

Forschungsberichte aus dem
wbk Institut für Produktionstechnik
Karlsruher Institut für Technologie (KIT)

Andreas Kuhnle

Adaptive Order Dispatching based on Reinforcement Learning

Application in a Complex Job Shop in the Semiconductor Industry

Band 241

wbk
Institut für Produktionstechnik

Forschungsberichte aus dem
wbk Institut für Produktionstechnik
Karlsruher Institut für Technologie (KIT)

Hrsg.: Prof. Dr.-Ing. Jürgen Fleischer
Prof. Dr.-Ing. Gisela Lanza
Prof. Dr.-Ing. habil. Volker Schulze

Andreas Hermann Kuhnle

Adaptive Order Dispatching based on Reinforcement Learning

Application in a Complex Job Shop in the Semiconductor Industry

Band 241

KIT
Karlsruher Institut für Technologie

wbk
Institut für Produktionstechnik

Adaptive Order Dispatching based on Reinforcement Learning
Application in a Complex Job Shop in the Semiconductor Industry

Zur Erlangung des akademischen Grades eines

Doktors der Ingenieurwissenschaften (Dr.-Ing.)

bei der KIT-Fakultät für Maschinenbau des

Karlsruher Instituts für Technologie (KIT)

angenommene

Dissertation

von

M.Sc. Andreas Hermann Kuhnle

Tag der mündlichen Prüfung:	08.10.2020
Hauptreferentin:	Prof. Dr.-Ing. Gisela Lanza
Korreferenten:	Prof. Dr.-Ing. Michael Freitag
	Prof. Dr. rer. nat. Stefan Nickel

KIT
Karlsruher Institut für Technologie

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the internet at http://dnb.d-nb.de.

Zugl.: Karlsruhe, Karlsruher Institut für Technologie, Diss., 2020

Copyright Shaker Verlag 2020
All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publishers.

Printed in Germany.

ISBN 978-3-8440-7762-9
ISSN 0724-4967

Shaker Verlag GmbH • Am Langen Graben 15a • 52353 Düren
Phone: 0049/2421/99011-0 • Telefax: 0049/2421/99011-9
Internet: www.shaker.de • e-mail: info@shaker.de

wbk
Institut für Produktionstechnik

Vorwort der Herausgeber

Die schnelle und effiziente Umsetzung innovativer Technologien wird vor dem Hintergrund der Globalisierung der Wirtschaft der entscheidende Wirtschaftsfaktor für produzierende Unternehmen. Universitäten können als “Wertschöpfungspartner” einen wesentlichen Beitrag zur Wettbewerbsfähigkeit der Industrie leisten, in dem sie wissenschaftliche Grundlagen sowie neue Methoden und Technologien erarbeiten und aktiv den Umsetzungsprozess in die praktische Anwendung unterstützen.

Vor diesem Hintergrund soll im Rahmen dieser Schriftenreihe über aktuelle Forschungsergebnisse des Instituts für Produktionstechnik (wbk) des Karlsruher Instituts für Technologie (KIT) berichtet werden. Unsere Forschungsarbeiten beschäftigen sich sowohl mit der Leistungssteigerung von Fertigungsverfahren und zugehörigen Werkzeugmaschinen- und Handhabungstechnologien als auch mit der ganzheitlichen Betrachtung und Optimierung des gesamten Produktionssystems. Hierbei werden jeweils technologische wie auch organisatorische Aspekte betrachtet.

Prof. Dr.-Ing. Jürgen Fleischer
Prof. Dr.-Ing. Gisela Lanza
Prof. Dr.-Ing. habil. Volker Schulze

KIT
Karlsruher Institut für Technologie

wbk
Institut für Produktionstechnik

Vorwort des Verfassers

Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am wbk Institut für Produktionstechnik des Karlsruher Instituts für Technologie (KIT).

