Proceedings of the 2022 Winter Simulation Conference
B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C.G. Corlu, L.H. Lee, E.P. Chew, T. Roeder, and P. Lendermann, eds.

APPLICATION OF DEEP REINFORCEMENT LEARNING FOR PLANNING OF VENDOR-MANAGED INVENTORY FOR SEMICONDUCTORS

Fazail Ahmad
Marco Ratusny
Hans Ehm

Infineon Technologies AG
Am Campeon 1 - 15
Neubiberg, 85579, GERMANY

Santiago Nieto-Isaza

Department of Industrial Engineering
Universidad del Norte
Kilómetro 5 Vía, Puerto Colombia
Barranquilla, COLOMBIA

ABSTRACT

Previous research has shown that Deep Reinforcement Learning (DRL) can be applied to the planning of Vendor-Managed Inventory (VMI) for semiconductors. This study extends the research in this direction by overcoming the limitations of existing work and evaluating the approach through a case study at a semiconductor supplier. The results strengthen the potential of DRL for VMI planning. Moreover, the developed framework can be deployed with other advanced DRL algorithms.

1 INTRODUCTION

Compared to other supply chains, the semiconductor industry suffers from more intense and longer-lasting disturbances in demand and supply due to its unique feature of long production lead times. This challenge has been laid bare by the global shortage of microchips in the wake of the Covid-19 outbreak. One strategy to mitigate these disturbances is Vendor-Managed Inventory (VMI), which works like a minibar in a hotel room: suppliers maintain a suitable level of inventory, and customers can pull from it as desired. In the semiconductor industry, suppliers rely on customer demand forecasts for VMI planning due to long lead times. Therefore, collaboration between the two sides is necessary for functioning of VMI. A step towards enhancing customers' trust in VMI is optimizing VMI planning at the suppliers' side, which is largely human-based. A potential method to improve VMI planning is reinforcement learning (RL), since RL can be applied on the VMI model and has seen tremendous development in recent years. Afridi et al. (2020) demonstrate this potential by using Deep Q-Network (DQN), a deep reinforcement learning (DRL) algorithm. Nonetheless, Afridi et al. (2020)'s work has limitations: Their approach requires long training times, generates an increased number of replenishments and lacks a framework for hyperparameter tuning. This work builds on Afridi et al. (2020) by developing a heuristic for consolidating the replenishments and an approach for tuning the hyperparameters. It also compares DRL with two myopic policies and presents an industrial case study.

2 METHODOLOGY

For the application of DRL, the definitions of state, action and reward of VMI from Afridi et al. (2020) are employed. Unlike Afridi et al. (2020), the DQN algorithm is programmed using PyTorch and trained using an NVIDIA Tesla P4 GPU to reduce the training duration. The trained algorithm is used in the testing phase to make the best possible replenishment decisions. A simple heuristic that assumes that some days in a week are restricted for replenishment is applied to control the increased number of replenishments. In the first step, the replenishments are planned by DRL, ignoring the restricted days. Next, the replenishments are consolidated based on the heuristic that holds shipments on restricted days and dispatches them combined on the next available unrestricted day.

Following Gijsbrechts et al. (2022), an approach to tune the hyperparameters of DQN is designed. This approach involves selecting some hyperparameters to tune and defining a suitable domain for each of them. Then, these hyperparameters are tuned by running several training trials of DQN with the aim to optimize the action-values. For each trial, hyperparameter values are chosen by a Bayesian optimization algorithm called Tree-Structured Parzen Estimator by Bergstra et al. (2011). This algorithm selects the values using a simulated model of the action-values, which is sequentially improved during an experiment. Finally, the DRL approach is compared with two myopic policies for VMI planning. The first policy replenishes VMI to the mean level if the inventory level is forecasted to be equal to or below it. The second policy is a variation of the first one and has the VMI minimum limit as the reorder point.

3 CASE STUDY AND RESULTS

The methodology is applied to three products (A, B, and C) of a semiconductor supplier. Product A is also used by Afridi et al. (2020), and following them, a lead time of 4 weeks is assumed. A horizon of 847 and 280 days is employed for training and testing, respectively. The performance is analyzed using percentage no-violation (PNV), which is the percentage of days when VMI is within the pre-defined limits.

This implementation reduces training duration to around 25 minutes per run, while Afridi et al. (2020) need almost 24 hours. In terms of performance, this study outperforms not only Afridi et al. (2020) but also the Existing Practices (EP) at the supplier for A. Using the heuristic for consolidating replenishments, this work achieves comparable PNV (59%) with less than half the shipments as Afridi et al. (2020). DRL also gives improved performance than EP with B and C. For C, DRL gives PNV of 60%, which is 25% better than EP. However, DRL needs more replenishments due to this product's exceptionally high demand volumes. Hyperparameter tuning helps improve the performance of DRL, for example, it increases PNV from 55% to 59% for B. Regarding comparison with the myopic policies, DRL performs inferior for A but outperforms the policies for the other two products. With A, which has stable demand, myopic policies achieve PNV of above 70% with fewer replenishments than DRL, which gives PNV of 59%. However, for B and C with volatile demand, DRL surpasses the first policy in terms of both PNV (59% vs. 55% for B and 60% vs. 40% for C) and replenishments (84 vs. 133 for B and 120 vs. 195 for C). DRL also outperforms the second policy in terms of PNV (59% vs. 39% for B and 60% vs. 51% for C).

4 CONCLUSION AND MANAGERIAL INSIGHTS

The findings show that DRL surpasses human-based practices for VMI planning. Nonetheless, myopic policies might be more suitable for some products. Hence, a comparison should be performed for a product before implementation of DRL. The developed framework should work with other DRL algorithms and help exploit the ever-growing realm of RL methods. With a VMI measurement method, DRL can help enhance collaboration between suppliers and customers in the semiconductor industry.

REFERENCES

Afridi, M. T., S. Nieto-Isaza, H. Ehm, T. Ponsignon, and A. Hamed. 2020. "A Deep Reinforcement Learning Approach for Optimal Replenishment Policy in a Vendor Managed Inventory Setting for Semiconductors". In *Proceedings of the 2020 Winter Simulation Conference*, edited by K-H. G. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R.Thiesing, 1753-1764. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Bergstra, J., R. Bardenet, Y. Bengio, and B. Kègl. 2011. "Algorithms for Hyper-Parameter Optimization". In *Proceedings of the 24th International Conference on Neural Information Processing Systems*, edited by J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, and K. Q. Weinberger, 2546-2554. Red Hook, New York: Curran Associates, Inc.

Gijsbrechts, J., R. N. Boute, J. A. Van Mieghem, and D. J. Zhang. 2022. "Can Deep Reinforcement Learning Improve Inventory Management? Performance on Lost Sales, Dual Sourcing, and Multi-Echelon Problems". *Manufacturing & Service Operations Management* 24(3):1349-1368.