InforMARL: Scalable Multi-Agent Reinforcement Learning through Intelligent Information Aggregation

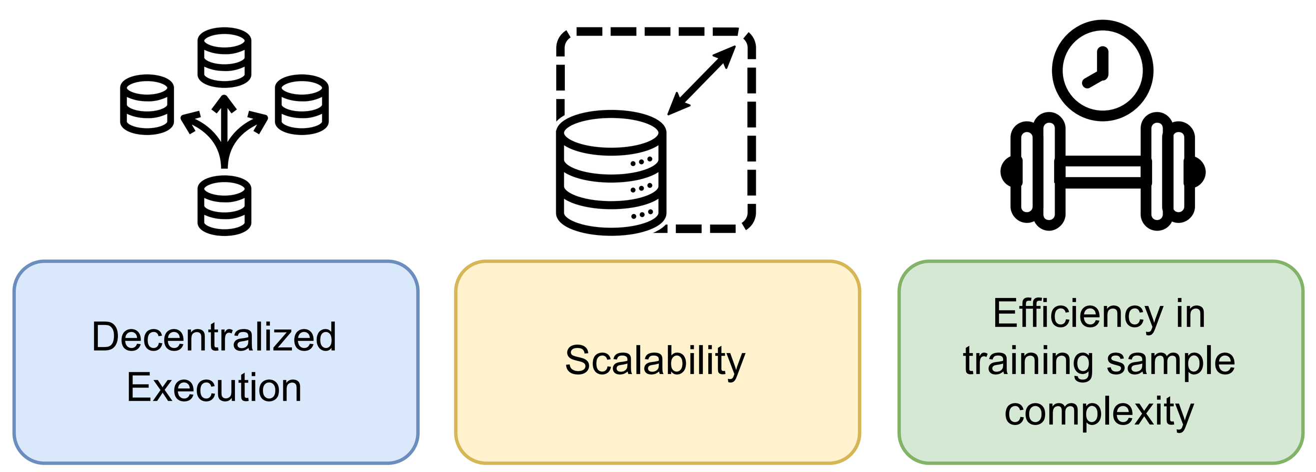
Siddharth Nayak [1] Kenneth Choi [1] Wenqi Ding [1] Sydney Dolan [1] Karthik Gopalakrishnan [2] Hamsa Balakrishnan [1]

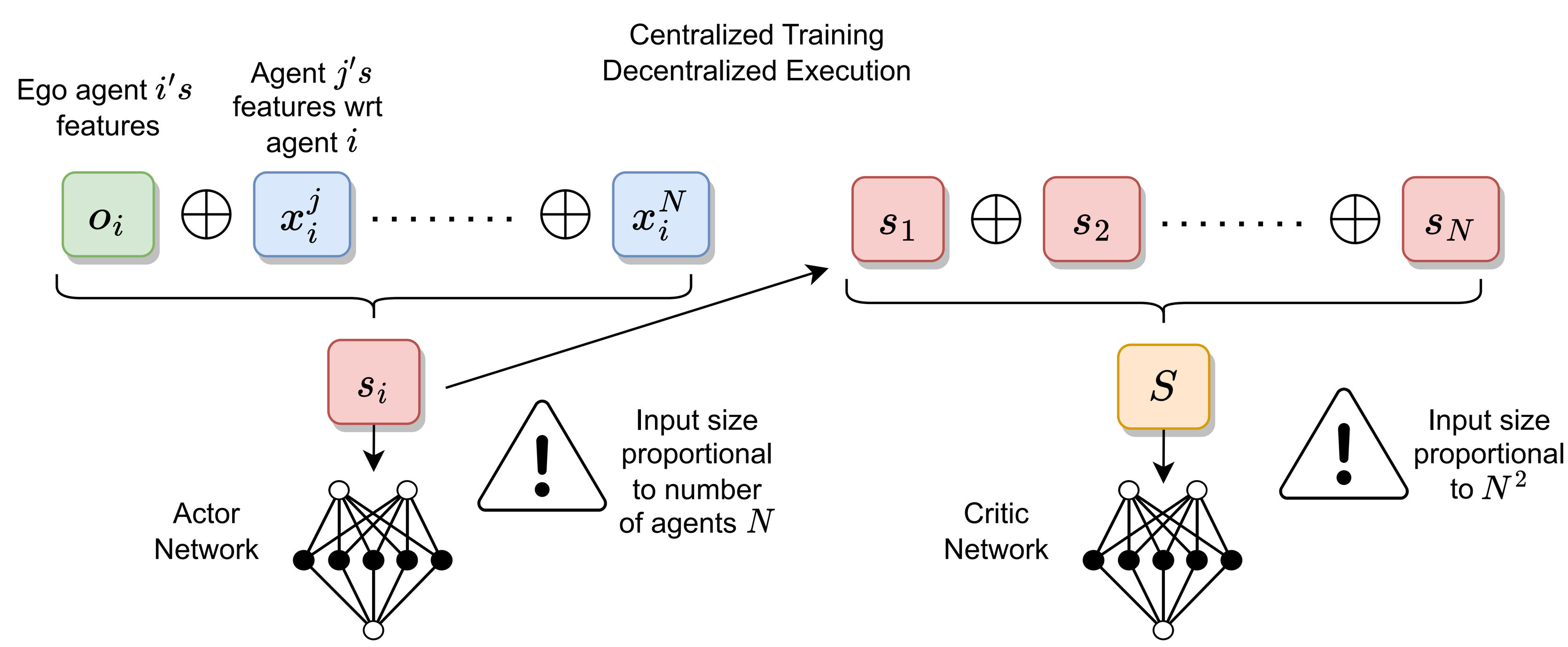
[1]Massachusetts Institute of Technology [2]Stanford University

