Review of A multi-action deep reinforcement learning framework for flexible Job-shop scheduling problem

Kun Leia, Peng Guoa,b, Wenchao Zhaoa, Yi Wangc,
Linmao Qiana, Xiangyin Menga, Liansheng Tang

Expert Systems With Applications(2022)

경영과학연구실 김지원

Background

Flexible Job-shop scheduling problem

$J = \{J_1, \cdots, J_n\} of\ n\ jobs\ and\ M = \{M_1, \cdots, M_m\} of\ m\ machines$

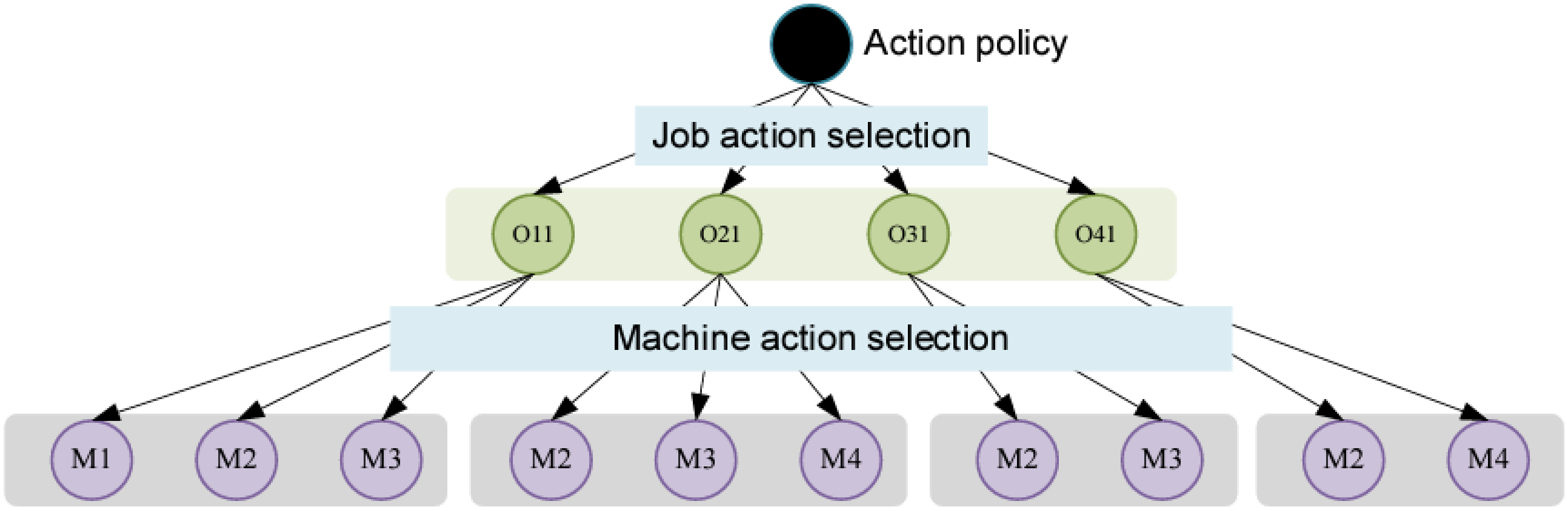
Job Ji consists of a specific sequence of ni consecutive operations $O_i = \{O_{i1}, O_{i2}, \cdots, O_{in_i}\}$ with precedence constraints

- Dispatching rules used in solving FJSP can be divided into two basic categories: the job selection rules and the machine selection rules
- FJSP objectives: to minimize scheduling objectives such as mean flow time, mean tardiness, and maximum tardiness

