Advancing Supply Chain Inventory Management through Mathematical and Deep Reinforcement Learning Approaches

Francesco Stranieri

The increasing complexity of modern supply chains, driven by industry 4.0 principles and advances in artificial intelligence (AI), has created a growing need for adaptive, data-driven supply chain management strategies. While simple and widely recognized, traditional inventory policies often struggle to address the multidimensional, uncertain, and dynamic challenges inherent in supply chain inventory management (SCIM). In this context, deep reinforcement learning (DRL) has emerged as a promising alternative capable of directly learning effective policies by interacting with simulated supply chain environments. However, its practical adoption remains constrained by several barriers, including computational complexity, scalability issues, limited real-world data, and interpretability concerns.

This thesis explores the potential of DRL to improve SCIM across diverse structural attributes and operational factors, focusing on three primary research objectives: 1) evaluating the performance of state-of-the-art DRL algorithms in two-echelon systems subjected to stochastic and seasonal demand, varying numbers of product types and warehouses, and different lead times, thus addressing scalability issues; 2) developing a hybrid heuristic that combines DRL with multi-stage (MS) stochastic programming to mitigate the computational complexity associated with the curse of dimensionality, while accounting for production limits; and 3) assessing the applicability and interpretability of DRL in a real-world pharmaceutical supply chain characterized by perishability, production yields, batch limits, lead times, and lost sales under non-stationary demand.

The first study demonstrates that DRL, particularly the proximal policy optimization (PPO) algorithm, consistently outperforms base-stock and (s, Q)-policies in capacitated two-echelon systems. By leveraging continuous action spaces and designing a balanced allocation rule, PPO effectively handles increasing complexity, scales across large state-action spaces, and achieves training times that are comparable to traditional inventory policies. The second study introduces a hybrid heuristic in which DRL determines long-horizon production planning while MS stochastic programming optimizes short-term logistics decisions. This combined approach results

in robust, computationally tractable solutions that outperform both standalone methods. The third study focuses on a pharmaceutical supply chain, where we evaluate PPO with the order-up-to policy and one of its variants based on projected inventory levels, for which we derive and validate bounds-based procedures to optimize their parameters. All three policies surpass human-driven policies in minimizing costs under complex conditions, providing managerial insights, and interpreting the impact of costs and lost sales in relation to ethical and legal constraints.

These findings collectively highlight the transformative potential of DRL, demonstrating its ability to tackle uncertain and dynamic SCIM problems, exploit high-dimensional state and action spaces, and outperform traditional inventory policies while also extending and identifying best practices for integrating DRL into practical SCIM decision-making. Furthermore, the open-source SCIMAI-Gym Python library, developed as part of this research, provides a flexible simulation environment that encourages further experimentation, collaboration, and practical implementation.

In conclusion, this thesis lays a solid foundation for future research. Potential directions include implementing action masking and reward shaping techniques, integrating large language models, leveraging advanced computational frameworks, and modeling even more complex supply chain configurations. Ultimately, these contributions advance the field toward more robust, efficient, and adaptive supply chains that align with the principles of AI and industry 4.0.