Novelty Assessment Report

Paper: Distributional value gradients for stochastic environments

PDF URL: https://openreview.net/pdf?id=6hZAo6fZvJ

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2026-01-07

Abstract

Gradient-regularized value learning methods improve sample efficiency by leveraging learned models of transition dynamics and rewards to estimate return gradients. However, existing approaches, such as MAGE, struggle in stochastic or noisy environments, limiting their applicability. In this work, we address these limitations by extending distributional reinforcement learning on continuous state-action spaces to model not only the distribution over scalar state-action value functions but also over their gradients. We refer to this approach as Distributional Sobolev Training. Inspired by Stochastic Value Gradients (SVG), our method utilizes a one-step world model of reward and transition distributions implemented via a conditional Variational Autoencoder (cVAE). The proposed framework is sample-based and employs Max-sliced Maximum Mean Discrepancy (MSMMD) to instantiate the distributional Bellman operator. We prove that the Sobolev-augmented Bellman operator is a contraction with a unique fixed point, and highlight a fundamental smoothness trade-off underlying contraction in gradient-aware RL. To validate our method, we first showcase its effectiveness on a simple stochastic reinforcement-learning toy problem, then benchmark its performance on several MuJoCo environments.

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: **Modeling Return Distributions and Their Gradients in Stochastic Reinforcement Learning**

A total of **50 papers** were analyzed and organized into a taxonomy with **19 categories**.

Taxonomy Overview

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

- **Distributional Value Function Learning**
- **Distributional Policy Gradient Methods**
- **Model-Based Distributional RL and Gradient Methods**
- **Risk-Sensitive and Constrained Distributional RL**
- **Exploration and Uncertainty Quantification**
- **Specialized Applications and Extensions**
- **Theoretical Advances and Algorithmic Foundations**

Complete Taxonomy Tree

- Modeling Return Distributions and Their Gradients in Stochastic Reinforcement Learning Survey Taxonomy
- Distributional Value Function Learning
  - Categorical and Quantile-Based Distributional RL (4 papers)
  - [7] How Does Return Distribution in Distributional Reinforcement Learning Help Optimization? (Sun Ke, 2022) View paper
  - [20] Nonparametric return distribution approximation for reinforcement learning (Tetsuro Morimura, 2010) View paper
  - [29] IGN : Implicit Generative Networks (Haozheng Luo, 2022) View paper
  - [30] Diverse Projection Ensembles for Distributional Reinforcement Learning (Boehmer, 2023) View paper
  - Flow-Based and Continuous Distributional Representations (2 papers)
  - [22] Flow Models for Unbounded and Geometry-Aware Distributional Reinforcement Learning (C, 2025) View paper
  - [43] PACER: A Fully Push-forward-based Distributional Reinforcement Learning Algorithm (Wensong Bai, 2023) View paper
  - Theoretical Foundations of Distributional RL (2 papers)
  - [26] Intrinsic Benefits of Categorical Distributional Loss: Uncertainty-aware Regularized Exploration in Reinforcement Learning (Sun Ke, 2022) View paper
  - [42] Convergence Theorems for Entropy-Regularized and Distributional Reinforcement Learning (Jhaveri, 2025) View paper
- Distributional Policy Gradient Methods
  - Actor-Critic with Distributional Critics (6 papers)
  - [5] Distributed Distributional Deterministic Policy Gradients (Gabriel Barth-Maron, 2022) View paper
  - [8] Distributional policy gradient with distributional value function (Qi Liu, 2024) View paper
  - [11] Sample-based distributional policy gradient (Rahul Singh, 2022) View paper
  - [12] Distributional soft actor-critic: Off-policy reinforcement learning for addressing value estimation errors (Duan, 2021) View paper
  - [18] Bayesian distributional policy gradients (Faisal, 2021) View paper
  - [35] PG-Rainbow: Using Distributional Reinforcement Learning in Policy Gradient Methods (Jeon, 2024) View paper
  - Distributional Advantage Estimation (2 papers)
  - [3] Distributional Meta-Gradient Reinforcement Learning (H Yin, 2023) View paper
  - [4] Generalized Advantage Estimation for Distributional Policy Gradients (Smereka, 2025) View paper
  - Stochastic Policy Parameterization and Gradient Estimation (4 papers)
  - [6] Estimating or Propagating Gradients Through Stochastic Neurons (Bengio, 2022) View paper