Multi-action space for FJSP

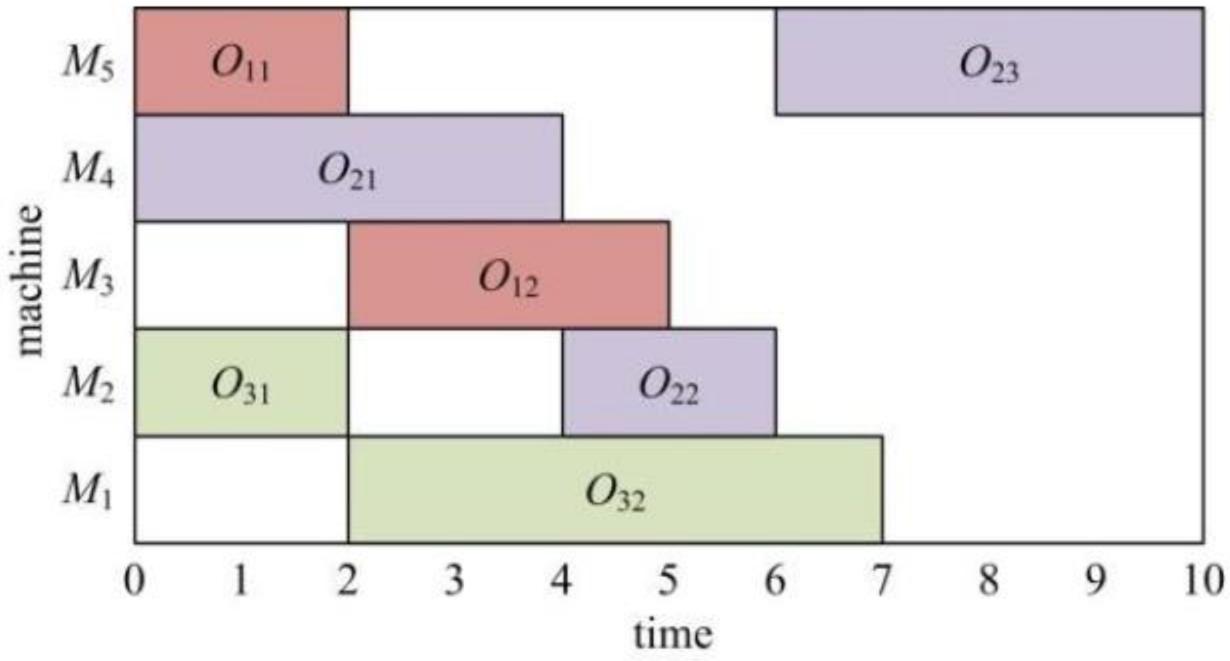
- Hierarchical multi-action space of FJSP involves with a job operation action space and machine action space

: At each timestep, the RL agent selects an operation action from its eligible operation action space and then chooses a machine action for the selected operation action from its compatible machine action space.



Problem Statement

Optimization goal of FJSP: to assign operations to compatible machines and determine a sequence of operations on a machine for minimizing the makespan



Literature review

1) Solving FJSP via mathematical programming and heuristics

- Approximate methods such as swarm intelligence (SI) and evolutionary algorithms (EA) are employed to solve scheduling problems in recent years
- Doh, et al.(2013) suggested a heuristic approach that combines machine assignment rules and job sequencing rules for solving FJSP with multiple process plans
- Zhang, Mei, & Zhang (2019) proposed genetic programming(GP_ for FJSP and dynamic flexible job shop scheduling problem

2) Solving optimal problems via DRL

- Wang, et al.(2021) proposed a DRL approach for dynamic Job-shop scheduling in intelligent manufacturing and showed their method outperforms heuristic rules and meta-heuristic algorithms.
- Waschneck, et al. (2018) proposed cooperative agents based on Deep Q- Network (DQN) designed for production scheduling

Key Idea

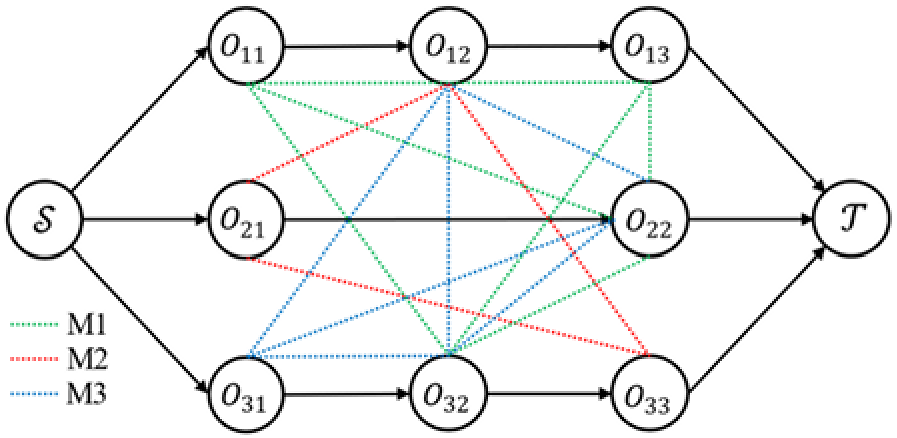
The paper proposed DRL architecture on FJSP on multi-action space

