Revisiting Some Common Practices in Cooperative Multi-Agent Reinforcement Learning

Wei Fu, Chao Yu, Zelai Xu, Jiaqi Yang, Yi Wu
Proceedings of the 39th International Conference on Machine Learning, PMLR, 2022.

Goal

Revisiting common design principles in
Cooperative Multi-Agent ML

- Value Decomposition (VD)
- parameter sharing

Propose method to resolve limitation of common methods

Background: Value-Decomposition (VD)

$$Q_{\text{tot}}\left(s, \mathbf{a}\right) = f_{\text{mix}}\left(Q_1\left(o^1, a^1\right), \ldots, Q_n\left(o^n, a^n\right); s\right),$$

XOR game:

	1	2
1	0	1
2	1	1

Can’t solve multi-modal problem

Policy Gradient

- parameter sharing → can’t solve
- Individual policy → single optimal mode &
  may challenge optimization
- ID-conditioned policy → single optimal mode

Policy for Multi-modal

Executing Individual PG with every agent order can achieve this, but need large computation

Covering all modes with single policy??

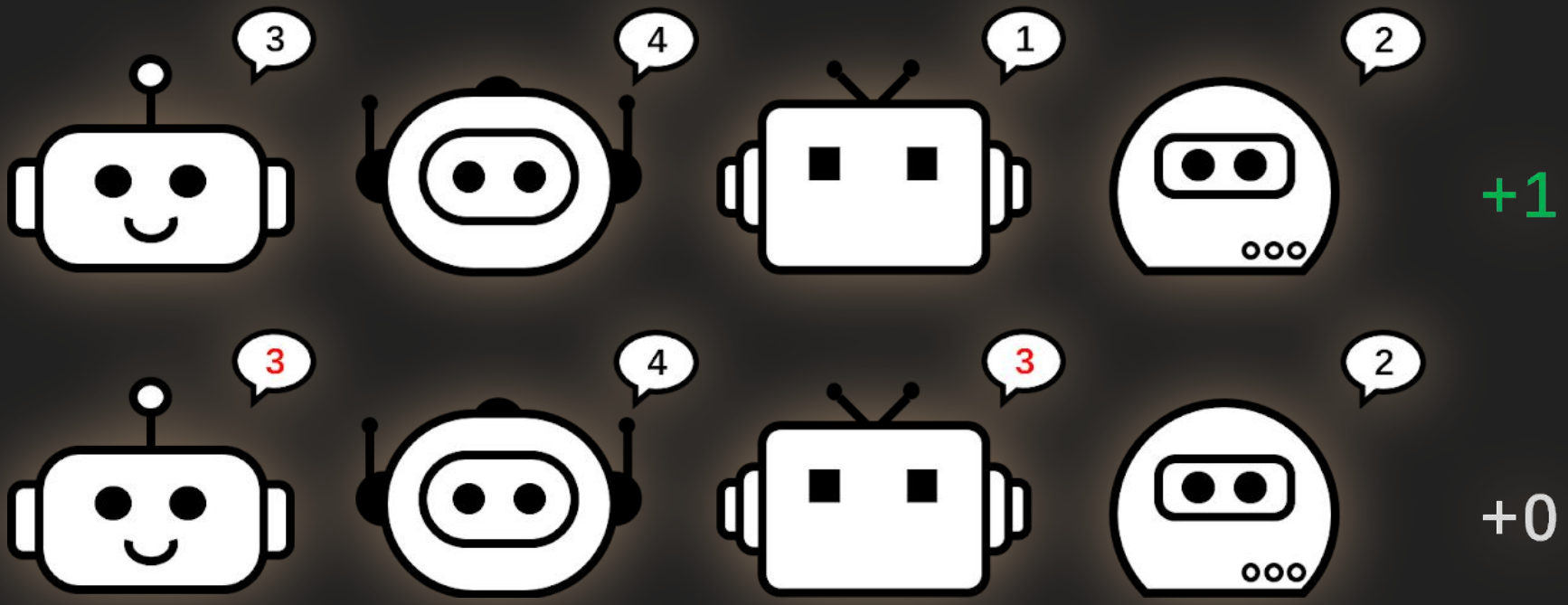
PG-AR (Auto-Regressive)

Individual: $\pi(\mathbf{a} \mid \mathbf{o}) \approx \prod_{i=1} \pi_{\theta_i}\left(a^i \mid o^i\right)$

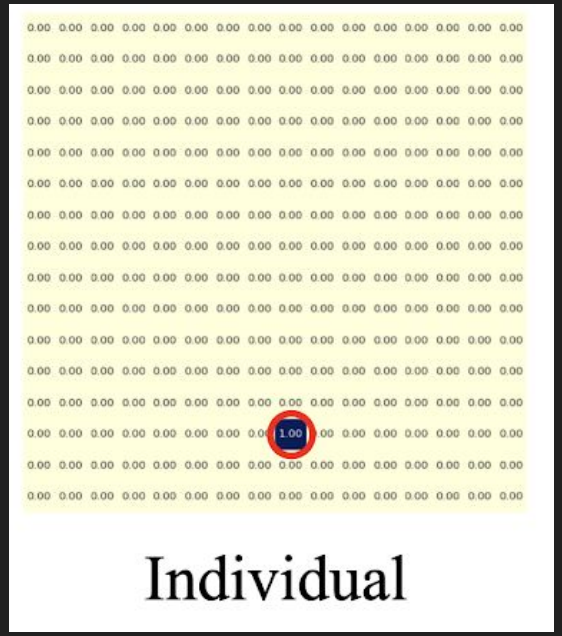
PG-AR: $\pi_\theta(\mathbf{a} \mid \mathbf{o}) \approx \prod_{i=1}^{n} \pi_{\theta^{x_i}}\left(a^{x_i} \mid o^{x_i}, a^{x_1}, \ldots, a^{x_{i-1}}\right)$

