Dynamic Scheduling Method for Job-Shop Manufacturing Systems by Deep Reinforcement Learning with Proximal Policy Optimization

RECLAIM

@reclaim_FoF

RECLAIM Manufacturing

Ming Zhang, Yang Lu, Chao Liu and Yuchun Xu

https://doi.org/10.3390/su14095177

THE FRAMEWORK PROGRAMME FOR RESEARCH AND INNOVATION HORIZON 2020

Aston University BIRMINGHAM UK

1. Summary

For increasingly complex modern manufacturing production systems, operational decision making encounters more challenges in terms of having **sustainable manufacturing to satisfy customers and markets' rapidly changing demands**. Nowadays, the efficiency of decision making could not be guaranteed nor meet the **dynamic scheduling requirement** in the job-shop manufacturing environment based on the traditional knowledge-based method. We propose using **AI-enhanced deep reinforcement learning methods** to tackle the dynamic scheduling problem in the **job-shop manufacturing system with unexpected machine failure**. The proximal policy optimization algorithm was used in the DRL framework to accelerate the learning process and improve performance.

2. Objectives

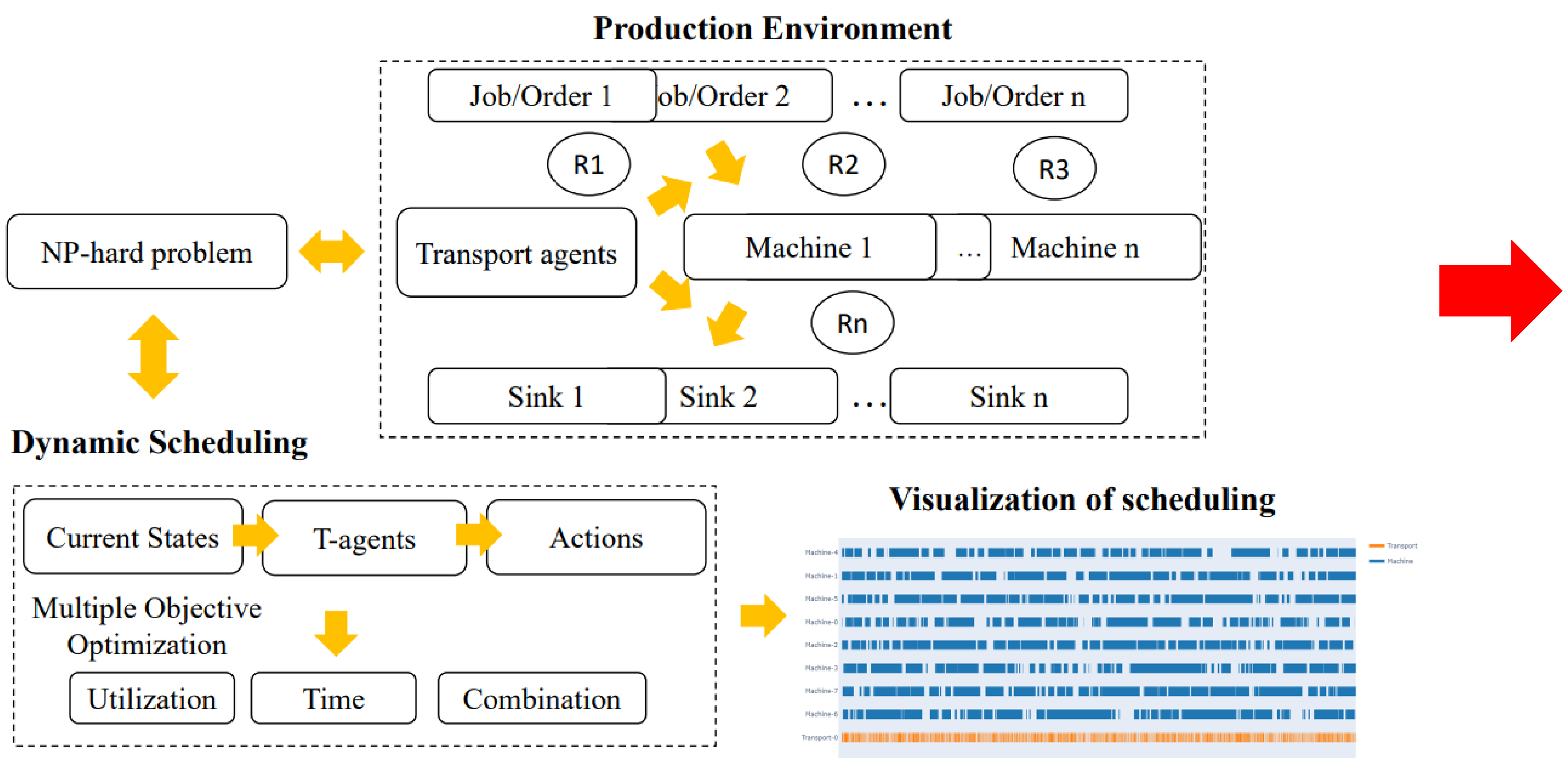
AI-enhanced Data-driven Method

Sustainable Manufacturing

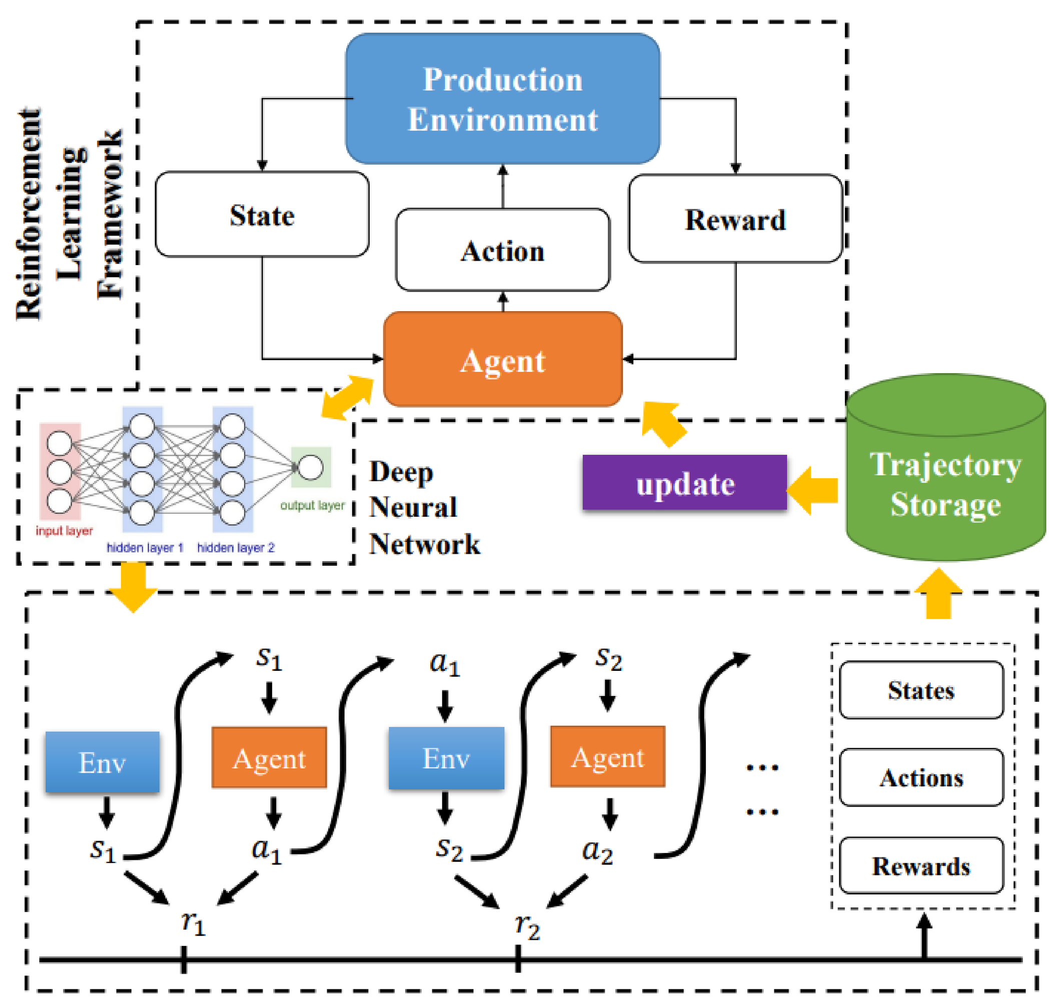
Production High Efficiency

3. Method

Dynamic Job-Shop Scheduling Problem

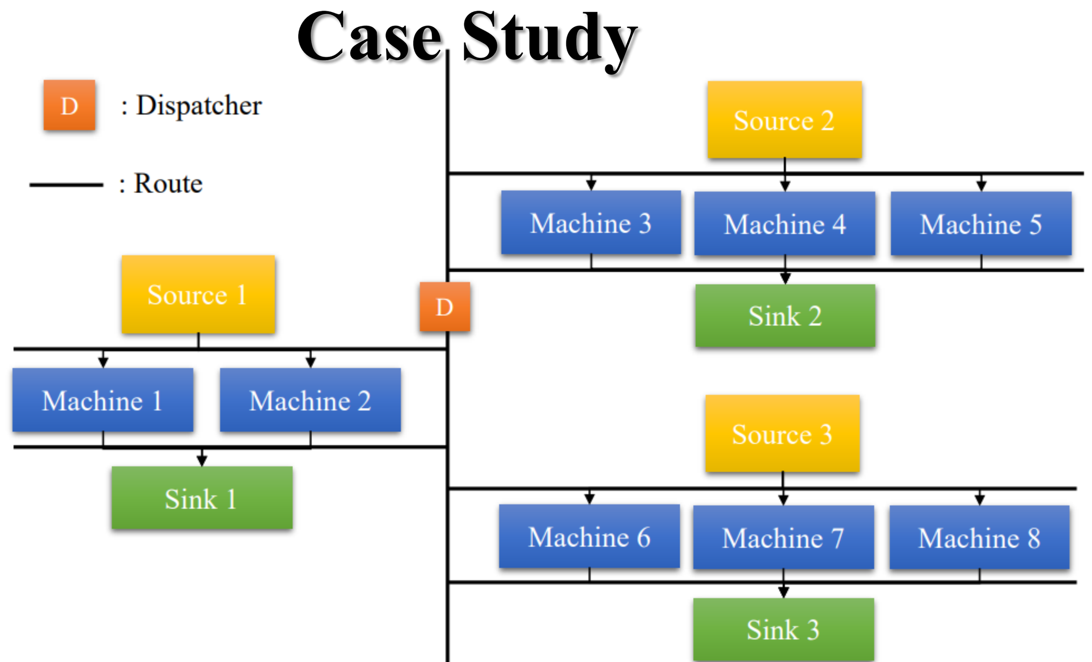


Deep Reinforcement Learning Framework



4. Results

Case Study



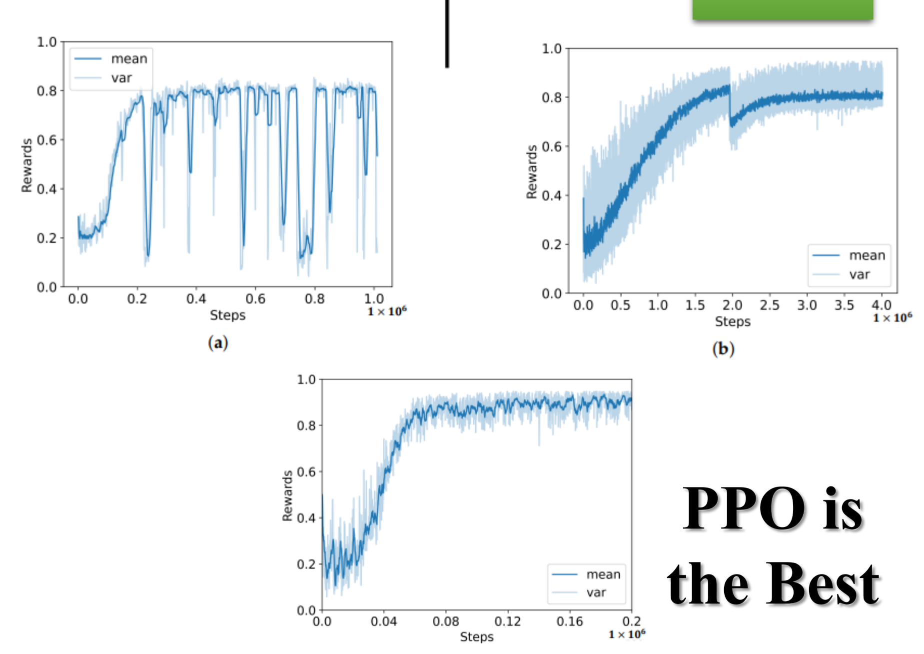


Figure 4. Learning process of different algorithms; (a) policy gradient (PG), (b) trust region policy optimization (TRPO), and (c) proximal policy optimization (PPO).