1) Multiple Markov decision processes (MMDP) to represent both job operation and machine states
2) Multi pointer graph network (MPGN) to define the job operation action policy and the machine action policy
3) multi-Proximal Policy Optimization (multi-PPO) to learn two sub-policies, including a job operation action policy and a machine action policy

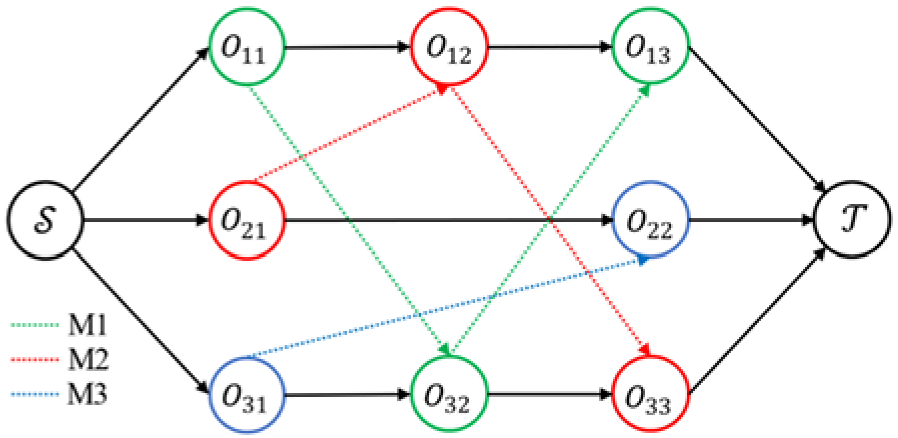
Multiple Markov decision processes (MMDP)

Disjunctive graph for Flexible Job-shop Scheduling problem

G=(O, C,D). Here, O={Oij|∀i,j}∪{S,T}is a set of all operations (S, T : dummy nodes)



(a) Disjunctive graph for an FJSP instance



(b) Example of a feasible solution

Black arrows represent conjunctive arcs representing the precedence constraints
Colored lines represent disjunctive arcs representing eligible machine cliques

Multiple Markov decision processes (MMDP)

State: local states of operations and machines

1) Local state of operation Oij

a disjunctive graph on the previous page

Nodes: Each node contains two features
① the completion time of scheduled operation or the estimated completion time of unscheduled operation
② binary variable representing whether the operation is scheduled or not

Arcs: the set of arcs which have been assigned directions till timestep t and the set of remaining disjunction arcs

