REINFORCEMENT LEARNING-BASED SIMULATION OPTIMIZATION FOR AN INTEGRATED MANUFACTURING-WAREHOUSE SYSTEM: A TWO-STAGE APPROACH

Maryam Hosseini, Systems Science and Industrial Engineering, Ph.D. Candidate

Time: 2:30 pm - 4:30 pm on 11/19/2024

Zoom Link: https://binghamton.zoom.us/j/98851948680

Abstract

This study addresses the complexities of supply chain management in integrated warehouse-manufacturing systems, where the accuracy of demand forecasting and effective inventory optimization are critical to efficient operations. We developed a two-stage approach to address these issues. In the first stage, we focused on time series forecasting using both traditional statistical methods and advanced deep learning models. Through preprocessing steps, including outlier removal and feature engineering, we observed that the forecasting accuracy improved even in the presence of limited external features. Specifically, deep learning models Bi-GRU and Bi-LSTM demonstrated notable accuracy improvements, achieving R-squared values of 38% and 26% on the test set, respectively.

In the second stage, we created a Material Requirements Planning model and implemented it in FlexSim's OptQuest interface, running simulations across 25 different scenarios, each over 25 replications. The analysis reveals strategic variations in order quantities, costs, inventory, and stay times that underscore key trade-offs in managing an integrated warehouse-manufacturing system. Initial and later scenarios favor conservative ordering, with quantities ranging between 200-400 units, keeping total costs low at around 1000-2000 units. In contrast, mid-scenarios with higher order quantities (600-1200 units) result in increased costs, reaching 5000-6500 units. These simulations provided a range of ordering frequencies and costs, guiding us in finding cost-minimizing strategies.

To further enhance decision-making, we designed two reward functions and integrated a reinforcement learning agent into the FlexSim simulation environment. Using a Taguchi experimental design, we evaluated the agent's learning behavior under varying penalty schemes. These two reward functions, testing alpha (α) and gamma (γ) at levels of [0.0001, 0.005, 0.01] to allow efficient testing of all combinations, minimizing the number of experiments needed while examining the effects of α and γ on reward performance. The first reward function penalized overdue items on shelves with alpha (α), while the second reward function penalized large discrepancies between target and actual. Initial training shows low, stable rewards during random exploration, followed by a steady increase in average rewards as learning progresses. Configurations with higher α and γ values yield faster reward increases, indicating quicker learning, while lower values show slower, more gradual convergence. Although both reward functions share underlying similarities, the variations in α and γ significantly influence reward growth stability and the exploration-exploitation balance in the model. Both reward functions revealed fluctuations in the agent's performance, suggesting additional factors may influence stability.

This study contributes both practical insights for inventory management and a novel two-stage framework for supply chain optimization. Our approach highlights the importance of data preparation, advanced forecasting, and RL-driven optimization for businesses aiming to enhance operational efficiency in complex supply chain systems.