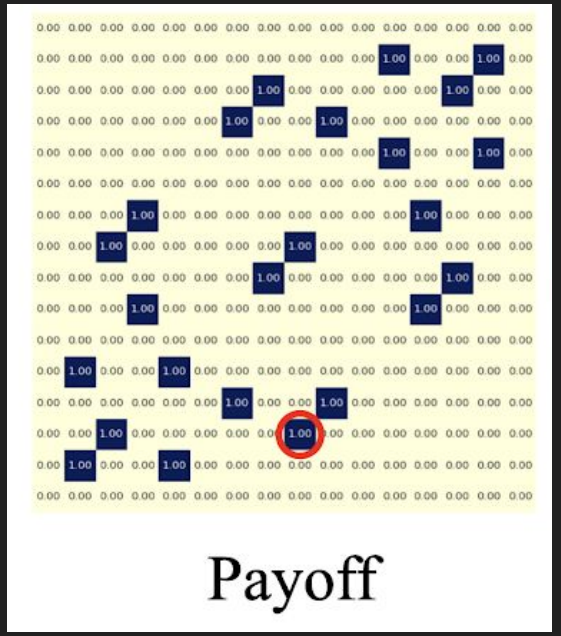
PG-AR
In XOR game: $\pi\left(a^1, a^2\right) = \pi\left(a^1\right) \pi\left(a^2 \mid a^1\right)$

Permutation game



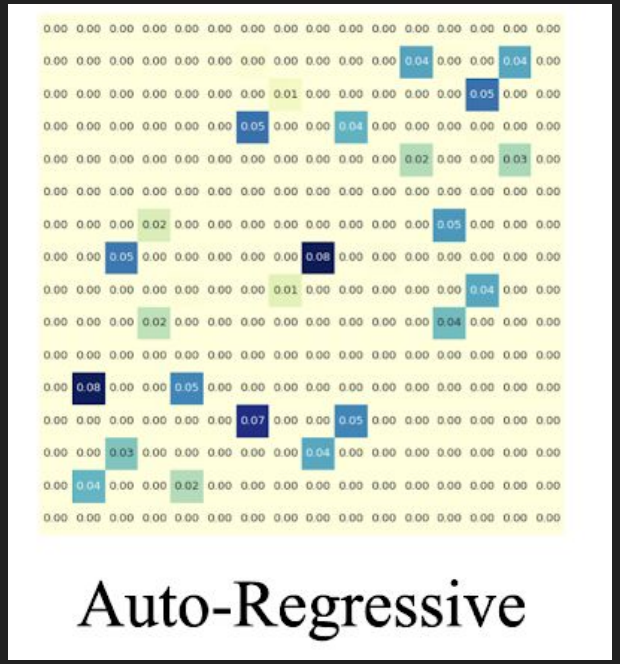
Individual policy: 4-agent permutation game

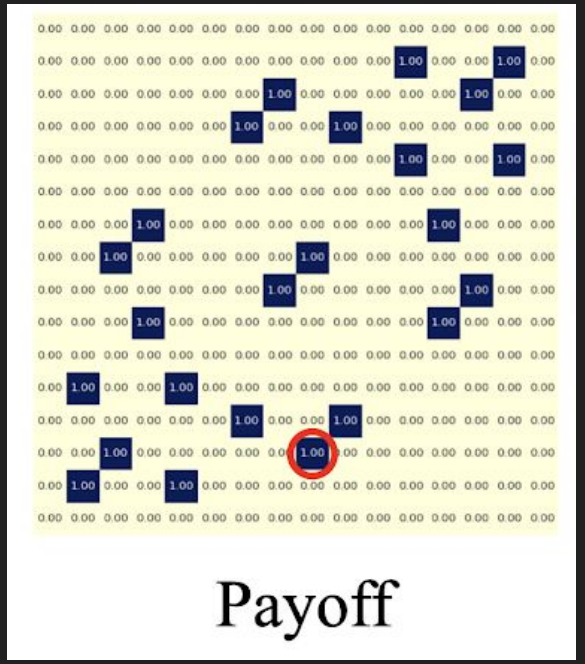




PG-AR (Auto-Regressive)

Multi-modal policy





Training Paradigms

- Multi-step optimization
  minimize calculation

- Randomized execution order
  prevent overfitting

PG-AR on Popular Testbeds

StarCraft

- Alternate attacking with other agents

Google Research Foot-ball

- keeps short passing the ball to other

It took a lot to optimize, but learned new behavior

Conclusion

In multi-modal scenarios, Value-decomposition and parameter sharing can lead unsatisfying behavior.

Policy gradient can learn multi-modal behavior by using auto-regressive.

Thank you for your attention.