2) Local state of machine Mk

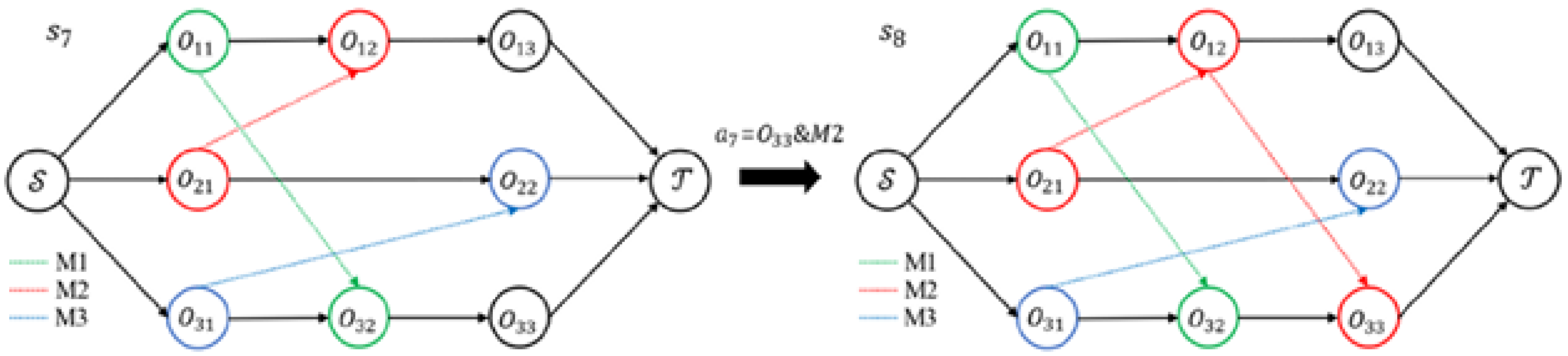
① the completion time for machine Mk
② the processing time of operation Oij on machine Mk if machine Mk is compatible or the average processing time of other compatible machines otherwise

Multiple Markov decision processes (MMDP)

Actions: The actions at timestep t are composed of job operation action and machine action

Transition: the directions of disjunctive arcs are updated based on the current job operation action and machine action

Reward: the negative value of the makespan gap between two continuous timesteps t and t+1



Multi- pointer graph network (MPGN)

Two encoder-decoder components, which define the job operation action policy and the machine action policy, respectively.