  - [23] Stochastic variance-reduced policy gradient (Matteo Papini, 2018) View paper
  - [25] Improving stochastic policy gradients in continuous control with deep reinforcement learning using the beta distribution (Poâ□□Wei Chou, 2017) View paper
  - [37] Total stochastic gradient algorithms and applications in reinforcement learning (Parmas, 2018) View paper
- Model-Based Distributional RL and Gradient Methods
  - Bayesian Model-Based Distributional RL (3 papers)
  - [1] Value-distributional model-based reinforcement learning (Luis, 2024) View paper
  - [41] Offline Bayesian Aleatoric and Epistemic Uncertainty Quantification and Posterior Value Optimisation in Finite-State MDPs (Faisal, 2024) View paper
  - [47] Bayesian policy gradient algorithms (M Ghavamzadeh, 2006) View paper
  - Stochastic Value Gradients and World Models ★ (3 papers)
  - [0] Distributional value gradients for stochastic environments (Anon et al., 2026) View paper
  - [9] Stochastic optimization methods for policy evaluation in reinforcement learning (Yi Zhou, 2024) View paper
  - [16] Gradient estimation in model-based reinforcement learning: a study on linear quadratic environments (Ã□G Lovatto, 2021) View paper
- Risk-Sensitive and Constrained Distributional RL
  - Stochastic Dominance and Risk Measures (3 papers)
  - [2] Optimizing return distributions with distributional dynamic programming (Pires, 2025) View paper
  - [10] Stochastically Dominant Distributional Reinforcement Learning (Martin, 2022) View paper
  - [44] A Simple Mixture Policy Parameterization for Improving Sample Efficiency of CVaR Optimization (Luo Y, 2024) View paper
  - Safe and Constrained Distributional RL (4 papers)
  - [14] Safety-Optimized Fast Charging of Lithium-ion Battery Based on Distributional SAC-Conservative Augmented Lagrangian SDRL Algorithm (Liping chen, 2025) View paper
  - [15] Distributional constrained reinforcement learning for supply chain optimization (JS BermÃºdez, 2023) View paper
  - [21] Conservative offline distributional reinforcement learning (Ma, 2021) View paper
  - [24] Off-Policy Conservative Distributional Reinforcement Learning With Safety Constraints (Heng-Rui Zhang, 2024) View paper
- Exploration and Uncertainty Quantification
  - Deep Exploration via Distributional Ensembles (2 papers)
  - [46] Bag of Policies for Distributional Deep Exploration (Luchen Li, 2023) View paper
  - [48] MERL: Multi-Head Reinforcement Learning (Flet-Berliac, 2019) View paper
  - Return Distribution Regularization and Stability (2 papers)
  - [28] Policy optimization in a noisy neighborhood: On return landscapes in continuous control (Nate Rahn, 2023) View paper
  - [36] Moments Matter:Stabilizing Policy Optimization using Return Distributions (Dennis Jabs, 2026) View paper
- Specialized Applications and Extensions
  - Financial Portfolio Optimization (5 papers)
  - [19] From deterministic to stochastic: an interpretable stochastic model-free reinforcement learning framework for portfolio optimization (Zitao Song, 2022) View paper
  - [27] Reinforcement Learning for Risk-Aware Portfolio Optimization (Ahmed, 2025) View paper
  - [33] Optimizing Return Policies: A Bayesian Learning Approach (B Hajek, 2025) View paper
  - [38] APMPO: A Portfolio Management Policy Optimization Framework with Adaptive Reinforcement Learning Algorithm (F Gu, 2025) View paper
  - [40] RegimeFolio: A Regime Aware ML System for Sectoral Portfolio Optimization in Dynamic Markets (Yiyao Zhang, 2025) View paper
  - Multi-Objective and Pareto-Optimal Distributional RL (1 papers)
  - [17] Distributional pareto-optimal multi-objective reinforcement learning (XQ Cai, 2023) View paper
  - Continuous-Time and Specialized Control Settings (2 papers)
  - [13] On the Evolution of Return Distributions in Continuous-Time Reinforcement Learning (Wiltzer, 2021) View paper
  - [31] Policy Evaluation in Distributional LQR (Yulong Gao, 2023) View paper
  - Opponent Modeling and Multi-Agent Distributional RL (2 papers)
  - [49] Decision-making with Speculative Opponent Models (Jing Sun, 2022) View paper
  - [50] Zero-touch Continuous Network Slicing Control via Scalable Actor-Critic Learning (Rezazadeh, 2021) View paper
  - Offline and Post-Training Distributional RL (2 papers)
  - [32] Distributional policy optimization: An alternative approach for continuous control (Chen Tessler, 2019) View paper
  - [34] DVPO: Distributional Value Modeling-based Policy Optimization for LLM Post-Training (Dingwei Zhu, 2025) View paper
  - Stochastic Path Integration and Non-Standard Domains (1 papers)
  - [45] Search and return model for stochastic path integrators. (Macau, 2018) View paper
- Theoretical Advances and Algorithmic Foundations (1 papers)
  - [39] Advances in Distributional Reinforcement Learning: Bridging Theory with Algorithmic Practice (Ke, 2024) View paper