Mein besonderer Dank gilt Frau Prof. Dr.-Ing. Gisela Lanza für die Betreuung meiner wissenschaftlichen Arbeit als Hauptreferentin, die persönliche Unterstützung sowie Förderung, das mir entgegengebrachte Vertrauen und die Freiheiten in den vergangenen Jahren, welche weit über diese Arbeit hinausgingen. Weiter danke ich Herrn Prof. Dr.-Ing. Michael Freitag und Herrn Prof. Dr. rer. nat. Stefan Nickel für das Interesse an meiner Arbeit und die Übernahme des Korreferats sowie Herrn Prof. Dr.-Ing. Wolfgang Seemann für den Prüfungsvorsitz.

Dem Karlsruher House of Young Scientists danke ich für die Förderung meines Forschungsaufenthaltes am Massachusetts Institute of Technology. Ebenso gilt mein Dank Prof. Stephen C. Graves für die herzliche Gastfreundschaft und die wissenschaftlichen Diskussionen.

Allen Kolleginnen und Kollegen des wbk in Karlsruhe, insbesondere im Bereich PRO, danke ich für die fortwährende Unterstützung in den vergangenen Jahren. Namentlich seien an dieser Stelle Christoph Liebrecht, Stefan Treber, Niclas Eschner, Constantin Hofmann, Tom Stähr, Robin Kopf, Sebastian Greinacher und Nicole Stricker erwähnt. Unsere gemeinsamen Unternehmungen bleiben mir lange in Erinnerung. Für das Korrekturlesen dieser Arbeit danke ich Carmen Krahe und Marvin May. Darüber hinaus, gilt mein besonderer Dank Andreas Fischer, Johannes Jakubik, Nicole Röhrig, Leonard Overbeck, Henri Felsenstein, Louis Schäfer, Kathrina Traub, Katharina König, Felix Hettich, Jan-Philipp Kaiser, Marvin May, Felix Theiss, Sebastian Behrendt, Patrick Haberkern, Tillmann Stüker, Alexander Valet und Alexander Heilig, welche durch ihre Arbeit sowie zahlreiche Diskussionen einen deutlichen Beitrag zum Gelingen der Arbeit geleistet haben.

Den Kollegen der Infineon Technologies AG aus Regensburg und München möchte ich an dieser Stelle ebenfalls herzlich für die tolle und vertrauensvolle Zusammenarbeit danken.

Abschließend möchte ich mich bei meinen Eltern, Geschwistern und Ann-Kathrin dafür bedanken, dass sie mir zu jeder Zeit den nötigen Rückhalt und stets wertvolle Unterstützung gegeben haben.

Karlsruhe, 20.10.2020

Andreas Hermann Kuhnle

KIT
Karlsruher Institut für Technologie

Kurzfassung

Heutige Produktionssysteme tendieren durch die Marktanforderungen getrieben zu immer kleineren Losgrößen, höherer Produktvielfalt und größerer Komplexität der Materialflusssysteme. Diese Entwicklungen stellen bestehende Produktionssteuerungsmethoden in Frage. Im Zuge der Digitalisierung bieten datenbasierte Algorithmen des maschinellen Lernens einen alternativen Ansatz zur Optimierung von Produktionsabläufen. Aktuelle Forschungsergebnisse zeigen eine hohe Leistungsfähigkeit von Verfahren des Reinforcement Learning (RL) in einem breiten Anwendungsspektrum. Im Bereich der Produktionssteuerung haben sich jedoch bisher nur wenige Autoren damit befasst. Eine umfassende Untersuchung verschiedener RL-Ansätze sowie eine Anwendung in der Praxis wurden noch nicht durchgeführt.

Unter den Aufgaben der Produktionsplanung und -steuerung gewährleistet die Auftragssteuerung (*order dispatching*) eine hohe Leistungsfähigkeit und Flexibilität der Produktionsabläufe, um eine hohe Kapazitätsauslastung und kurze Durchlaufzeiten zu erreichen. Motiviert durch komplexe Werkstattfertigungssysteme, wie sie in der Halbleiterindustrie zu finden sind, schließt diese Arbeit die Forschungslücke und befasst sich mit der Anwendung von RL für eine adaptive Auftragssteuerung. Die Einbeziehung realer Systemdaten ermöglicht eine genauere Erfassung des Systemverhaltens als statische Heuristiken oder mathematische Optimierungsverfahren. Zusätzlich wird der manuelle Aufwand reduziert, indem auf die Inferenzfähigkeiten des RL zurückgegriffen wird.