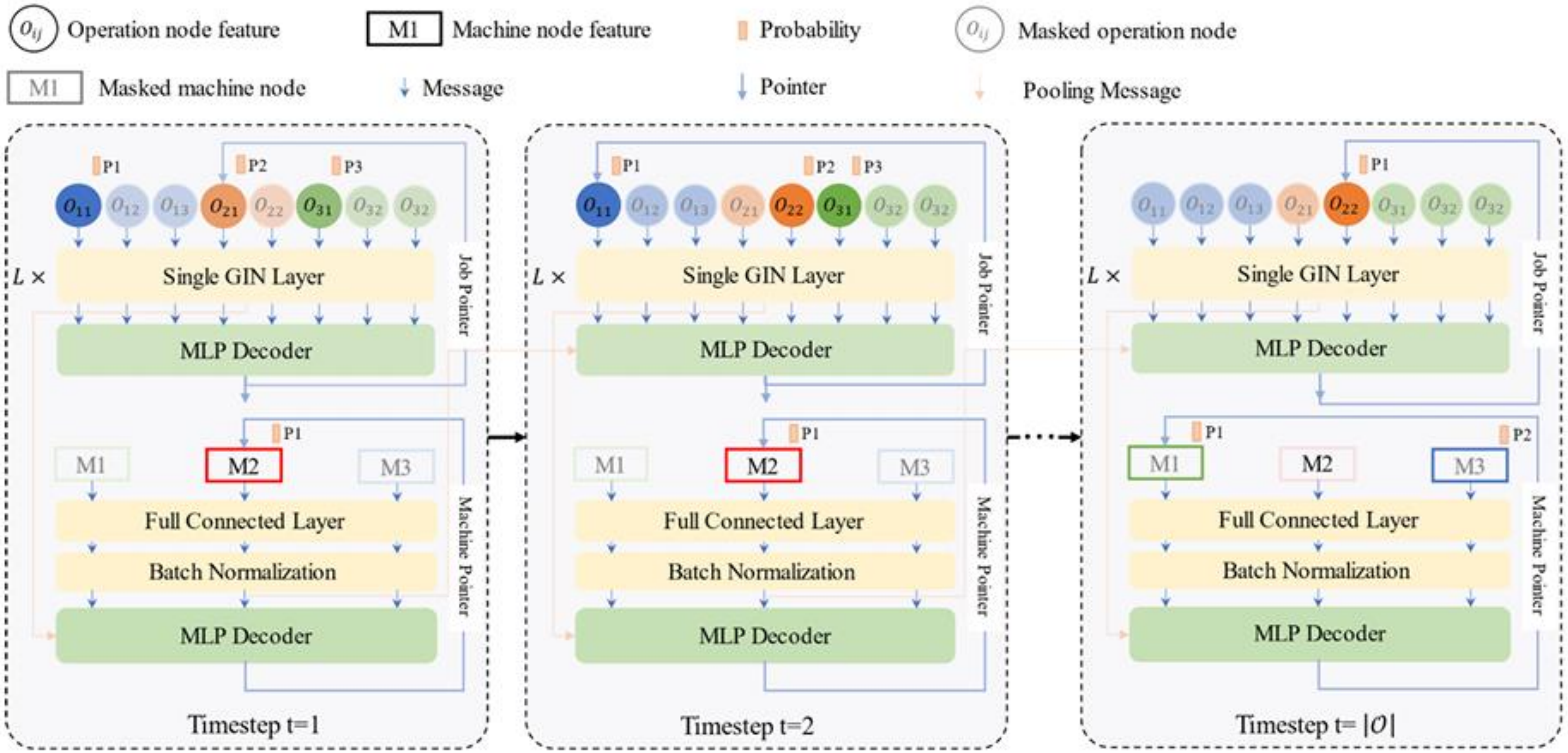


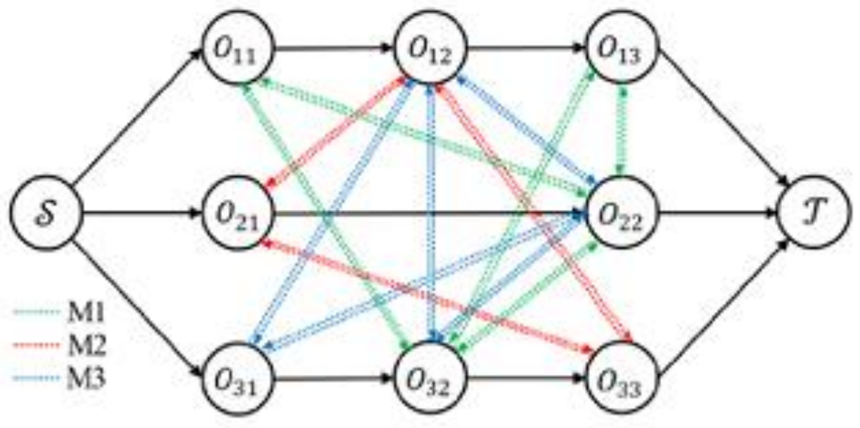
Fig. 3. The MPGN architecture for the FJSP.

Multi- pointer graph network (MPGN)

(1) Job operation encoder (Graph embedding)

- The complex graph state is embedded by exploiting a graph neural network (GNN)
- Each node is encoded via a L-layer Graph Isomorphism Network (GIN)

*GIN: GNN variant designed to maximize representational power of a GNN



Multi- pointer graph network (MPGN)

(2) Machine encoder (node embedding)

- There is no graph structure in the machine's state information
- Therefore, the paper adopted a full connected layer to encode the local state of machine

Multi- pointer graph network (MPGN)

(3) Decoders (action selection)

- At each timestep t, the job decoder selects a job operation action and the machine decoder selects a machine action
- Each decoder is based on MLP layers
- In decoding, each decoder computes a probability distribution over either the job operation action space or the machine action space

Multi-Proximal Policy Optimization (multi-PPO) algorithm

Multi-Proximal Policy Optimization (multi-PPO) algorithm

The proposed multi-PPO architecture includes two actor networks (job operation and machine encoder-decoders)

Each actor learns a stochastic policy to select operation and machine action respectively