Sustainability Comparison

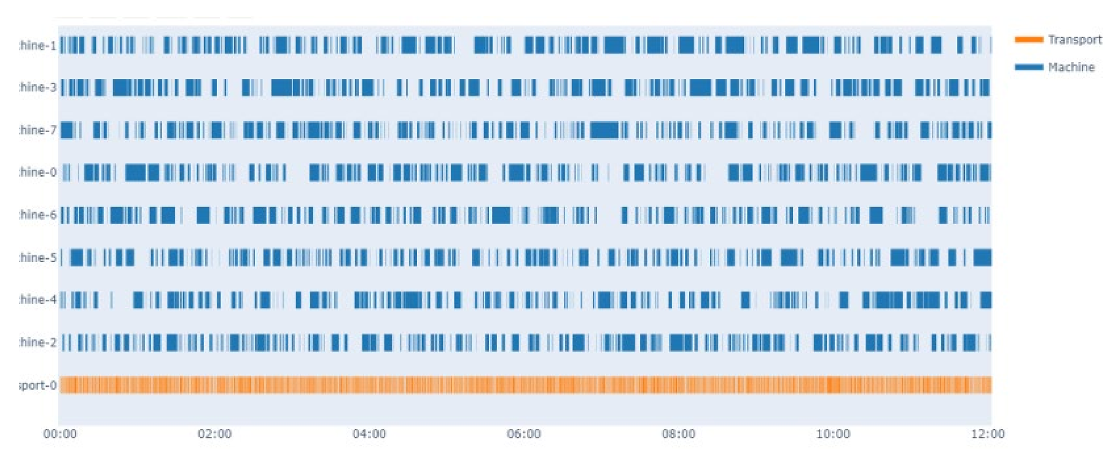


Figure 6. Gantt chart of operating state working with random policy

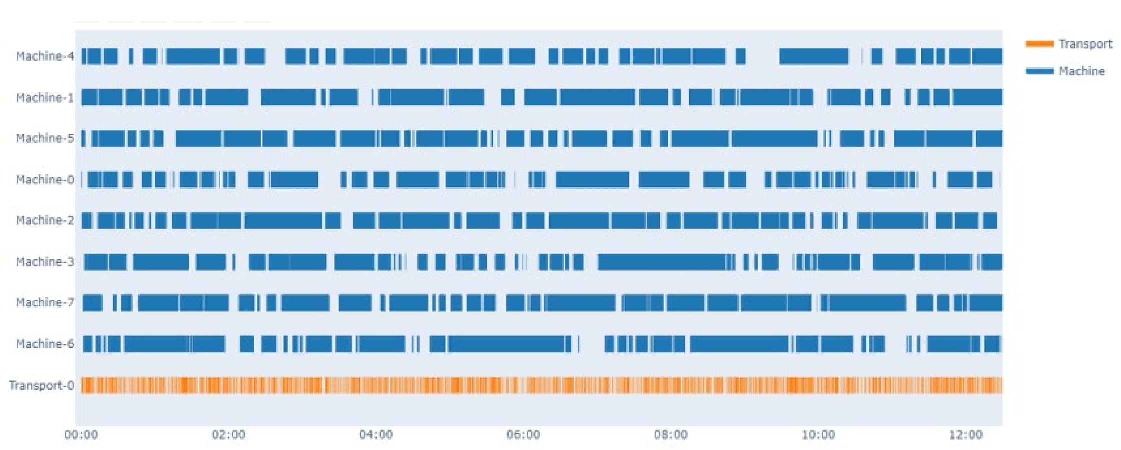


Figure 7. Gantt chart of operating state working with optimal policy trained by the PPO algorithm

Reward Function Comparison

Utilization:

$$R_{\omega-uti}(S_t, A_t) = \begin{cases} \omega_1 R_{uti}(S_t, A_t) & A_t \in A_{S\to M} \\ \omega_2 R_{uti}(S_t, A_t) & A_t \in A_{M\to S} \\ 0 & else \end{cases}$$

Waiting time:

$$R_{\omega-wt}(S_t, A_t) = \begin{cases} \omega_1 R_{wt}(S_t, A_t) & A_t \in A_{S\to M} \\ \omega_2 R_{wt}(S_t, A_t) & A_t \in A_{M\to S} \\ 0 & else \end{cases}$$

Table 5. Results for PPO dispatching approaches under different reward function in both production scenarios

PPO	Scenario 1		
	$U(\%)$	$WT(s)$	α
R_{const}	43.20 ± 3.72	119.30 ± 11.04	2.30 ± 0.63
$R_{\omega-uti}$	44.21 ± 3.60	130.65 ± 11.51	2.37 ± 0.59
$R_{\omega-wt}$	43.68 ± 4.11	126.61 ± 12.02	2.38 ± 0.71
R_{hybird}	43.35 ± 3.67	124.53 ± 19.15	2.32 ± 0.62
PPO	**Scenario 2**		
	$U(\%)$	$WT(s)$	α
R_{const}	62.29 ± 5.02	80.79 ± 14.87	0.56 ± 0.15
$R_{\omega-uti}$	66.31 ± 7.09	99.87 ± 20.55	0.54 ± 0.18