Narrative

Core task: Modeling return distributions and their gradients in stochastic reinforcement learning. The field has evolved into several major branches that reflect different ways of exploiting distributional information. Distributional Value Function Learning focuses on representing the full return distribution rather than just its mean, enabling richer value estimates and improved stability. Distributional Policy Gradient Methods extend gradient-based policy optimization to leverage return variability, with works like Distributional Policy Gradient[8] and Distributed D4PG[5] demonstrating how distributional critics can guide policy updates. Model-Based Distributional RL and Gradient Methods integrate learned world models with distributional representations, allowing agents to propagate uncertainty through planning. Risk-Sensitive and Constrained Distributional RL addresses safety and robustness by optimizing risk measures beyond expected return, while Exploration and Uncertainty Quantification uses distributional information to guide exploration strategies. Specialized Applications and Extensions apply these ideas to domains like finance and control, and Theoretical Advances and Algorithmic Foundations provide convergence guarantees and algorithmic principles.

A particularly active line of work explores how to compute and utilize gradients of return distributions within model-based settings, where stochastic dynamics and value uncertainty interact. Distributional Value Gradients[0] sits squarely in this space, emphasizing stochastic value gradients and world models to enable end-to-end differentiation through learned distributional predictions. This contrasts with purely model-free approaches like Distributional Meta Gradient[3], which meta-learns distributional features without explicit environment models, and with methods such as Stochastic Policy Evaluation[9] or Gradient Estimation Model[16] that focus on variance reduction or gradient estimation techniques in stochastic settings. The interplay between model-based planning, gradient propagation, and distributional representations remains an open question, with ongoing work examining trade-offs between sample efficiency, computational cost, and the fidelity of uncertainty estimates across these different methodological branches.

