Multi-Agent Deep Reinforcement Learning for Collaborative Task Scheduling

Mali Imre Gergely
Babeș-Bolyai University

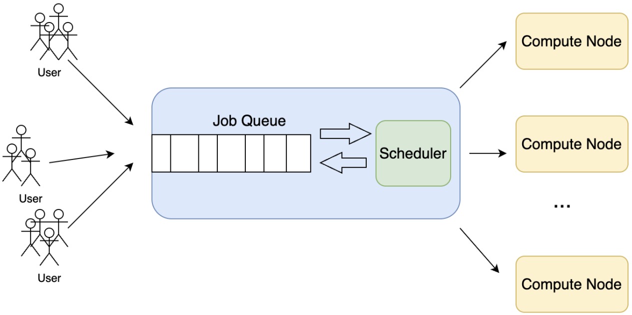
WeADL 2024 Workshop

The workshop is organized by the Machine Learning research group (www.cs.ubbcluj.ro/ml) and the Romanian Meteorological Administration (https://www.meteoromania.ro/)

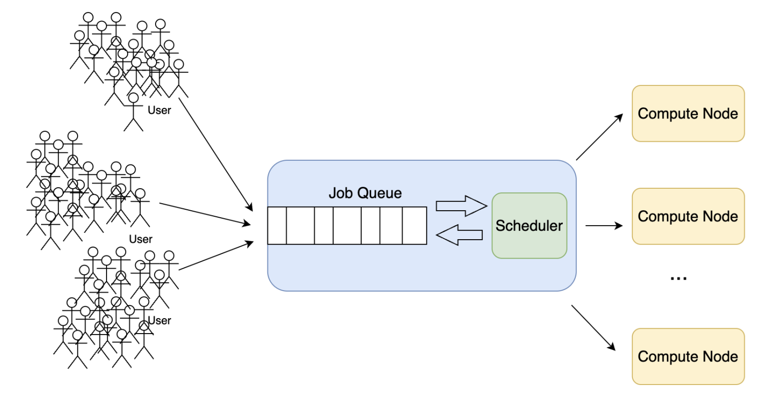
Machine Learning Research Group
MLyRE

Summary

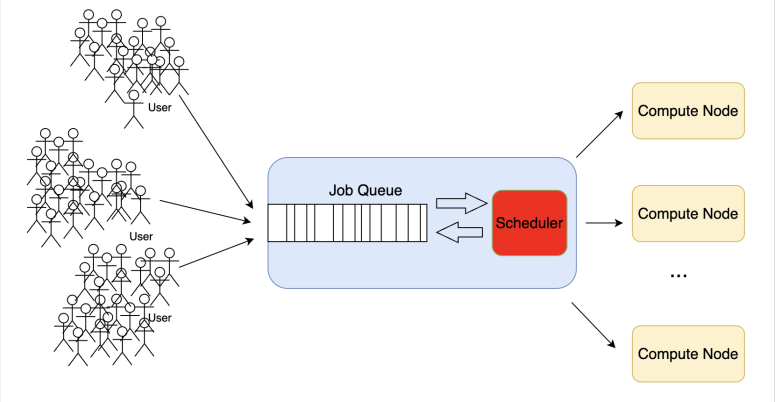
Task Scheduling



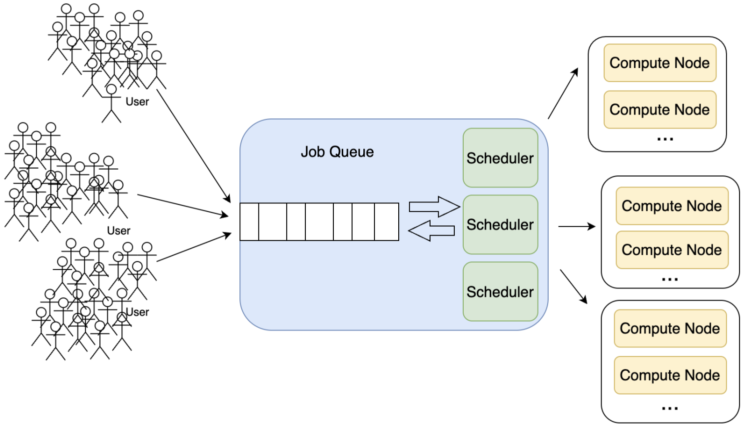
Task Scheduling



Task Scheduling



Task Scheduling



Related Work, Inspiration

- DeepRM [Mao et al., 2016]
- DRAS [Fan et al., 2022]
- DeepMAG [Zhadan et al., 2023]