Die vorgestellte Methodik fokussiert die Modellierung und Implementierung von RL-Agenten als Dispatching-Entscheidungseinheit. Bekannte Herausforderungen der RL-Modellierung in Bezug auf Zustand, Aktion und Belohnungsfunktion werden untersucht. Die Modellierungsalternativen werden auf der Grundlage von zwei realen Produktionsszenarien eines Halbleiterherstellers analysiert. Die Ergebnisse zeigen, dass RL-Agenten adaptive Steuerungsstrategien erlernen können und bestehende regelbasierte Benchmarkheuristiken übertreffen. Die Erweiterung der Zustandsrepräsentation verbessert die Leistung deutlich, wenn ein Zusammenhang mit den Belohnungszielen besteht. Die Belohnung kann so gestaltet werden, dass sie die Optimierung mehrerer Zielgrößen ermöglicht. Schließlich erreichen spezifische RL-Agenten-Konfigurationen nicht nur eine hohe Leistung in einem Szenario, sondern weisen eine Robustheit bei sich ändernden Systemeigenschaften auf.

Damit stellt die Forschungsarbeit einen wesentlichen Beitrag in Richtung selbstoptimierender und autonomer Produktionssysteme dar. Produktionsingenieure müssen das Potenzial datenbasierter, lernender Verfahren bewerten, um in Bezug auf Flexibilität wettbewerbsfähig zu bleiben und gleichzeitig den Aufwand für den Entwurf, den Betrieb und die Überwachung von Produktionssteuerungssystemen in einem vernünftigen Gleichgewicht zu halten.

Contents

Contents I

Abbreviations IV

1 Introduction 1
- 1.1 Motivation . . . 1
- 1.2 Problem statement . . . 3
- 1.3 Research goal . . . 4
- 1.4 Structure of this work . . . 5

2 Fundamentals 6
- 2.1 Production planning and control . . . 6
  - 2.1.1 Complex manufacturing systems . . . 8
  - 2.1.2 Logistical performance indicators . . . 11
  - 2.1.3 Production planning and control tasks . . . 16
  - 2.1.4 Order dispatching . . . 18
  - 2.1.5 Summary: Solution approach requirements . . . 19
- 2.2 Quantitative optimization methods . . . 20
  - 2.2.1 Mathematical optimization . . . 21
  - 2.2.2 Heuristics and metaheuristics . . . 22
  - 2.2.3 Machine learning . . . 25
  - 2.2.4 Simulation . . . 27
  - 2.2.5 Summary: Quantitative methods for an adaptive order dispatching . . . 28
- 2.3 Reinforcement Learning . . . 29
  - 2.3.1 Agent and environment . . . 29
  - 2.3.2 Markov Decision Process . . . 32
  - 2.3.3 Basic concepts of Reinforcement Learning . . . 33
  - 2.3.4 Model-free methods . . . 37
  - 2.3.5 Value function approximation . . . 41
  - 2.3.6 Summary: Adaptive order dispatching based on Reinforcement Learning 43

3 State-of-the-art literature review 44
- 3.1 Literature review of research focus areas . . . 44
  - 3.1.1 Dynamic selection and parametrization of dispatching rules . . . 45
  - 3.1.2 Applications of Reinforcement Learning to adaptive scheduling and dispatching . . . 48

3.1.3 Advanced applications of Reinforcement Learning not related to manufacturing . . . 51
3.2 Research deficit . . . 52