Related Works in Same Category

The following **2 sibling papers** share the same taxonomy leaf node with the original paper:

1. Stochastic optimization methods for policy evaluation in reinforcement learning

Authors: Yi Zhou, Shaocong Ma | **Year/Venue**: 2024 | **URL**: View paper

Abstract

â▯¦ Here, we discuss two types of RL â▯¦ gradient estimates obtained through stochastic gradient descent (SGD). Specifically focusing on SVRG, to reduce the variance in the gradient estimatesâ▯¦

Relationship Analysis

Both papers belong to the category of stochastic value gradients and world models, sharing a focus on gradient estimation methods in stochastic RL environments. The original paper develops distributional Sobolev training using conditional VAEs to model return distributions and their gradients jointly, while the candidate paper focuses on stochastic optimization methods (specifically SVRG) for policy evaluation through variance-reduced gradient descent. The key difference is that the original paper addresses distributional RL with learned world models for gradient estimation, whereas the candidate paper concentrates on optimization techniques for reducing variance in standard gradient-based policy evaluation.

2. Gradient estimation in model-based reinforcement learning: a study on linear quadratic environments

Authors: Ã▯G Lovatto, TP Bueno, LN de Barros | **Year/Venue**: 2021 | **URL**: View paper

Abstract

Stochastic Value Gradient (SVG) methods underlie many recent achievements of model-based Reinforcement Learning agents in continuous state-action spaces. Despite their practical significance, many algorithm design choices still lack rigorous theoretical or empirical justification. In this work, we analyze one such design choice: the gradient estimator formula. We conduct our analysis on randomized Linear Quadratic Gaussian environments, allowing us to empirically assess gradient estimation quali...

Relationship Analysis

Both papers belong to the Stochastic Value Gradients and World Models category, focusing on gradient estimation methods using learned stochastic models. The candidate paper analyzes gradient estimator formulas in SVG methods on Linear Quadratic Gaussian environments to justify design choices and understand bias-variance tradeoffs. The original paper extends this framework by introducing Distributional Sobolev Training, which models distributions over both returns and their gradients using conditional VAEs, whereas the candidate focuses on empirical analysis of existing SVG gradient estimators rather than distributional extensions.

Contributions Analysis

Overall novelty summary. The paper proposes Distributional Sobolev Training, which extends distributional RL to model both value distributions and their gradients in continuous state-action spaces. It resides in the 'Stochastic Value Gradients and World Models' leaf, which contains only three papers total including this one. This is a notably sparse research direction within the broader taxonomy of 50 papers, suggesting the specific combination of gradient-aware distributional learning with stochastic world models remains relatively underexplored compared to more populated branches like categorical distributional RL or actor-critic methods.

The taxonomy reveals that neighboring leaves pursue related but distinct approaches. The sibling category 'Bayesian Model-Based Distributional RL' focuses on epistemic uncertainty quantification through Bayesian inference rather than gradient modeling. Meanwhile, the parent category's other branch addresses policy gradient methods with distributional critics, which leverage return distributions for policy updates but do not explicitly model value gradients. The paper's use of cVAE-based world models and gradient propagation distinguishes it from purely model-free distributional methods in adjacent branches, positioning it at the intersection of model-based planning and gradient-regularized value learning.

Among 26 candidates examined across three contributions, no clearly refuting prior work was identified. The Distributional Sobolev framework examined six candidates with zero refutations, the contraction proofs examined ten candidates with zero refutations, and the MSMMD metric examined ten candidates with zero refutations. This suggests that within the limited search scope, the specific combination of distributional Bellman operators augmented with gradient information, contraction guarantees for Sobolev-augmented operators, and the MSMMD instantiation appear relatively novel. However, the modest search scale means potentially relevant work outside the top-26 semantic matches may exist.

Based on the limited literature search covering 26 candidates, the work appears to occupy a sparsely populated niche combining gradient-aware distributional learning with stochastic world models. The absence of refuting candidates across all contributions suggests novelty within the examined scope, though the small search scale and the paper's position in a three-paper taxonomy leaf indicate this assessment reflects top-K semantic proximity rather than exhaustive field coverage.