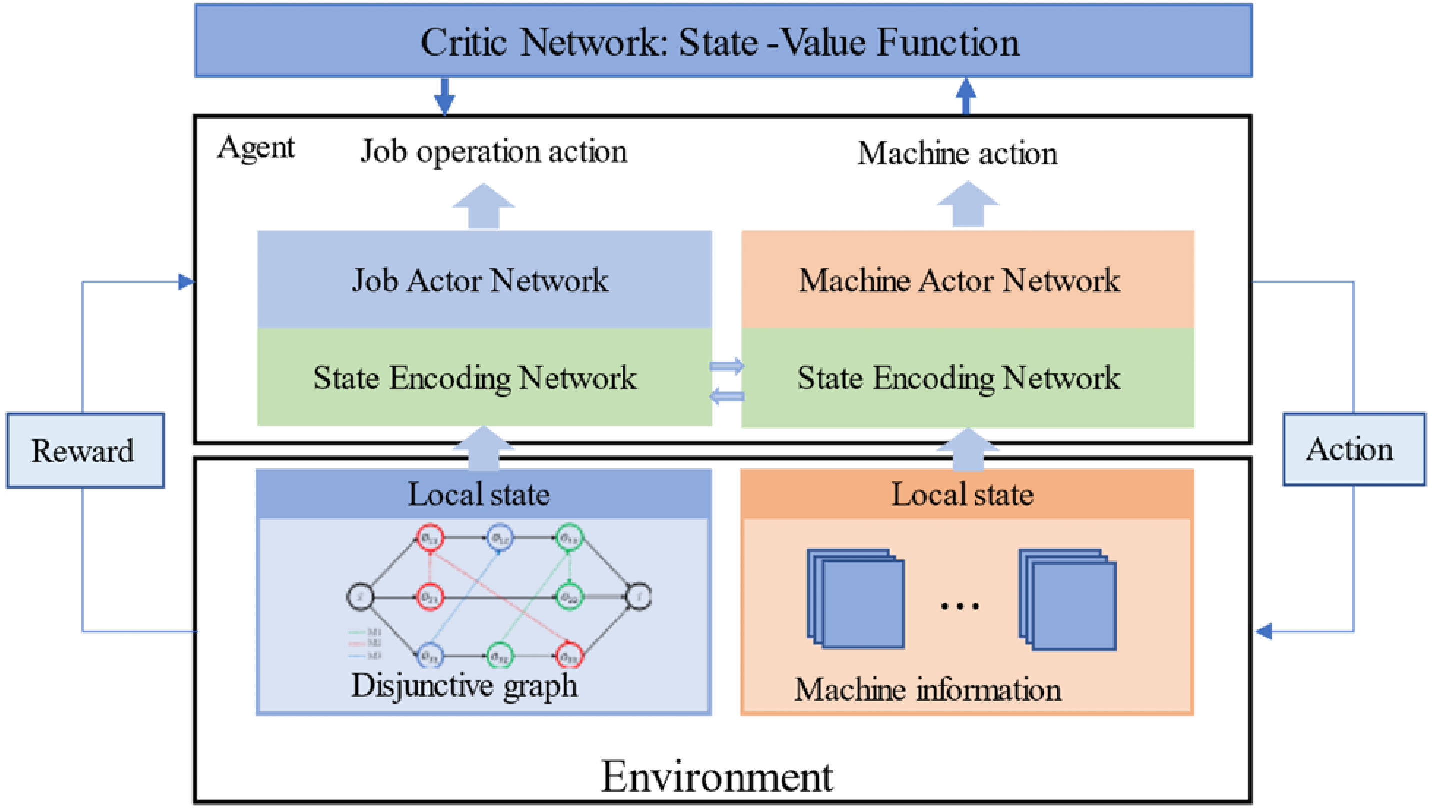


Fig. 5. Multiple actor-critic architecture for a multi-action space scheduling problem.

Computational experiment

Dataset for small and middle-scale experiments

Training set with 12,800 FJSP instances, validation set with 128 FJSP instances, and testing set with 128 FJSP instances

An example 3 × 3 FJSP instance.

p_{ijk}	Job 1			Job 2			Job 3		
	O_{11}	O_{12}	O_{13}	O_{21}	O_{22}	O_{23}	O_{31}	O_{32}	O_{33}
Machine 1	–	–	56.4	–	66.1	–	–	69.5	37.8
Machine 2	45.3	22.5	–	35.8	–	65.4	–	–	–
Machine 3	–	9.8	–	–	78.7	26.3	34.9	54.4	–