4 Adaptive order dispatching based on Reinforcement Learning 54
4.1 Scope and assumptions . . . 55
4.2 Production system modelling and virtual training environment . . . 56
4.2.1 Organizational structure . . . 57
4.2.2 Process control structure . . . 62
4.2.3 Categorization of complexity drivers . . . 64
4.3 Adaptive order dispatching architecture . . . 64
4.3.1 Formal control architecture . . . 65
4.3.2 Order dispatching decision-making based on Reinforcement Learning 66
4.3.3 Learning and operation phase . . . 68
4.4 Reinforcement Learning modelling . . . 69
4.4.1 Agent type – Optimization algorithm . . . 69
4.4.2 Action representation . . . 73
4.4.3 State representation . . . 75
4.4.4 Reward function . . . 82
4.4.5 Function approximation and hyper-parameters . . . 92
4.5 Evaluation and benchmarking . . . 94
4.5.1 Analysis of convergence . . . 95
4.5.2 Performance evaluation criteria . . . 96
4.5.3 Benchmark algorithms . . . 99
4.6 Implementation of adaptive order dispatching system . . . 100
4.7 Summary of approach and framework . . . 101

5 Evaluation and computational results 103
5.1 Complex job shop application use case . . . 105
5.1.1 Production system parameters . . . 105
5.1.2 Use case evaluation scenarios . . . 108
5.1.3 Analysis of system dimensions . . . 109
5.1.4 Benchmark heuristics and validation of simulation model . . . 110
5.2 Performance evaluation of Reinforcement Learning algorithm . . . 114
5.2.1 Validity of action selection . . . 116
5.2.2 State, action, and reward design alternatives . . . 117
5.2.3 Performance variance analysis . . . 123
5.3 Analysis of multi-objective performance . . . 124

5.3.1 Multi-objective reward functions . . . 124
5.3.2 Pareto-optimal Reinforcement Learning configurations . . . 126
5.3.3 Computational effort . . . 130
5.4 Transferability of dispatching policy . . . 130
5.4.1 Integrated comparison of multiple scenarios . . . 130
5.4.2 Investigation of changing scenarios . . . 133
5.5 Dispatching policy plausibility analysis . . . 134
5.5.1 Imitation of heuristic behaviour . . . 134
5.5.2 Decision-making comparison on an action-level . . . 136
5.6 Summary of evaluation and computational results . . . 137

6 Discussion and outlook 139
6.1 Discussion . . . 139
6.2 Outlook and further considerations . . . 143

7 Conclusion 145

References 147

List of Figures 169

List of Tables 173

List of own publications 175

Appendix X
A1 Explanation of quantitative optimization methods . . . X
A2 List of simulation parameters . . . XIV
A3 Methodological approach of modelling RL-based adaptive order dispatching . XVI
A4 Computational results for varying hyper-parameter settings . . . XVII
A5 Theoretical computations of the system's capacity dimensions . . . XIX
A6 Recursion limit, forced idle time, and episode length parameter analysis . . . XXI
A7 Overview of feasible actions . . . XXIII
A8 Correlation of performance indicators . . . XXIV
A9 Computational results of RL-agents for Scenario 2 . . . XXV
A10 Additional computational results . . . XXX

Abbreviations

Abbreviation	Description
ANN	Artificial Neural Network
ATC	Apparent Tardiness Cost
CI	Confidence Interval
CNN	Convolutional Neural Network
CV	Coefficient of Variation
DQN	Deep-Q Network
DU	Distance Unit (arbitrary)
EDD	Earliest Due Date
FIFO	First In First Out
FLNQ	Fewest Lots in the Next Queue
KPI	Key Performance Indicator
LWT	Longest Waiting Time
MDP	Markov Decision Process
ML	Machine Learning
MLP	Multi-Layer Perceptron
MS	Minimum Slack
MTBF	Mean Time Between Failure
MTOL	Mean Time Off Line
NJF	Nearest Job First
OEE	Overall Equipment Effectiveness
PPC	Production Planning and Control
POMDP	Partially Observable Markov Decision Process
ReLU	Rectified Linear Unit
RL	Reinforcement Learning
SARSA	State, Action, Reward, State, and Action
SPT	Shortest Processing Time
SRPT	Shortest Remaining Processing Time
SST	Shortest Setup Time
TD	Temporal-Difference
TRPO	Trust Region Policy Optimization
TU	Time Unit (arbitrary)
WIP	Work In Process
WSPT	Weighted Shortest Processing Time