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

Contribution 1: Distributional Sobolev Reinforcement Learning framework

Description: The authors introduce a framework that models the joint distribution over both returns and their action-gradients, rather than treating gradients as auxiliary regularization. This is formalized through a novel Sobolev Bellman operator that bootstraps both return and gradient distributions simultaneously.

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. Distributional Meta-Gradient Reinforcement Learning

URL: View paper

Brief Assessment

Distributional Meta Gradient[3] focuses on learning adaptive distributional returns for meta-gradient RL in actor-critic settings, not on modeling joint distributions over returns and their action-gradients through a Sobolev Bellman operator as in the original paper.

2. Distributional reinforcement learning

URL: View paper

Brief Assessment

Distributional Reinforcement Learning[61] focuses on modeling return distributions without incorporating gradient information. The candidate does not address joint modeling of returns and their action-gradients through a Sobolev Bellman operator.

3. Beyond Marginals: Capturing Correlated Returns through Joint Distributional Reinforcement Learning

URL: View paper

Brief Assessment

Joint Distributional RL[71] focuses on modeling joint distributions over returns for different actions at the same state to capture cross-action correlations, not on jointly modeling returns and their action-gradients as in the original paper's Sobolev framework.

4. Distributional policy gradient with distributional value function

URL: View paper

Brief Assessment

Distributional Policy Gradient[8] focuses on sampling policy-gradient values from return distributions for exploration enhancement, not on jointly modeling return and gradient distributions through a Sobolev Bellman operator that bootstraps both simultaneously.

5. Using Exact Models to Analyze Policy Gradient Algorithms

URL: View paper

Brief Assessment

Exact Models Policy[70] focuses on analytically computing policy gradient landscapes for exactly solvable POMDPs using Markov chain models, not on distributional modeling of returns and gradients jointly.

6. Foundations of multivariate distributional reinforcement learning

URL: View paper

Brief Assessment

Multivariate Distributional RL[52] focuses on multivariate return distributions for multi-objective RL and zero-shot evaluation, not on jointly modeling return and gradient distributions for policy optimization as in the original paper's Sobolev framework.

Contribution 2: Contraction proofs for Sobolev Temporal Difference

Description: The authors provide the first contraction results for gradient-aware reinforcement learning, establishing that their Sobolev Bellman operator is contractive under both Wasserstein and max-sliced MMD metrics. They reveal a fundamental trade-off between smoothness constraints and discount factor for achieving contraction.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Stability and Generalization for Bellman Residuals

URL: View paper

Brief Assessment

Bellman Residuals Stability[67] focuses on stability and generalization bounds for Bellman residual minimization in offline RL, not on contraction properties of gradient-aware Bellman operators or Sobolev temporal difference methods.

2. Multi-Bellman operator for convergence of -learning with linear function approximation

URL: View paper

Brief Assessment

Multi Bellman Operator[64] focuses on contraction properties of multi-step Bellman operators in standard Q-learning with linear function approximation, not gradient-aware operators. The candidate does not address Sobolev operators or gradient-based temporal difference methods.

3. Exploring the Training Robustness of Distributional Reinforcement Learning Against Noisy State Observations

URL: View paper

Brief Assessment

Training Robustness Noisy[66] focuses on contraction properties of distributional Bellman operators under state observation noise in tabular and function approximation settings, not on gradient-aware Bellman operators or Sobolev temporal difference methods.

4. Implicit Constraint-Aware Off-Policy Correction for Offline Reinforcement Learning

URL: View paper

Brief Assessment

Implicit Constraint Aware[65] focuses on constraint-aware offline RL via proximal projections onto convex constraint sets, not gradient-aware Bellman operators or Sobolev temporal difference methods.

5. Robust Reinforcement Learning for Continuous Control with Model Misspecification

URL: View paper

Brief Assessment

Robust Model Misspecification[68] focuses on robustness to transition dynamics perturbations in continuous control, not gradient-aware Bellman operators or Sobolev temporal difference methods.

