Hannu Rummukainen*, Jukka K. Nurminen

Reinforcement learning for economic lot scheduling

Abstract: We investigated how to apply deep reinforcement learning to discrete-event production control on a stochastic single-machine production scheduling case. We modified the proximal policy optimization (PPO) algorithm for a discrete event model in continuous time, and implemented two state-value approximation methods and control policies, a linear model and a two-layer neural network. Compared to earlier published results on the same case, we improved the average cost rate by 2 %, and moreover, our control policy is entirely learned and is not based on any explicit control heuristics.

Keywords: reinforcement learning, production control, stochastic economic lot scheduling

*Corresponding Author: Hannu Rummukainen: VTT, E-mail: hannu.rummukainen@vtt.fi

Second Author: VTT (currently University of Helsinki), E-mail: jukka.k.nurminen@helsinki.fi

1 Background

Reinforcement learning holds the promise of automatically learning optimal control by trial and error without knowing the system structure. However, applying the approach to systems with multi-dimensional observation spaces is often challenging, and practical applications are rare. Following recent successes in applying deep reinforcement learning to games, we take another look on reinforcement learning as a method for discrete production control.

2 Aims

Our aim is to investigate how to apply a state-of-the-art deep reinforcement learning method in a production control application, what kind of challenges there are, and whether modern methods can improve on earlier results achieved by reinforcement learning.

3 Materials and methods

We performed experiments on a relatively simple stochastic scheduling problem based on research literature. The problem involves scheduling the production of multiple products on a single machine with stochastic demand. Switching the machine from one product to another is both slow and costly, and there are additional costs for holding products in stock, and for unsatisfied customer orders queued in backlog. Although simple, the model is quite challenging to solve, and no closed-form solution is known. In practice production is often controlled by heuristics such as a so-called *base-stock policy*, in which each product is manufactured until its stock level reaches a fixed base-stock level, and then the machine is switched to the next product in a fixed sequence.

We applied an open source implementation of the proximal policy optimization (PPO) algorithm of Schulman et al. [1]. As the lot scheduling model is a discrete event model in continuous time, we modified the PPO algorithm to apply continuous-time discounting. We implemented two different state-value approximation models and stochastic control policies: in the first model, the state-value approximation and the log-probability of an action are linear functions of the observed system state, and in the second model, we use a two-layer neural network.

4 Results

We compared our approach to two base-stock control policies that were published earlier for the same case study: the first policy was a base-stock policy derived by infinitesimal perturbation analysis in [2], and the second policy was a combination of lot-sizing by a base-stock policy and product choice by reinforcement learning [3]. Our model based on a neural network produced a 9 % better average cost rate than the base-stock policy of [2], and a 2 % better average cost rate than the best known earlier solution in [3]. More importantly, our reinforcement learning solution is more general than either of the earlier approaches, since our method is not restricted to policies with pre-specified base-stock parameters, but learns a general policy for every control decision.

Finding the best reinforcement learning policy required a lot of experimentation with different approximation functions and algorithm control parameters, and even then the algorithm required very long training runs to

converge. We conclude that reinforcement learning is still best suited to research problems where no satisfactory control policy is known.

References

[1] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O., Proximal policy optimization algorithms, In: 5th International Conference on Learning Representations, 2017

[2] Anupindi, R., Tayur, S., Managing stochastic multiproduct systems: model, measures, and analysis, Oper. Res., 1998, 46, S98–S111

[3] Paternina-Arboleda, C.D., Das, T.K., A multi-agent reinforcement learning approach to obtaining dynamic control policies for stochastic lot scheduling problem, Simul. Model. Pract. Th., 2005, 13, 389–406