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.

EVALUATING SCHEDULING PERIOD IN TRAINING DATA COLLECTION FOR SITUATION AWARE DISPATCHING

Chew Wye Chan
Boon Ping Gan

D-SIMLAB Technologies Pte Ltd
8 Jurong Town Hall Road, #23-05
Singapore, 609434, SINGAPORE

Wentong Cai

School of Computer Science and Engineering
Nanyang Technological University
Singapore, 639798, SINGAPORE

ABSTRACT

Earlier studies have shown that adapting dispatch rules to manufacturing situations improves factory performance. A trained machine learning model has been used to learn the relationship between manufacturing situations and dispatch rules. In order to generate training data for the model, a multi-pass simulation technique has been used to evaluate candidate dispatch rules under certain manufacturing situations. The candidate dispatch rule is applied within a fixed scheduling period, such as daily or weekly, to evaluate the factory's performance in the scheduling period. However, the scheduling period chosen should be able to differentiate the impact on the factory's performance of candidate dispatch rules. A short scheduling period will result in redundant training data, but a long scheduling period will result in insufficient training data. In this study, we evaluate the effect of the scheduling period on the accuracy of the trained machine learning model.

1 INTRODUCTION

In a complex manufacturing environment such as the semiconductor industry, scheduling is usually done using dispatching rules. Dispatching rules are used to assign a priority ranking to the lot in the queue of a machine. The lot will be dispatched to the available machine in sequence based on the priority ranking. However, it is concluded that there are no single dispatch rules that consistently generate better factory performance (Shiue and Su 2003). It is common to use discrete event simulation to evaluate different candidate dispatch rules to find the best dispatch rule. In practice, a dispatch decision is required in near real-time to avoid the machine being idle while waiting for the dispatch decision. On the other hand, machine learning techniques have gained popularity in learning the best dispatch rules for factory performance and are used to adapt the dispatch rules (Shiue and Su 2003).

Training data is needed to train a supervised machine learning model that adapts dispatch rules to manufacturing situations. A triplet of $\{\mathbf{P}, \mathbf{S}, \mathbf{D}\}$ was proposed to represent the training data (Shiue and Su 2003). P represents the user-defined factory performance, S represents the features of the manufacturing situations, and D represents the dispatch rules. In order to generate the training data, a multi-pass simulation technique (Wu and Wysk 1989) is used to simulate different candidate dispatch rules for the same scheduling period. The manufacturing situations' system state is recorded at the start of the scheduling period. At the end of the scheduling period, the dispatch rule that resulted in the best factory performance for the scheduling period is recorded. The process will be repeated for the next scheduling period by simulating the previous scheduling period with the recorded best dispatch rule.

The limitation of a multi-pass simulation is the need to define a reasonable scheduling period to evaluate the performance of a dispatch rule. Depending on the simulated model used, a scheduling period can be defined as a factor of total process time, a constant of 2500, 5000, or 10000 minutes (Wu and Wysk 1989; Shiue et al. 2020). As training data for the supervised machine learning is generated via multi-pass simulation, the training data's quality would impact the prediction accuracy of the supervised machine learning. A similar manufacturing situations will be recorded when the scheduling period is too short and causes data redundancy and negatively impact the machine learning model. The objective of the study is to evaluate the effect of the scheduling period on the accuracy of the trained machine learning model.

2 SCHEDULING PERIOD AND TRAINING DATA GENERATION

In an example of multi-pass simulation with two candidate dispatch rules (d_1 ,d_2) and two multi-pass periods with a constant scheduling period ($t_1 - t_0$,$t_2 - t_1$). At time t_0, the manufacturing situations are recorded. Candidate dispatch rules (d_1 ,d_2) are applied to the entire factory for the scheduling period t_0 to t_1 on the same manufacturing situations at t_0. At the end of the scheduling period, the target factor performance, ex: average cycle time, is used to decide the best dispatch rule for the scheduling period t_0 to t_1. Then the manufacturing situations are recorded at t_1 by simulating the best dispatch rule for the scheduling period t_0 to t_1. The exact process of simulating candidate dispatch rules will be performed for the scheduling period t_1 to t_2 . At the end of this multi-pass simulation example, two training data will be collected with manufacturing situations at the start of the scheduling period and the best dispatch rule to be applied for the scheduling period.

There might exist situations where the impact of factory performance between the best and worst performing dispatch rule are marginal. As we are using multi-pass simulation to generate training data for the machine learning model, a candidate dispatch rule that results in marginal factory performance difference should be omitted from the training data as not conclusive. These training data that are not conclusive will increase noise and random fluctuations of the machine learning model, negatively affecting the generalization ability and lowering the prediction accuracy of the model.

The length of the scheduling period would impact the occurrence of marginal factory performance between candidate dispatch rules. When a short scheduling period is used, many training data with marginal factory performance will occur. However, a longer simulation period is needed to collect sufficient training data if a longer scheduling period is used. For example, suppose a scheduling period is one month. In that case, a multi-pass simulation with one random seed will generate twelve training data for the simulation period of one year. Depending on the factory behavior, a longer scheduling period might still show marginal factory performance between candidate dispatch rules. Furthermore, generating a factory model for a long simulation period might also be unrealistic as customer demand changes.

Hence, deciding the scheduling period to generate training data for supervised machine learning is important. We propose to conduct experiments on the impact of scheduling period and factory performance between candidate dispatch rules. The outcome of this study is to develop an approach to dynamically decide a scheduling period that could show the distinguishable difference between candidate dispatch rules that improves training data collection quality. Better training data will positively impact the prediction accuracy of the trained machine learning model. In return, this would improve the ability of the model to adapt dispatch rule in a complex manufacturing environment.

REFERENCES

Shiue, Y. R., K. C. Lee, and C. T. Su. 2020. "A Reinforcement Learning Approach to Dynamic Scheduling in a Product-mix Flexibility Environment". *IEEE Access* 8:106542–106553.

Shiue, Y. R., and C. T. Su. 2003. "An Enhanced Knowledge Representation for Decision-tree Based Learning Adaptive Scheduling". *International Journal of Computer Integrated Manufacturing* 16(1):48–60.

Wu, S. D., and R. A. Wysk. 1989. "An Application of Discrete-event Simulation to On-line Control and Scheduling in Flexible Manufacturing". *International Journal of Production Research* 27(9):1603–1623.