6. Distributional reinforcement learning

URL: View paper

Brief Assessment

Distributional Reinforcement Learning[61] provides contraction results for standard distributional Bellman operators, not for gradient-aware operators. The candidate does not establish contraction properties for Sobolev Bellman operators that bootstrap gradient distributions.

7. On the convergence of smooth regularized approximate value iteration schemes

URL: View paper

Brief Assessment

Smooth Regularized Convergence[69] focuses on smooth regularized approximate value iteration schemes with entropy regularization and value smoothing, not gradient-aware Bellman operators or Sobolev temporal difference methods.

8. Iterated -Network: Beyond One-Step Bellman Updates in Deep Reinforcement Learning

URL: View paper

Brief Assessment

Iterated Network[62] focuses on learning multiple consecutive Bellman updates simultaneously in standard value-based RL, not on gradient-aware operators or contraction properties of Sobolev Bellman operators with Wasserstein or max-sliced MMD metrics.

9. Bridging hamilton-jacobi safety analysis and reinforcement learning

URL: View paper

Brief Assessment

Hamilton Jacobi Safety[63] focuses on safety analysis through Hamilton-Jacobi reachability with a time-discounted modification for minimum payoff problems, not on gradient-aware Bellman operators or Sobolev temporal difference methods in reinforcement learning.

10. Distributional reinforcement learning via moment matching

URL: View paper

Brief Assessment

Moment Matching DRL[57] focuses on standard distributional RL with MMD metrics for scalar returns, not gradient-aware Bellman operators or Sobolev methods.

Contribution 3: Max-sliced Maximum Mean Discrepancy metric

Description: The authors propose a tractable distributional metric called max-sliced MMD that maintains contraction properties while being computationally feasible for training distributional critics. This metric addresses the computational challenges of using Wasserstein distances in practice.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Distributional reinforcement learning via sinkhorn iterations

URL: View paper

Brief Assessment

Sinkhorn Iterations DRL[60] focuses on using Sinkhorn iterations for distributional RL, not on developing max-sliced MMD metrics for temporal difference learning. The candidate's approach is fundamentally different from the original paper's max-sliced MMD contribution.

2. Distributional Reinforcement Learning with Maximum Mean Discrepancy

URL: View paper

Brief Assessment

Maximum Mean Discrepancy[58] uses standard MMD with various kernels for distributional RL, not the max-sliced variant. The candidate focuses on moment matching via MMD but does not propose or analyze the max-sliced construction that is central to the original paper's tractable metric contribution.

3. Utilizing Maximum Mean Discrepancy Barycenter for Propagating the Uncertainty of Value Functions in Reinforcement Learning

URL: View paper

Brief Assessment

MMD Barycenter Uncertainty[59] focuses on using MMD barycenter for temporal difference updates in value-based Q-learning, not on developing max-sliced MMD as a tractable distributional metric for training distributional critics in actor-critic settings.

4. Distributional bellman operators over mean embeddings

URL: View paper

Brief Assessment

Bellman Mean Embeddings[55] focuses on mean embedding sketches for distributional RL using standard MMD and max-sliced variants, but does not address the specific computational challenges of using Wasserstein distances that the original paper's max-sliced MMD metric was designed to solve. The candidate's framework operates in a different technical context (sketch-based distributional RL) than the original's gradient-aware temporal difference learning.

5. A distributional analogue to the successor representation

URL: View paper

Brief Assessment

Distributional Successor Representation[53] uses MMD for learning distributional models in a different context (successor measures for zero-shot policy evaluation), not for distributional temporal difference learning with value gradients as in the original paper.

6. Foundations of multivariate distributional reinforcement learning

URL: View paper

Brief Assessment

Multivariate Distributional RL[52] uses standard MMD for multivariate distributions but does not propose or analyze the max-sliced MMD variant with contraction properties for gradient-aware temporal difference learning.