Symbol	Description	Unit	Value range
∇	Differential operator	-	-
$\varnothing$	Empty set	-	-
α	Learning rate	-	$[0, 1]$
$\alpha\text{-}value$	Alpha value according to the theory of operating curves	-	$\mathbb{R}^{\geq 0}$
$\mathcal{A}$	Set of all possible actions	-	-
$A(s)$	Set of all possible actions in state s	-	$\mathcal{A}$
a, a'	Action, next action	-	$\mathcal{A}$
A_{empty}	Action subset of all actions for moving without dispatching an order	-	$\mathcal{A}$
A_{idle}	Action subset of all actions for waiting and idling for a certain period	-	$\mathcal{A}$
$A_{M \to S}$	Action subset of all actions for dispatching an order from a machine to a sink resource	-	$\mathcal{A}$
$A_{O \to D}$	Definition of a dispatching action from resource O to resource D	-	$\mathcal{A}$
$A_{S \to M}$	Action subset of all actions for dispatching an order from a source to a machine resource	-	$\mathcal{A}$
$A_{valid,t}$	Action subset of all actions that are valid at time t	-	$\mathcal{A}$
a_j	Attribute of order j used for priority index determination	-	$\mathbb{R}$
A_t	Action at time t	-	$\mathcal{A}$
APT_i	Average processing time for machine i	TU	$\mathbb{R}^{\geq 0}$
at_{max}	Highest possible time to perform an action within the system	TU	$\mathbb{R}^{\geq 0}$
at_i	State variable indicating the time it takes to perform a specific action a_i	-	$[0, 1]$
ben_i	State variable indicating the relative inbound buffer availability at machine i	-	$[0, 1]$
bex_i	State variable indicating the relative outbound buffer availability at machine i	-	$[0, 1]$
bpt_i	State variable indicating the total processing time at machine i	-	$\mathbb{R}^{\geq 0}$
γ	MDP discount rate for future rewards	-	$[0, 1]$
c_j	Completion time of order j	-	$\mathbb{R}^{\geq 0}$
CAP_i^{EN}	Capacity of inbound buffer at machine i	-	$\mathbb{N}^{\geq 0}$

Symbol	Description	Unit	Value range
CAP_i^{EX}	Capacity of outbound buffer at machine i	-	$\mathbb{N}^{\geq 0}$
CT_j	Cycle time of order j	TU	$\mathbb{R}^{\geq 0}$
$d_{i,j}$	Transportation distance from resource i to resource j	DU	$\mathbb{R}^{\geq 0}$
d_j	Planned completion time of order j	-	$\mathbb{R}^{\geq 0}$
DT_i	Down time of machine i	TU	$\mathbb{R}^{\geq 0}$
ϵ	Probability of ϵ-greedy action selection	-	$[0, 1]$
ϵ_{conv}	Threshold for convergence of the RL-agent	-	$\mathbb{R}^{\geq 0}$
$E(\bullet)$	Expected value function, i.e. arithmetic mean	-	$\mathbb{R}$
$f(x)$	Objective function of an optimization problem	-	$\mathbb{R}$
$f(Z)$	Activation function of a perceptron	-	$\mathbb{R}$
f_{aggr}	Aggregation function of the tree-based reward	-	$\mathbb{R}$
G_t	Expected discounted reward at time t	-	$\mathbb{R}$
$h(s, a, \theta)$	Parameterized numerical preferences	-	$\mathbb{R}$
θ	Policy parameter vector	-	$\mathbb{R}^{\lvert\theta\rvert}$
i	Multi-purpose index used in several contexts, if just a single index is required	-	-
I	Number of orders, inventory inside the manufacturing system	-	$\mathbb{N}^{\geq 0}$
I_j	Priority rule index of order j	-	$\mathbb{N}^{\geq 0}$
j	Production order index (job, lot, batch), J total number of orders	-	$\{1, ..., J\}$
$J(\theta)$	Scalar performance measure of policy parameters θ	-	$\mathbb{R}$
$\widehat{J(\theta)}$	Estimate of the scalar performance measure of policy parameters θ	-	$\mathbb{R}$
k	Operation index, k_j number of all operations of order j	-	$\{1, ..., k_j\}$
λ	Rate parameter of the exponential distribution	-	$\mathbb{R}^{>0}$
l_i	State variable indicating whether the dispatching resource's is located at resource i	-	$\{0, 1\}$
L_j	Lateness of order j	TU	$\mathbb{R}^{\geq 0}$
μ	Mean time of a stochastic process	TU	$\mathbb{R}^{\geq 0}$
M	Machine resources	-	-
MA	Maintenance resource, n_{MA} number of maintenance workers	-	-

