction, representation learning targets state encoding, and partial observability work addresses belief state inference. Dig-DEC connects to these areas by providing a complexity measure that can apply across diverse structured environments, but diverges by focusing specifically on information-driven exploration without optimism. The scope note for this leaf emphasizes 'decision-estimation coefficients or information-theoretic principles,' distinguishing it from unstructured exploration methods found elsewhere in the taxonomy.

Among the three contributions analyzed, the literature search examined seventeen candidates total. The first contribution (Dig-DEC framework) examined two candidates with no refutations found. The second contribution (hybrid MDP regret bounds) examined six candidates and found one refutable match, suggesting some prior work exists in this specific setting. The third contribution (online function-estimation procedures) examined nine candidates with no refutations, indicating relatively novel technical machinery. The limited search scope means these findings reflect top-K semantic matches rather than exhaustive coverage.

Based on the analysis of seventeen candidates, the work appears to occupy a sparsely populated research niche with moderate novelty across its contributions. The removal of optimism from decision-estimation frameworks represents a conceptual shift, though the hybrid MDP results show some overlap with existing literature. The analysis does not cover the full breadth of reinforcement learning theory, so additional related work may exist outside the examined candidate set.

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

Contribution 1: Dig-DEC: a model-free decision-estimation coefficient removing optimism

Description: The authors propose Dig-DEC, a new complexity measure for decision making with structured observations that eliminates the optimism principle used in prior work and instead relies solely on information gain for exploration. This measure is always no larger than optimistic DEC and can be significantly smaller in special cases.

This contribution was assessed against **2 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Regret Minimization via Saddle Point Optimization

URL: View paper

Brief Assessment

Saddle Point Regret[68] focuses on parametrizing the decision-estimation coefficient via confidence radius and solving saddle-point problems for anytime algorithms, rather than removing optimism from the DEC framework or replacing it with information gain exploration as the original paper proposes.

---

2. Unified Algorithms for RL with Decision-Estimation Coefficients: PAC, Reward-Free, Preference-Based Learning, and Beyond

URL: View paper

Brief Assessment

Unified Decision Estimation[67] focuses on extending DEC to multiple learning goals (PAC, reward-free, preference-based) rather than proposing a new complexity measure that removes optimism. The candidate's generalized DEC framework is orthogonal to the original paper's Dig-DEC contribution.

Contribution 2: First model-free regret bounds for hybrid MDPs with bandit feedback

Description: The authors establish the first sublinear regret bounds for model-free learning in hybrid MDPs (stochastic transitions with adversarial rewards) under bandit feedback, addressing an open problem from prior work that only handled full-information feedback.

This contribution was assessed against **6 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Optimal Hybrid Feedback-Driven Learning for Wireless Interactive Panoramic Scene Delivery

URL: View paper

Brief Assessment

Hybrid Feedback Panoramic[53] focuses on wireless panoramic scene delivery with two-level feedback (prediction and transmission), not on general hybrid MDPs with stochastic transitions and adversarial rewards.

---

2. Near-Optimal Regret Bounds for Model-Free RL in Non-Stationary Episodic MDPs

URL: View paper

Brief Assessment

Non Stationary Episodic[56] focuses on non-stationary MDPs where both rewards and transitions vary over time, not the hybrid MDP setting (fixed stochastic transitions with adversarial rewards) addressed by the original paper.

---

3. Near-optimal dynamic regret for adversarial linear mixture mdps

URL: View paper

Brief Assessment

Adversarial Linear Mixture[51] focuses on adversarial linear mixture MDPs with full-information feedback, not bandit feedback. The candidate explicitly states 'full-information feedback' in its problem setting, which is fundamentally different from the bandit feedback setting claimed as novel in the original paper.

---

4. Near-Optimal Regret for Adversarial MDP with Delayed Bandit Feedback

URL: View paper

Brief Assessment

Delayed Bandit Feedback[54] focuses on adversarial MDPs with delayed feedback, not hybrid MDPs (stochastic transitions with adversarial rewards). The candidate addresses a different problem setting involving delays in feedback observation rather than the hybrid MDP structure.