7. Distributional reinforcement learning via moment matching

URL: View paper

Brief Assessment

Moment Matching DRL[57] uses standard MMD with Gaussian kernels for distributional TD, not max-sliced variants for gradient distributions.

8. From Wasserstein to Maximum Mean Discrepancy Barycenters: A Novel Framework for Uncertainty Propagation in Model-Free Reinforcement Learning

URL: View paper

Brief Assessment

Wasserstein MMD Barycenters[51] focuses on uncertainty propagation in model-free RL using MMD-based distributional methods, but does not propose the max-sliced MMD metric. The original paper's max-sliced MMD is a novel tractable metric designed specifically for training distributional critics with contraction properties in gradient-aware RL.

9. Distributional reinforcement learning with regularized wasserstein loss

URL: View paper

Brief Assessment

Regularized Wasserstein Loss[54] focuses on Sinkhorn divergence for distributional RL, not max-sliced MMD. The candidate uses standard MMD with multiquadric kernels rather than proposing max-sliced variants as a tractable metric for distributional critics.

10. Distributional reinforcement learning for multi-dimensional reward functions

URL: View paper

Brief Assessment

Multi Dimensional Reward[56] uses standard MMD with Gaussian kernels for multi-dimensional distributional RL, not the max-sliced variant. The candidate focuses on modeling joint return distributions from multiple reward sources rather than developing novel MMD-based metrics for temporal difference learning.

Appendix: Text Similarity Detection

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

References

- [0] Distributional value gradients for stochastic environments View paper
- [1] Value-distributional model-based reinforcement learning View paper
- [2] Optimizing return distributions with distributional dynamic programming View paper
- [3] Distributional Meta-Gradient Reinforcement Learning View paper
- [4] Generalized Advantage Estimation for Distributional Policy Gradients View paper
- [5] Distributed Distributional Deterministic Policy Gradients View paper
- [6] Estimating or Propagating Gradients Through Stochastic Neurons View paper
- [7] How Does Return Distribution in Distributional Reinforcement Learning Help Optimization? View paper
- [8] Distributional policy gradient with distributional value function View paper
- [9] Stochastic optimization methods for policy evaluation in reinforcement learning View paper
- [10] Stochastically Dominant Distributional Reinforcement Learning View paper
- [11] Sample-based distributional policy gradient View paper
- [12] Distributional soft actor-critic: Off-policy reinforcement learning for addressing value estimation errors View paper
- [13] On the Evolution of Return Distributions in Continuous-Time Reinforcement Learning View paper
- [14] Safety-Optimized Fast Charging of Lithium-ion Battery Based on Distributional SAC-Conservative Augmented Lagrangian SDRL Algorithm View paper
- [15] Distributional constrained reinforcement learning for supply chain optimization View paper
- [16] Gradient estimation in model-based reinforcement learning: a study on linear quadratic environments View paper
- [17] Distributional pareto-optimal multi-objective reinforcement learning View paper
- [18] Bayesian distributional policy gradients View paper
- [19] From deterministic to stochastic: an interpretable stochastic model-free reinforcement learning framework for portfolio optimization View paper
- [20] Nonparametric return distribution approximation for reinforcement learning View paper
- [21] Conservative offline distributional reinforcement learning View paper
- [22] Flow Models for Unbounded and Geometry-Aware Distributional Reinforcement Learning View paper
- [23] Stochastic variance-reduced policy gradient View paper
- [24] Off-Policy Conservative Distributional Reinforcement Learning With Safety Constraints View paper
- [25] Improving stochastic policy gradients in continuous control with deep reinforcement learning using the beta distribution View paper
- [26] Intrinsic Benefits of Categorical Distributional Loss: Uncertainty-aware Regularized Exploration in Reinforcement Learning View paper
- [27] Reinforcement Learning for Risk-Aware Portfolio Optimization View paper
- [28] Policy optimization in a noisy neighborhood: On return landscapes in continuous control View paper
- [29] IGN : Implicit Generative Networks View paper
- [30] Diverse Projection Ensembles for Distributional Reinforcement Learning View paper
- [31] Policy Evaluation in Distributional LQR View paper
- [32] Distributional policy optimization: An alternative approach for continuous control View paper
- [33] Optimizing Return Policies: A Bayesian Learning Approach View paper
- [34] DVPO: Distributional Value Modeling-based Policy Optimization for LLM Post-Training View paper
- [35] PG-Rainbow: Using Distributional Reinforcement Learning in Policy Gradient Methods View paper
- [36] Moments Matter:Stabilizing Policy Optimization using Return Distributions View paper
- [37] Total stochastic gradient algorithms and applications in reinforcement learning View paper
- [38] APMPO: A Portfolio Management Policy Optimization Framework with Adaptive Reinforcement Learning Algorithm View paper
- [39] Advances in Distributional Reinforcement Learning: Bridging Theory with Algorithmic Practice View paper
- [40] RegimeFolio: A Regime Aware ML System for Sectoral Portfolio Optimization in Dynamic Markets View paper
- [41] Offline Bayesian Aleatoric and Epistemic Uncertainty Quantification and Posterior Value Optimisation in Finite-State MDPs View paper
- [42] Convergence Theorems for Entropy-Regularized and Distributional Reinforcement Learning View paper
- [43] PACER: A Fully Push-forward-based Distributional Reinforcement Learning Algorithm View paper
- [44] A Simple Mixture Policy Parameterization for Improving Sample Efficiency of CVaR Optimization View paper
- [45] Search and return model for stochastic path integrators. View paper
- [46] Bag of Policies for Distributional Deep Exploration View paper