$R_{\omega-wt}$	62.03 ± 5.98	80.10 ± 15.63	0.57 ± 0.18
R_{hybird}	62.75 ± 6.99	80.56 ± 17.12	0.54 ± 0.19

Multiple-objective: $R_{hybird}(S_t, A_t) = w_1 R_{uti} + w_2 R_{wt}$

Table 6. Results for different combination of parameters ω_1 and ω_2 under reword function R_{hybird} in production scenario 2.

	$U(\%)$	$WT(s)$	α
$\omega_1 = 0.1, \omega_2 = 0.9$	61.89 ± 5.81	80.99 ± 16.14	0.57 ± 0.16
$\omega_1 = 0.25, \omega_2 = 0.75$	62.30 ± 6.08	80.35 ± 14.69	0.56 ± 0.17
$\omega_1 = 0.5, \omega_2 = 0.5$	62.75 ± 6.99	80.56 ± 17.12	0.54 ± 0.19
$\omega_1 = 0.75, \omega_2 = 0.25$	68.46 ± 7.02	106.22 ± 19.30	0.48 ± 0.16
$\omega_1 = 0.9, \omega_2 = 0.1$	69.79 ± 7.16	104.88 ± 20.29	0.44 ± 0.16

5. Conclusion

The deep reinforcement learning framework with the PPO algorithm has been approved as a suitable solution to the dynamic scheduling problems in the manufacturing environment. This research is still in the initial phase. However, it shows **the powerful potential of data-driven AI-based methods to significantly enhance the manufacturing process**.

Dr Ming Zhang
m.zhang21@aston.ac.uk

Prof. Yuchun Xu
y.xu16@aston.ac.uk

Dr Chao Liu
liuc16@aston.ac.uk