Environment

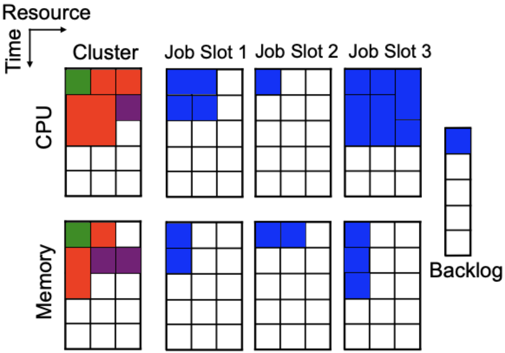


Figure: DeepRM environment

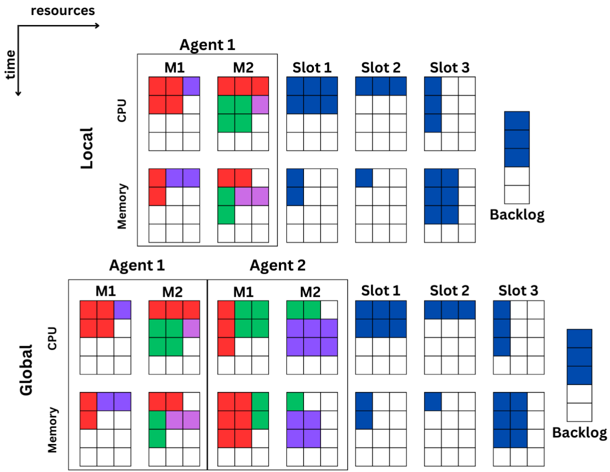
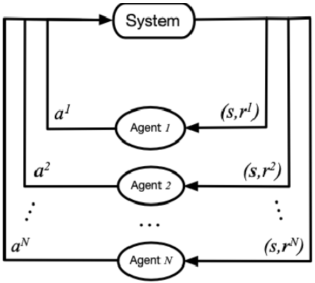


Figure: Our environment

Agents - PPO [Schulman et al., 2017]



$$L(\theta) = \hat{\mathbb{E}}[min(r_t(\theta)\hat{A}_t, clip(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t)]$$

Job Slot - shared resource

AEC games [Terry et al., 2021]:
$(S, s_0, N, (A_i)_{i\in[N]}, (T_i)_{i\in[N]}, P, (R_i)_{i\in[N]}, (\Omega_i)_{i\in[N]}, (O_i)_{i\in[N]}, v)$, where:

- S - states, s_0 is the initial state.
- N - number of agents; agents from 1 to N;environment = agent 0.
- A_i - actions for agent i. For convenience, A_0 is generally void.
- T_i - agent i's state transition function
- P - environment transition function.
- R_i - possible rewards for agent i.
- Ω_i - possible observations for agent i, while O_i observation function.
- v - compute next agent

Questions

- can independent, fully decentralised PPO agents learn task scheduling?
- what effect does the locality/globality of observations have?
- how do these agents perform against heuristics?

Environment Parameters

Parameter	Value
time horizon	20
job slot size	5
number of resources	2
resource capacity	10
backlog size	60
number of machines per agent	1,2
number of agents	2,3

Table: Environment parameters

Agent Parameters

Parameter	Value	Description
learning rate	$3 \cdot 10^{-3}$	
batch_size	64	
γ	0.99	discount factor
gae-λ	0.95	bias-variance tradeoff factor [Schulman et al., 2015]
clip-range	0.2	clipping parameter for the surrogate loss
entropy coeff.	0.0	Used in computing the loss
value-function coeff.	0.5	