Computational experiment

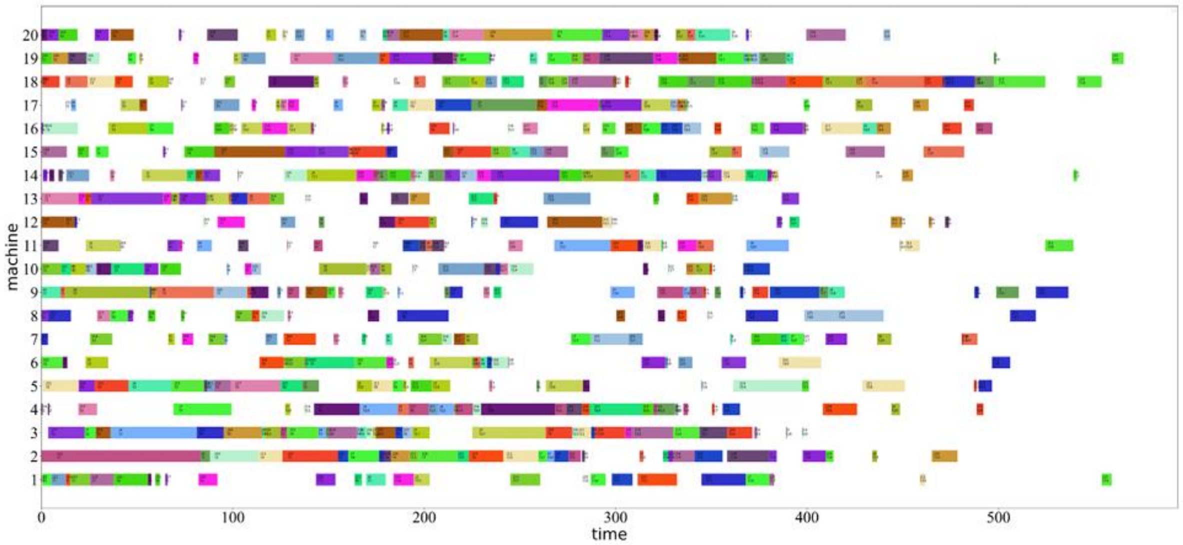
Experimental Results of small-sized experiments

Table 3
Results of all methods on randomly generated instances.

