Session 8: New OM Research Areas / reinforcement learning

Chair: Maxi Udenio

Arnoud Wellens (KU Leuven) - The Impact of Improving Demand Forecasts with Machine Learning Methods on a Retailer's Inventory Control

Despite the recent outperformance of forecasting with machine learning (ML) based methods, simple statistical forecasting techniques remain the standard approach in retail. We propose and validate a simple-to-use decision-tree framework with the goal to democratize ML forecasting for large-scale retail applications and investigate the impact of these improved forecasts on a retailer's inventory control. Based on a test set of 4,548 products of a leading Belgian retailer and a variety of external variables (e.g., promotions and national events), our method improves the forecast accuracy of commonly used statistical methods up to 25%. Using simulation, we investigate whether this forecast superiority translates to higher service levels, lower inventory costs and improvements regarding various bullwhip metrics. Hence, we demonstrate the trade-off between investing in more sophisticated forecasting methods and the operational benefits they yield.

Tarkan Temizoz (Eindhoven University of Technology) - Deep Controlled Learning for Large Scaled Inventory Control

Highly stochastic large-scaled Markov Decision Processes (MDPs) suffer from finding good value approximations and require efficiently allocating computational resources. The supply chain and inventory control models are examples of such MDPs, where the number of SKUs can be arbitrarily large, and the costs have high variance.

Deep Controlled Learning (DCL) is a supervised learning approach that maps states to actions, using neural networks to represent the policies. It originates from the following important fact of the policy iteration: the greedy policy improvement step does not require an ex-act calculation of the value functions; it suffices to determine which action has the highest Q-value. The estimates of the Q-values can be collected using rollouts, which brings about two sources of errors: the horizon length of each rollout and the number of rollouts allocated to each action.

DCL reduces these errors in two ways. It calculates the unbiased estimates of Q-values and introduces a variance reduction scheme that incorporates common random numbers (CRN) that produce the same events within the same rollout for each action. A significant variance reduction leads to selecting the best action with more confi-dence, hence avoiding using more rollouts. Moreover, DCL also uses two ranking-and-selection approaches for rollout allocation to choose the best action efficiently. Both methods eliminate actions that no longer promise the best one. The conservative approach provides confidence bounds with a prespecified probability, guaranteeing the creation of stable state-action pairs. The greedy approach discards actions when a certain threshold is exceeded.

We test DCL in 2D online bin packing and lost sales problems. In all experiments, we show that positive correlations are induced between actions, and both rollout allocation algorithms successfully avoid wasting resources on bad actions. Moreover, the generated neural network policies perform near-optimal while outperforming the heuristic benchmarks.

Given these results and the algorithm's applicability in modern hardware parallelization, we state that DCL is a promising approach to tackle large-scaled stochastic dynamic models in inventory control.

Robert van Steenbergen (University of Twente) - Reinforcement Learning for UAV-Aided Humanitarian Logistics

We introduce a vehicle routing problem with time windows in a humanitarian setting. We present an MDP formulation about this humanitarian problem including trucks and drones, and travel time uncertainty. Several heuristic methods are developed to solve the problem sequentially, including approximate value iteration and approximate policy iteration. We show that sequential decision methods improve results compared to the execution of deterministic plans. With different vehicle types and online decisions, it can anticipate on both earliness and lateness that unfolds during the execution.