---

5. Bias no more: high-probability data-dependent regret bounds for adversarial bandits and MDPs

URL: View paper

Brief Assessment

High Probability Adversarial[52] focuses on adversarial MDPs with unknown transitions and bandit feedback, but does not address hybrid MDPs (stochastic transitions with adversarial rewards). The candidate's setting differs fundamentally from the original paper's hybrid MDP framework.

---

6. Beating Adversarial Low-Rank MDPs with Unknown Transition and Bandit Feedback

URL: View paper

Prior Art Analysis

Low Rank Unknown[55] demonstrates prior work exists for model-free learning in adversarial low-rank MDPs with bandit feedback and unknown transitions. The candidate paper presents Algorithm 5 (model-free) and Algorithm 3/4 (model-free, oracle-efficient) that achieve sublinear regret bounds for bandit feedback settings with unknown transitions. These algorithms address the same problem space as the original paper's contribution, showing that model-free approaches with bandit feedback in hybrid/adversarial MDPs were already established before the original paper's submission.

Evidence

Evidence 1 - **Rationale**: The candidate describes Algorithm 5 as a model-free algorithm for the bandit feedback setting, demonstrating that model-free approaches to this problem existed prior to the original paper's claimed contribution. - **Original**: By applying dig-dec to hybrid mdps with stochastic transitions and adversarial rewards, we obtain the first model-free regret bounds for hybrid mdps with bandit feedback under linear reward and several general transition structures, resolving the main open problem left by [lwz25]. - **Candidate**: algorithm 5 starts with a different exploration phase, where it calls vox (mhammedi et al., 2023) to learn a policy cover; vox is a model-free, reward-free exploration algorithm. after this initial exploratory phase, the algorithm also applies exponential weights and utilizes the same loss estimator...

Evidence 2 - **Rationale**: Algorithm 3 is explicitly model-free (using only feature class φ, not transition models) and handles bandit feedback in adversarial low-rank MDPs, showing prior work on model-free bandit feedback approaches. - **Original**: we establish the first sublinear regret formodel-free learning in hybrid bilinear classes and bellmancomplete coverable mdps with linear reward and bandit feedback, resolving the open question in [lwz25]. - **Candidate**: algorithm 3 oracle efficient algorithm for adversarial low-rank mdps (oblivious adversary). input: number of rounds t, feature class φ, confidence parameter δ ∈(0, 1). 1: set ε ←t-1/3, nreg ←t2/3, ν ←n-1/2 reg , and t0 ←ε-2ad13h6 log(φ/δ). 2: get ψcov 1:h ←vox(φ, ε, δ). // compute policy cover with ...

Contribution 3: Improved online function-estimation procedures with sharper regret bounds

Description: The authors develop refined online function estimation procedures that achieve tighter concentration bounds. For average estimation error, they improve regret from T^(3/4) to T^(2/3) in on-policy settings and from T^(5/6) to T^(8/9) in off-policy settings. For squared error in Bellman-complete MDPs, they redesign the two-timescale procedure to improve regret from T^(2/3) to sqrt(T).

This contribution was assessed against **9 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. An Offlineâ□□Cascadeâ□□Online Learningâ□□Based Algorithm for Distal Trajectory Estimation of Medical Continuum Manipulators

URL: View paper

Brief Assessment

Offline Cascade Online[57] focuses on trajectory estimation for continuum manipulators using neural networks, not reinforcement learning or online function estimation with regret bounds. The technical domains are entirely different.

---

2. Universal Online Learning: an Optimistically Universal Learning Rule

URL: View paper

Brief Assessment

Universal Online Learning[62] focuses on universal consistency in online learning with non-i.i.d. processes and k-nearest neighbor algorithms, not on online function estimation procedures or regret bounds in reinforcement learning contexts.

---

3. Thompson Sampling in Online RLHF with General Function Approximation

URL: View paper

Brief Assessment

Thompson Sampling RLHF[65] focuses on online RLHF with preference data and action value function approximation using posterior sampling, not on general online function estimation procedures in adversarial or stochastic MDPs with the specific average/squared error frameworks discussed in the original paper.

---

4. Decentralized online learning with kernels

URL: View paper

Brief Assessment

Decentralized Kernel Learning[66] addresses multi-agent stochastic optimization in RKHS with distributed consensus constraints, not centralized online function estimation with concentration bounds for MDP value functions. The technical settings and objectives are fundamentally different.