Symbol	Description	Unit	Value range
mf_i	State variable indicating the machine availability at machine i	-	$\{0, 1\}$
MT_i	Manufacturing time of machine i	TU	$\mathbb{R}^{\geq 0}$
n	Machine index (production resource), N total number of machines	-	$\{1, ..., N\}$
OCC_i^{EN}	Number of occupied inbound buffer slots at machine i	-	$\mathbb{N}^{\geq 0}$
OCC_i^{EX}	Number of occupied outbound buffer slots at machine i	-	$\mathbb{N}^{\geq 0}$
$\pi_t(a \mid s)$	Probability according to policy π_t of $A_t = a$ if $S_t = s$	-	$[0, 1]$
$\pi_\theta(a \mid s, \theta)$	Parameterized policy π based on parameter vector θ	-	$\mathbb{R}$
$p(s', r \mid s, a)$	Transition probability function from state s and action a to s' and reward r	-	$[0, 1]$
PT_i	Total processing time at machine i, i.e. sum of processing times of orders in the inbound buffer	TU	$\mathbb{R}^{\geq 0}$
PT_j	Processing time of order j	TU	$\mathbb{R}^{\geq 0}$
$\Phi(\bullet)$	Real-value function	-	$\mathbb{R}$
$q_\pi(s, a)$	State-action value function of state s and action a under policy π	-	$\mathbb{R}$
$\mathcal{R}$	Set of all fixed resources (sources, machines, and sinks)	-	-
$r(s, a, s')$	Expected reward function for the transition from state s and action a to s' (r_{sparse} sparse reward function, $r_{modelled}$ modelled reward function)	-	$\mathbb{R}$
R_t	Reward signal at time t	-	$\mathbb{R}$
$r_{const,ep}$	Episodic constant reward function	-	$\mathbb{R}$
r_{const}	Reward function with constant reward ω for each action subset	-	$\mathbb{R}$
r_{NJF^*}	Reward function with constant reward when the NJF*-action a_{NJF^*} is selected	-	$\mathbb{R}$
$r_{res}(\bullet)$	Combined reward function of multiple weighted functions	-	$\mathbb{R}$
r_{tree}	Reward function for the tree-based reward	-	$\mathbb{R}$
r_{util}	Reward function with utilization as reward value	-	$\mathbb{R}$
$r_{uti,ep}$	Episodic utilization reward function	-	$\mathbb{R}$

