Industrializing Deep Reinforcement Learning for ASML's Service Network

J.F. van der Haar

Problem Definition

ASML's customer service network aims to ensure timely availability of the materials needed for machine maintenance. Most operational inventory control decisions in this network are automated by the internally developed NORA algorithm, which is known to have good yet suboptimal performance. More advanced methods exist in literature, but they struggle with scalability. Our goal is to propose a method that can outperform NORA while maintaining tractability.

Methodology

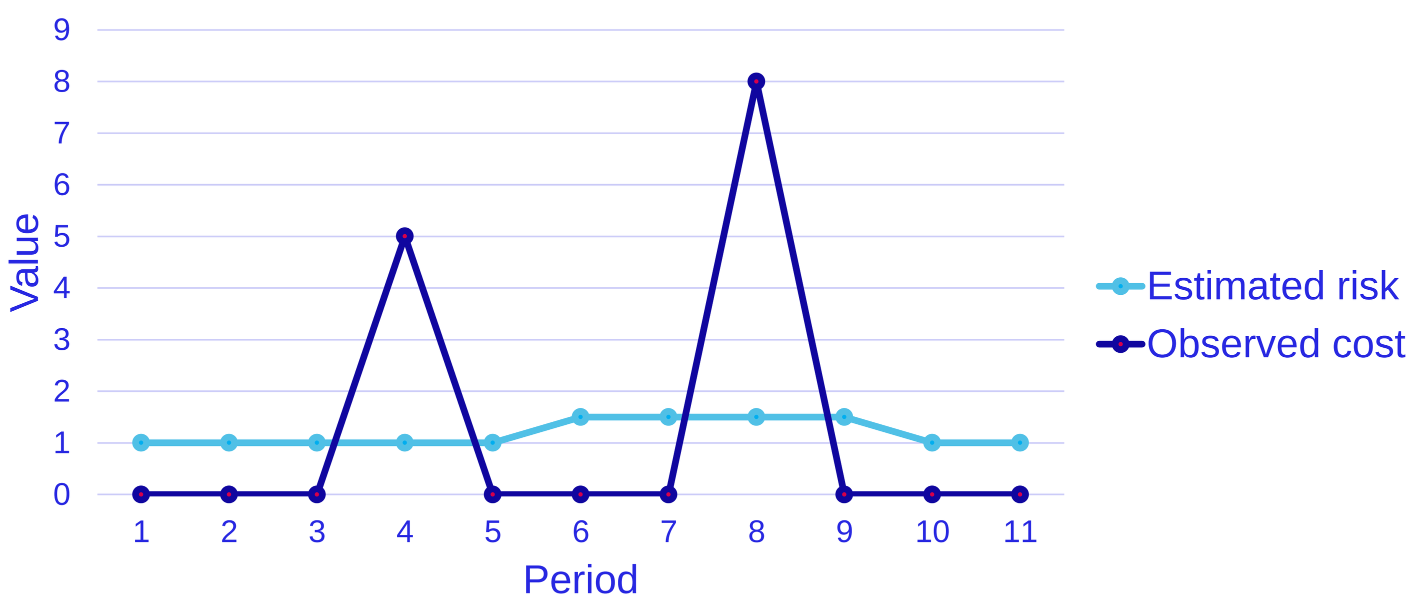
Deep reinforcement learning (DRL) is an approximate method that can explicitly account for future decisions and events, which makes it a potent technique for a wide range of inventory management problems. Moreover, it can take operational decisions in near real-time. We therefore choose to address the problem using DRL. However, research has thus far failed to train DRL models for realistic problems of comparable size and complexity.

Scaling DRL Model Training

We propose three techniques to scale DRL model training, namely reward shaping, action splitting and global models.

Reward Shaping

DRL models learn by optimizing a reward signal, which typically means it tries to minimize the incurred cost. We propose to use an estimate of the expected costs instead to stabilize the reward signal.