---

5. First-Order Regret in Reinforcement Learning with Linear Function Approximation: A Robust Estimation Approach

URL: View paper

Brief Assessment

First Order Regret[64] focuses on first-order regret bounds in linear MDPs using robust estimation techniques, while the original paper addresses general online function estimation in decision-estimation coefficient frameworks with different technical approaches (batching and two-timescale procedures).

---

6. Concentration bounds for temporal difference learning with linear function approximation: the case of batch data and uniform sampling

URL: View paper

Brief Assessment

Temporal Difference Concentration[63] focuses on batch data settings with uniform sampling for policy evaluation, not online function estimation with regret bounds in adversarial or stochastic MDPs as in the original paper.

---

7. Martingale methods for sequential estimation of convex functionals and divergences

URL: View paper

Brief Assessment

Martingale Sequential Estimation[61] focuses on sequential estimation of convex functionals and divergences using martingale methods, not on online function estimation in reinforcement learning contexts with regret bounds for MDPs.

---

8. Distributionally Robust Off-Dynamics Reinforcement Learning: Provable Efficiency with Linear Function Approximation

URL: View paper

Brief Assessment

Distributionally Robust Off Dynamics[60] focuses on off-dynamics RL with distributionally robust MDPs and linear function approximation, not on general online function estimation procedures or the specific regret improvements (T^(3/4) to T^(2/3), T^(5/6) to T^(8/9), T^(2/3) to sqrt(T)) claimed in the original paper.

---

9. On adaptive resampling strategies for sequential Monte Carlo methods

URL: View paper

Brief Assessment

Adaptive Resampling Sequential[59] focuses on Sequential Monte Carlo methods for sampling from probability distributions using importance sampling and resampling. This is fundamentally different from the original paper's work on online function estimation in reinforcement learning with regret bounds for MDPs.

Appendix: Text Similarity Detection

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

References

- [0] An Improved Model-free Decision-estimation Coefficient with Applications in Adversarial MDPs View paper
- [1] Structured State Space Models for In-Context Reinforcement Learning View paper
- [2] Reinforcement learning from hierarchical critics View paper
- [3] Reinforcement Learning with Anticipation: A Hierarchical Approach for Long-Horizon Tasks View paper
- [4] Learning to reinforcement learn View paper
- [5] Efficient entity-based reinforcement learning View paper
- [6] Blocks assemble! learning to assemble with large-scale structured reinforcement learning View paper
- [7] Hierarchical reinforcement learning: A comprehensive survey View paper
- [8] Intelligent problem-solving as integrated hierarchical reinforcement learning View paper
- [9] SigmaRL: A sample-efficient and generalizable multi-agent reinforcement learning framework for motion planning View paper
- [10] DEAR: Disentangled Environment and Agent Representations for Reinforcement Learning without Reconstruction View paper
- [11] Efficient Graph Bandit Learning with Side-Observations and Switching Constraints View paper
- [12] Action-sufficient state representation learning for control with structural constraints View paper
- [13] Latent space policies for hierarchical reinforcement learning View paper
- [14] On the role of information structure in reinforcement learning for partially-observable sequential teams and games View paper