Symbol	Description	Unit	Value range
$r_{w,util}$	Reward function with utilization time rewarded differently per action subset	-	$\mathbb{R}$
r_{wt}	Reward function with waiting time as reward value	-	$\mathbb{R}$
$r_{w,wt}$	Reward function with waiting time rewarded differently per action subset	-	$\mathbb{R}$
$r_{wt,ep}$	Episodic waiting time reward function	-	$\mathbb{R}$
r_j	Order release time of order j	-	$\mathbb{R}^{\geq 0}$
$\overline{r_{i \rightarrow end}}$	Reward at the intersection of the moving mean reward and $\overline{r_{lq}}$	-	$\mathbb{R}$
$\overline{r_{lq}}$	Average reward received in the last quarter of an experiment run	-	$\mathbb{R}$
RPT_i	Remaining processing time at machine i	TU	$\mathbb{R}$
rpt_i	State variable indicating the remaining processing time at machine i	-	$\mathbb{R}$
ρ	Utilization of a process according to the queuing theory	-	$[0, 1]$
σ	Standard deviation of a stochastic process time	-	$\mathbb{R}^{\geq 0}$
SO	Source resources, n_{SO} number of sources	-	-
SI	Sink resources, n_{SI} number of sinks	-	-
$\mathcal{S}$	Set of all possible states	-	-
s, s'	State, next state	-	$\mathcal{S}$
S_t	State representation at time t	-	$\mathcal{S}$
S_T	Terminal state of an episode	-	$\mathcal{S}$
S_{AT}	State element containing variable at_i	-	-
S_{BEN}	State element containing variable ben_i	-	-
S_{BEX}	State element containing variable bex_i	-	-
S_{BPT}	State element containing variable bpt_i	-	-
$S_{j,k}$	Sequence of k operations of order j	-	-
S_L	State element containing variable l_i	-	-
S_{MF}	State element containing variable mf_i	-	-
S_{RPT}	State element containing variable rpt_i	-	-
S_{VA}	State element containing variable va_i	-	-
S_{WT}	State element containing variable wt_i	-	-
ST_j	Setup time of order j	TU	$\mathbb{R}^{\geq 0}$
t	Discrete time step or point in time	-	$\mathbb{R}^{\geq 0}$
T	Time period	TU	$\mathbb{R}^{\geq 0}$

Symbol	Description	Unit	Value range
$t_{\rightarrow O}$	Time to move from the current location to the origin O of a dispatching action	TU	$\mathbb{R}^{\geq 0}$
t_{load}	Handling time to load a resource	TU	$\mathbb{R}^{\geq 0}$
$t_{O \rightarrow D}$	Time to perform dispatching from origin O to destination D	TU	$\mathbb{R}^{\geq 0}$
t_{unload}	Handling time to unload a resource	TU	$\mathbb{R}^{\geq 0}$
TP	Throughput of a manufacturing system	$\frac{\mathrm{orders}}{\mathrm{TU}}$	$\mathbb{R}^{\geq 0}$
TR	Transportation resources, n_{TR} number of transportation resources	-	-
TT_j	Transport time of order j	TU	$\mathbb{R}^{\geq 0}$
U	Average manufacturing uptime utilization of all machines	-	$[0, 1]$
U_{disp}	Utilization of the order dispatching resource	-	$[0, 1]$
U_i	Manufacturing uptime utilization of machine i	-	$[0, 1]$
v	Speed of the dispatching resource	$\frac{\mathrm{DU}}{\mathrm{TU}}$	$\mathbb{R}^{\geq 0}$
$v_\pi(s)$	State value function of state s under policy π	-	$\mathbb{R}$
va_i	State variable indicating valid actions	-	$\{0, 1\}$
W	Weight vector of a perceptron	-	$\mathbb{R}^{\lvert W \rvert}$
w_j	Weight of order j	-	$\mathbb{R}^{\geq 0}$
WT_i^{max}	Longest waiting time of an order at resource i	TU	$\mathbb{R}^{\geq 0}$
WT_i^{mean}	Mean waiting time at resource i	TU	$\mathbb{R}^{\geq 0}$
WT_i^{std}	Standard deviation of the waiting time at resource i	-	$\mathbb{R}^{\geq 0}$
wt_i	State variable indicating the waiting time at source or machine i	-	$\mathbb{R}^{\geq 0}$
WT_j	Waiting time of order j	TU	$\mathbb{R}^{\geq 0}$
$x \in \Omega$	Set of feasible solutions in an optimization problem	-	-
X	Input vector of a perceptron	-	$\mathbb{R}^{\lvert X \rvert}$
$x_{j,k,n}$	Assignment variable indicating if operation k of order j is assigned on machine n	-	$\{0, 1\}$
ξ	Initial state distribution	-	-
Y_{dist}	Distance-based evaluation of the tree-based reward	-	$\mathbb{R}^{\geq 0}$
Y_l^i	Evaluation of node i in layer l of the tree-based reward	-	$\mathbb{R}^{\geq 0}$
Y_{util}	Utilization-based evaluation of the tree-based reward	-	$\mathbb{R}^{\geq 0}$
$Z(X, W)$	Perceptron state based on X and W	-	$\mathbb{R}$