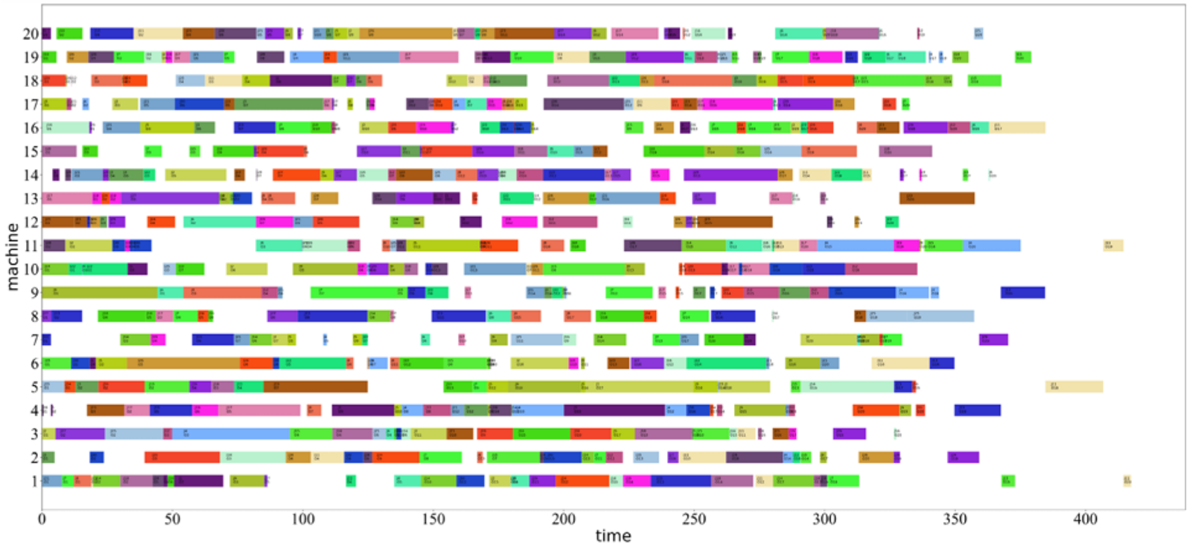
Size		MIP	FIFO + SPT	MOPNR + SPT	LWKR + SPT	MWKR + SPT	FIFO + EET	MOPNR + EET	LWKR + EET	MWKR + EET	Ours
6 × 6	Obj.	227.86	328.45	329.28	397.02	331.58	418.62	438.59	474.21	613.07	**272.32**
	Gap	0.00%	44.15%	44.51%	74.24%	45.52%	83.72%	92.48%	108.11%	169.06%	**19.51%**
	Time (s)	0.73	0.028	0.032	0.028	0.027	0.026	**0.025**	**0.025**	0.026	0.041
10 × 10	Obj.	255.88	385.82	377.60	495.94	382.43	661.25	711.71	821.53	1173.02	**320.45**
	Gap	0.00%	50.78%	47.57%	93.82%	49.46%	158.42%	178.14%	221.06%	358.43%	**25.23%**
	Time (s)	2962	0.084	0.092	0.085	0.087	**0.077**	0.086	0.079	0.080	0.14
15 × 15	Obj.	287.23	413.00	412.91	567.56	409.99	966.56	1046.48	1259.30	1957.39	**347.99**
	Gap	0.00%	43.79%	43.76%	97.60%	42.74%	236.51%	264.34%	338.43%	581.47%	**21.15%**
	Time (s)	3600	0.24	0.26	0.23	0.24	**0.21**	0.24	0.25	0.25	0.39
20 × 10	Obj.	391.41	566.32	569.36	733.16	567.37	1063.45	1107.93	1210.85	1815.82	**454.85**
	Gap	0.00%	44.69%	45.46%	87.31%	44.96%	171.70%	183.06%	209.36%	363.92%	**16.21%**
	Time (s)	3600	0.21	0.24	0.22	0.21	**0.19**	0.23	0.23	0.23	0.34
20 × 20	Obj.	322.54	434.48	430.79	609.96	430.72	1262.36	1384.12	1709.83	2762.01	**361.75**
	Gap	0.00%	34.71%	33.56%	89.11%	33.54%	291.38%	329.13%	430.11%	756.33%	**12.16%**
	Time (s)	3600	0.42	0.46	0.45	0.44	**0.38**	0.41	0.42	0.44	1.08
30 × 20	Obj.	–	528.51	525.08	741.08	522.93	1633.91	1732.54	2087.02	3462.27	**433.42**
	Gap		21.94%	21.15%	70.98%	20.65%	276.98%	299.74%	381.52%	698.83%	**0.00%**
	Time (s)		0.78	0.86	0.83	0.83	**0.69**	0.78	0.81	0.82	1.97

Computational experiment

Experimental Results of small-sized experiments



(a) The Gantt chart of the best dispatching rule



(b) The Gantt chart of our method

Computational experiment

Experimental Results of middle-sized experiments

Table 4
Results of all methods on randomly generated instances.

Size		FIFO + SPT	MOPNR + SPT	LWKR + SPT	MWKR + SPT	FIFO + EET	MOPNR + EET	LWKR + EET	MWKR + EET	Ours (20 × 20)	Ours (30 × 20)
50 × 20	Obj.	716.07	716.10	1002.88	716.61	2567.54	2631.44	2889.44	4829.83	590.22	**587.48**
	Gap	21.89%	21.89%	70.71%	21.98%	337.04%	347.92%	391.84%	722.13%	0.47%	**0.00%**
	Time (s)	1.34	1.48	1.39	1.41	**1.10**	1.32	1.35	1.36	4.12	4.12
100 × 20	Obj.	1201.32	1199.82	1574.44	1199.10	5004.78	5041.13	5184.85	7840.83	1071.03	**1054.70**
	Gap	13.90%	13.76%	49.28%	13.69%	374.52%	377.97%	221.06%	643.42%	1.55%	**0.00%**
	Time (s)	6.66	7.40	6.83	7.03	**4.62**	6.35	6.70	6.99	18.34	18.34