- [15] Hierarchical reinforcement learning: A survey and open research challenges View paper
- [16] Structured world belief for reinforcement learning in pomdp View paper
- [17] For sale: State-action representation learning for deep reinforcement learning View paper
- [18] Data-efficient hierarchical reinforcement learning View paper
- [19] iQRL - Implicitly Quantized Representations for Sample-efficient Reinforcement Learning View paper
- [20] Compositional reinforcement learning from logical specifications View paper
- [21] SURRL: Structural unsupervised representations for robot learning View paper
- [22] An Optimal Obstacle Avoidance Method Using Reinforcement Learning-Based Decision Parameterization for Autonomous Vehicles View paper
- [23] Local-Guided Global: Paired Similarity Representation for Visual Reinforcement Learning View paper
- [24] Survey on Graph-Based Reinforcement Learning for Networked Coordination and Control View paper
- [25] DynSyn: Dynamical Synergistic Representation for Efficient Learning and Control in Overactuated Embodied Systems View paper
- [26] Stochastic neural networks for hierarchical reinforcement learning View paper
- [27] MOOSS: Mask-Enhanced Temporal Contrastive Learning for Smooth State Evolution in Visual Reinforcement Learning View paper
- [28] Reinforcement Learning for Fuzzy Structured Adaptive Optimal Control of Discrete-Time Nonlinear Complex Networks View paper
- [29] Enhancing Reinforcement Learning via Transformer-Based State Predictive Representations View paper
- [30] StARformer: Transformer With State-Action-Reward Representations for Robot Learning View paper
- [31] Learn2Assemble with Structured Representations and Search for Robotic Architectural Construction View paper
- [32] ORAN-GUIDE: RAG-Driven Prompt Learning for LLM-Augmented Reinforcement Learning in O-RAN Network Slicing View paper
- [33] Kinematic state abstraction and provably efficient rich-observation reinforcement learning View paper
- [34] Hierarchical reinforcement learning by discovering intrinsic options View paper
- [35] Learning structures through reinforcement View paper
- [36] Dictionary Learning-Structured Reinforcement Learning With Adaptive-Sparsity Regularizer View paper
- [37] A structured prediction approach for generalization in cooperative multi-agent reinforcement learning View paper
- [38] Exploiting Exogenous Structure for Sample-Efficient Reinforcement Learning View paper
- [39] Adaptive Reinforcement Learning for Fault-Tolerant Optimal Consensus Control of Nonlinear Canonical Multiagent Systems With Actuator Loss of Effectiveness View paper
- [40] Composable deep reinforcement learning for robotic manipulation View paper
- [41] A deep hierarchical reinforcement learning algorithm in partially observable Markov decision processes View paper
- [42] StARformer: Transformer with State-Action-Reward Representations for Visual Reinforcement Learning View paper
- [43] Multi-agent hierarchical reinforcement learning for energy management View paper
- [44] Which mutual-information representation learning objectives are sufficient for control? View paper
- [45] Decoupling representation learning from reinforcement learning View paper
- [46] Multi-task reinforcement learning: a hierarchical bayesian approach View paper
- [47] Sample efficient reinforcement learning for building control: Leveraging physics informed latent representations View paper
- [48] Learning telic-controllable state representations View paper
- [49] Pves: Position-velocity encoders for unsupervised learning of structured state representations View paper
- [50] State Chrono Representation for Enhancing Generalization in Reinforcement Learning View paper
- [51] Near-optimal dynamic regret for adversarial linear mixture mdps View paper
- [52] Bias no more: high-probability data-dependent regret bounds for adversarial bandits and MDPs View paper
- [53] Optimal Hybrid Feedback-Driven Learning for Wireless Interactive Panoramic Scene Delivery View paper
- [54] Near-Optimal Regret for Adversarial MDP with Delayed Bandit Feedback View paper
- [55] Beating Adversarial Low-Rank MDPs with Unknown Transition and Bandit Feedback View paper
- [56] Near-Optimal Regret Bounds for Model-Free RL in Non-Stationary Episodic MDPs View paper
- [57] An Offlineâ□□Cascadeâ□□Online Learningâ□□Based Algorithm for Distal Trajectory Estimation of Medical Continuum Manipulators View paper
- [58] Online estimation for functional data View paper
- [59] On adaptive resampling strategies for sequential Monte Carlo methods View paper
- [60] Distributionally Robust Off-Dynamics Reinforcement Learning: Provable Efficiency with Linear Function Approximation View paper
- [61] Martingale methods for sequential estimation of convex functionals and divergences View paper
- [62] Universal Online Learning: an Optimistically Universal Learning Rule View paper
- [63] Concentration bounds for temporal difference learning with linear function approximation: the case of batch data and uniform sampling View paper
- [64] First-Order Regret in Reinforcement Learning with Linear Function Approximation: A Robust Estimation Approach View paper
- [65] Thompson Sampling in Online RLHF with General Function Approximation View paper
- [66] Decentralized online learning with kernels View paper
- [67] Unified Algorithms for RL with Decision-Estimation Coefficients: PAC, Reward-Free, Preference-Based Learning, and Beyond View paper
- [68] Regret Minimization via Saddle Point Optimization View paper