

Scaling Deep Reinforcement Learning and Stochastic Modeling: From Simulation to Real-World Deployment

Authors:

Ankit Anil Chaudhari, Dianzhao Li, Ostap Okhrin

The "Reinforcement Learning Dresden" project leverages High-Performance Computing (HPC) to address complex, data-driven optimization challenges across maritime navigation, automotive autonomy, and finance. Deploying advanced machine learning models into the real world requires overcoming severe environmental, computational, and data-processing bottlenecks.

In maritime autonomy, inland waterways introduce non-linear dynamics like shallow water effects and strong currents, causing standard RL agents to diverge. We resolved this using spatial-temporal recurrent networks and a 2-level hierarchical RL architecture to mitigate multi-objective conflicts between path following and collision avoidance. Furthermore, extracting clean trajectories from noisy, high-granularity AIS big data required a heavily parallelized preprocessing pipeline utilizing multi-core CPU architectures.

Within automotive autonomy, a primary challenge is bridging the sim2real gap caused by sensor noise, latency, and visual domain shifts, alongside overcoming the robotic driving profiles of pure RL. We developed a platform-agnostic sim2real framework and a modified DDPG model grounded in empirical driving data. Additionally, calibrating vehicle-following models for chaotic, non-lane-based traffic required processing the massive trajectory dataset. To ensure safety and explainability, we also worked on the integration of Large Language Models (LLMs) to interrogate and validate the decision-making of our autonomous agents. Executing continuous RL training loops coupled with high-fidelity CARLA rendering created severe CPU-to-GPU data starvation bottlenecks. Simultaneously, structural processing of terabytes of our traffic data (MiTra) introduced extreme memory and I/O limits, necessitating massive shared-memory CPU configurations and high-throughput disk access.

In finance, modeling high-dimensional functional data and correcting parameter biases in stochastic processes required generating massive synthetic datasets to prevent small-sample bias. We deployed artificial neural networks trained via millions of parallel Monte Carlo iterations, incurring immense memory overhead. Ultimately, overcoming these intersecting challenges demonstrates that tightly coupled HPC allocations are a fundamental prerequisite for advancing reliable, deployable decision science.