Table: PPO Agent Parameters

Global

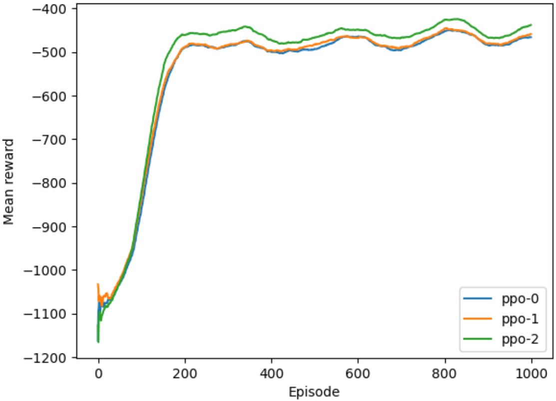


Figure: Mean reward obtained over time by three PPO agents, using global observations, global rewards and having one machine per agent.

Local

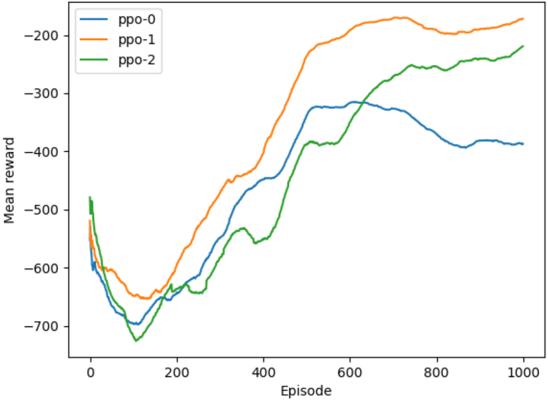


Figure: Running mean reward obtained over time by three PPO agents, using local observations, local rewards and having one machine per agent.

PPO-s vs Heuristics

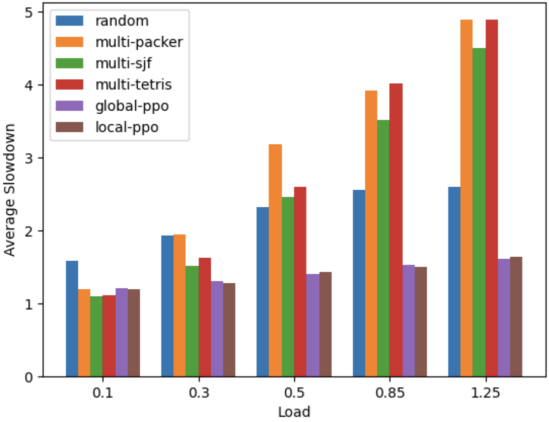


Figure: Average job slowdowns over load factors of different methods in an environment with 3 machines and 1 machine per agent.

Conclusions

- global vs local perform roughly the same
- training speed and convergence stability
- scalability

References I

Fan, Y., Li, B., Favorite, D., Singh, N., Childers, T., Rich, P., Allcock, W., Papka, M. E., and Lan, Z. (2022).
Dras: Deep reinforcement learning for cluster scheduling in high performance computing.
IEEE Transactions on Parallel and Distributed Systems, 33(12):4903–4917.

Mao, H., Alizadeh, M., Menache, I., and Kandula, S. (2016).
Resource management with deep reinforcement learning.
In *Proceedings of the 15th ACM workshop on hot topics in networks*, pages 50–56.

Schulman, J., Moritz, P., Levine, S., Jordan, M., and Abbeel, P. (2015).
High-dimensional continuous control using generalized advantage estimation.
arXiv preprint arXiv:1506.02438.

References II

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O. (2017).
Proximal policy optimization algorithms.
arXiv preprint arXiv:1707.06347.

Terry, J., Black, B., Grammel, N., Jayakumar, M., Hari, A., Sullivan, R., Santos, L. S., Dieffendahl, C., Horsch, C., Perez-Vicente, R., et al. (2021).
Pettingzoo: Gym for multi-agent reinforcement learning.
Advances in Neural Information Processing Systems, 34:15032–15043.

Zhadan, A., Allahverdyan, A., Kondratov, I., Mikheev, V., Petrosian, O., Romanovskii, A., and Kharin, V. (2023).
Multi-agent reinforcement learning-based adaptive heterogeneous dag scheduling.
ACM Transactions on Intelligent Systems and Technology, 14(5):1–26.