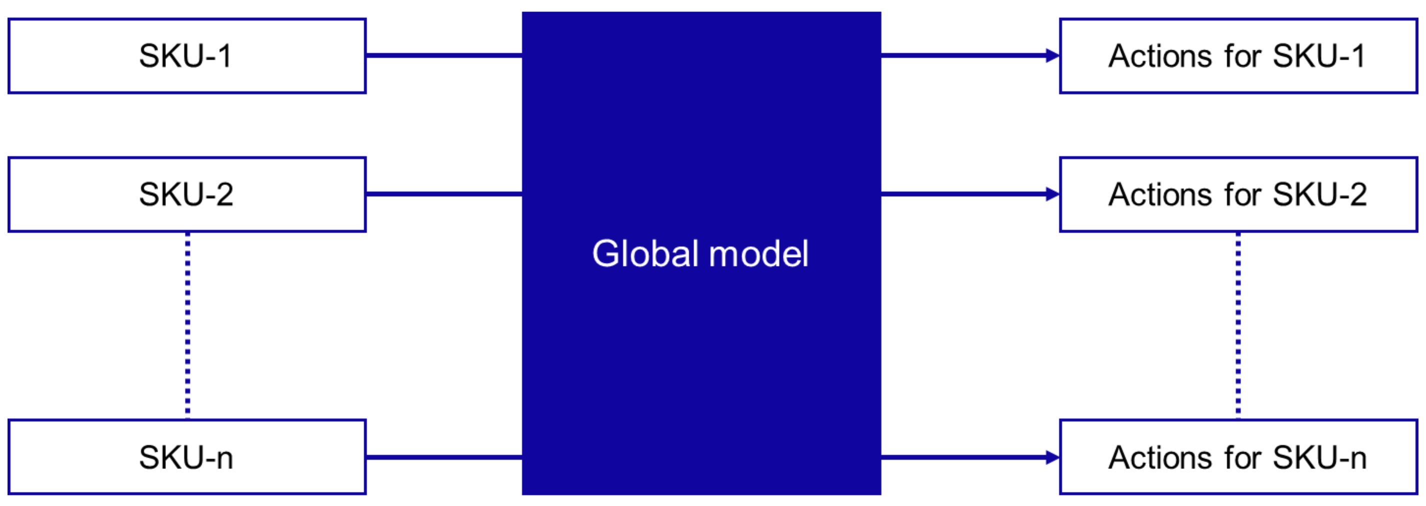


Action Splitting

We simplify decisions for the DRL model by only letting it take a decision on one shipment at a time. For each shipment, it first chooses a receiving location, after which it chooses what location to source from.

Global Model

We train a single model for all SKUs in the entire network, whereby this model takes decisions for each SKU separately



Scaling Study

The effectiveness of our proposed scaling method is assessed by training it on a synthetic dataset with 60 locations and 10,000 SKUs, split between low demand SKUs (≤0.0005 location/day) and high demand SKUs (>0.0005 location/day). We perform a simulation run on 100 randomly selected SKUs from both categories and obtain the following results:

- 2.24% cost savings on low-demand SKUs compared to NORA.
- 5.02% cost savings on high-demand SKUs compared to NORA.

Benchmark Study

To examine whether our method can also be used to address a problem with the amount of complexities present in a real-life service supply chain, we benchmark our method against NORA using simulation runs on more complex real data. We train on all SKUs in the dataset and evaluate on 100 randomly selected SKUs. The results indicate cost savings of 5.14%.

Based on one-sided independent samples t-tests, we find that:

- DRL significantly outperforms NORA in 48% of cases.
- NORA significantly outperforms DRL in 6% of cases.
- Neither significantly outperforms the other for 46% of cases.

Conclusions

- Our scaling study demonstrates that our proposed techniques can be used to train DRL models for realistic problems with a size of 60 locations and 10,000 SKUs.
- Our benchmark study demonstrates that they can also be used to train DRL models for problems of the required complexity.

We hence conclude that our method is able to outperform NORA while maintaining tractability.

Operations Planning, Accounting and Control. Department of Industrial Engineering & Innovation Sciences
Data and Artificial Intelligence. Department of Mathematics & Computer Science