- [47] Bayesian policy gradient algorithms View paper
- [48] MERL: Multi-Head Reinforcement Learning View paper
- [49] Decision-making with Speculative Opponent Models View paper
- [50] Zero-touch Continuous Network Slicing Control via Scalable Actor-Critic Learning View paper
- [51] From Wasserstein to Maximum Mean Discrepancy Barycenters: A Novel Framework for Uncertainty Propagation in Model-Free Reinforcement Learning View paper
- [52] Foundations of multivariate distributional reinforcement learning View paper
- [53] A distributional analogue to the successor representation View paper
- [54] Distributional reinforcement learning with regularized wasserstein loss View paper
- [55] Distributional bellman operators over mean embeddings View paper
- [56] Distributional reinforcement learning for multi-dimensional reward functions View paper
- [57] Distributional reinforcement learning via moment matching View paper
- [58] Distributional Reinforcement Learning with Maximum Mean Discrepancy View paper
- [59] Utilizing Maximum Mean Discrepancy Barycenter for Propagating the Uncertainty of Value Functions in Reinforcement Learning View paper
- [60] Distributional reinforcement learning via sinkhorn iterations View paper
- [61] Distributional reinforcement learning View paper
- [62] Iterated -Network: Beyond One-Step Bellman Updates in Deep Reinforcement Learning View paper
- [63] Bridging hamilton-jacobi safety analysis and reinforcement learning View paper
- [64] Multi-Bellman operator for convergence of -learning with linear function approximation View paper
- [65] Implicit Constraint-Aware Off-Policy Correction for Offline Reinforcement Learning View paper
- [66] Exploring the Training Robustness of Distributional Reinforcement Learning Against Noisy State Observations View paper
- [67] Stability and Generalization for Bellman Residuals View paper
- [68] Robust Reinforcement Learning for Continuous Control with Model Misspecification View paper
- [69] On the convergence of smooth regularized approximate value iteration schemes View paper
- [70] Using Exact Models to Analyze Policy Gradient Algorithms View paper
- [71] Beyond Marginals: Capturing Correlated Returns through Joint Distributional Reinforcement Learning View paper