

Neural Inventory Control in Networks via Hindsight Differentiable Policy Optimization

Tuesday, June 18, 2024 2:00 PM (30 minutes)

Inventory management offers unique opportunities for reliably evaluating and applying deep reinforcement learning (DRL). We introduce Hindsight Differentiable Policy Optimization (HDPO), facilitating direct optimization of a policy's hindsight performance using stochastic gradient descent. HDPO leverages two key elements: (i) an ability to backtest any policy's performance on a sample of historical "noise" traces, and (ii) the differentiability of the total cost incurred on any subsample with respect to policy parameters. We assess this approach in four problem classes where we can benchmark performance against the true optimum. Our algorithms consistently achieve near-optimal performance across all these classes, even when dealing with up to 60-dimensional raw state vectors. Moreover, we propose a natural neural network architecture to address problems with weak (or aggregate) coupling constraints between locations in an inventory network. This architecture utilizes weight duplication for "sibling" locations and state summarization. We demonstrate empirically that this design significantly enhances sample efficiency and provide justification through an asymptotic performance guarantee. Lastly, we assess our approach in a setting that incorporates real sales data from a retailer, demonstrating its substantial superiority over predict-then-optimize strategies.

Primary authors: RUSSO, Dan (Columbia Business School (Decision, Risk and Operations division)); ALVO, Matias (Columbia Business School (Decision, Risk and Operations division)); KANORIA, Yash (Columbia Business School (Decision, Risk and Operations division))

Presenter: ALVO, Matias (Columbia Business School (Decision, Risk and Operations division))

Session Classification: Parallel session: Reinforcement learning in MDPs with large state spaces