Project Website

Standard MARL Recipe

Key Features Expected from MARL algorithms:

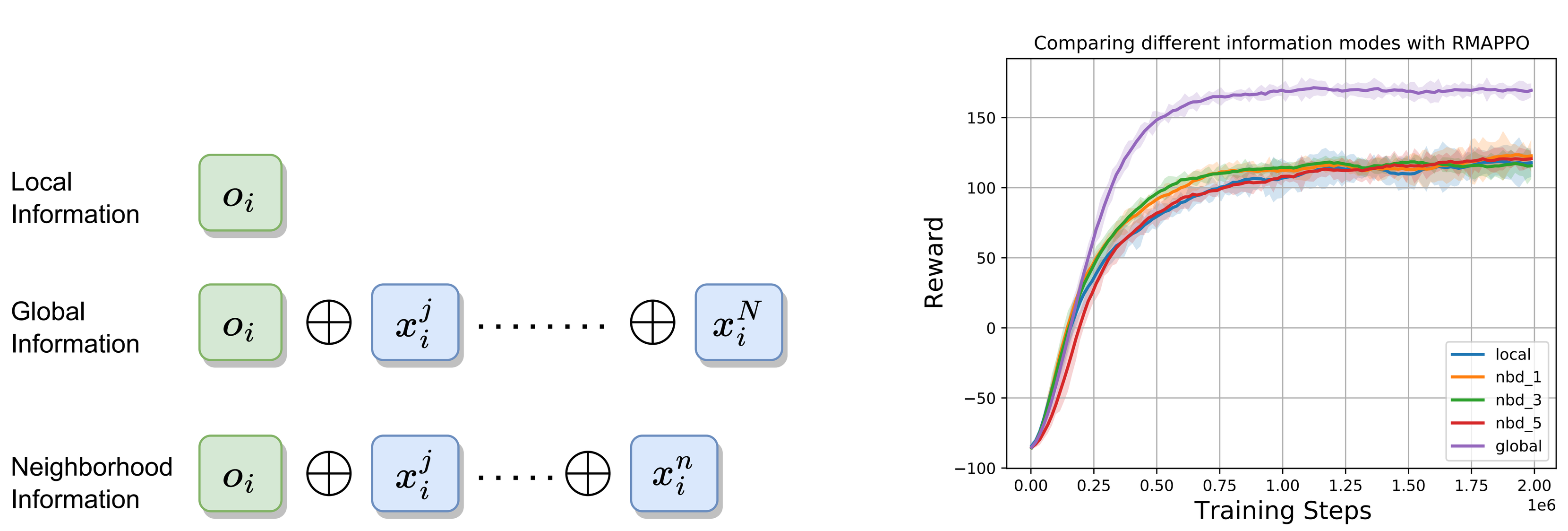




Need a method which is agnostic to number of entities in the environment

Motivation

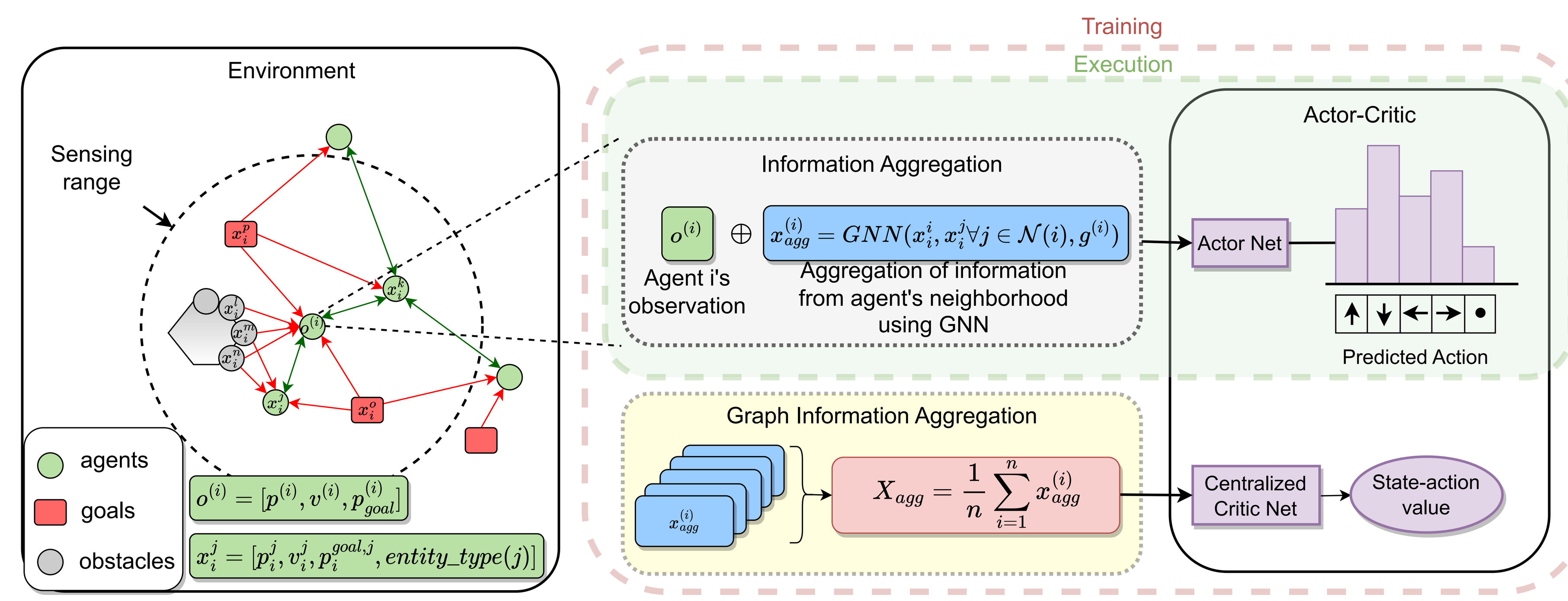
Consider MAPPO (Yu et al. 2022) with different amount of information included as inputs to the actor-critic networks.



(a) Information modes (b) Training MAPPO with different information modes

There is a significant improvement in performance when MAPPO has access to global information.

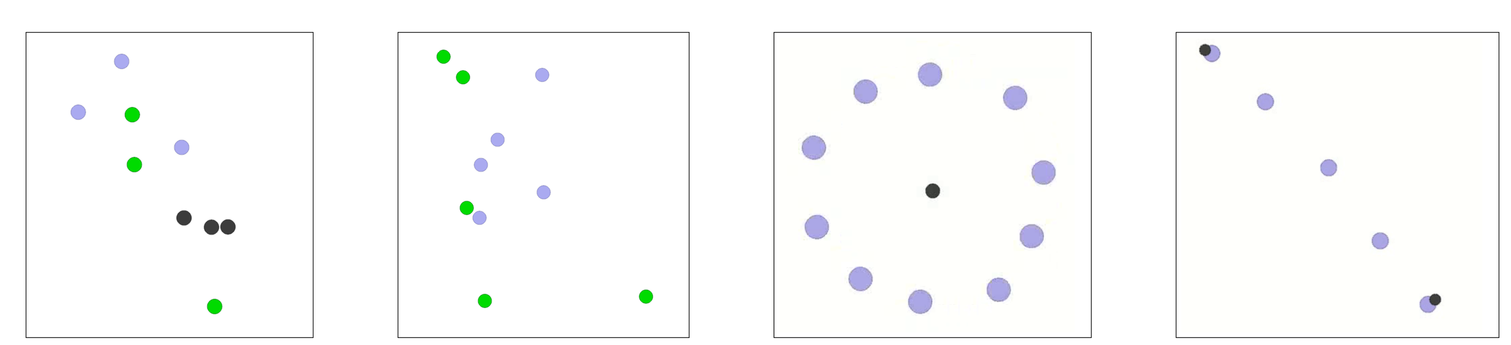
InforMARL



1. **Environment**: The agents are depicted by green circles, the goals by red rectangles, and the unknown obstacles by gray circles. A graph is created by connecting entities within the sensing-radius of the agents. The inter-agent edges are bidirectional, while the edges between agents and non-agent entities are unidirectional.
2. **Information Aggregation**: $x_{agg}^{(i)}$ represents the aggregated information from the neighborhood, which is the output of a GNN. Each agent's observation is concatenated with $x_{\text{agg}}^{(i)}$.
3. **Graph Information Aggregation**: The $x_{\text{agg}}^{(i)}$ from all the agents is averaged to get X_{agg}.
4. **Actor-Critic**: The concatenated vector $[o^{(i)}, x_{\text{agg}}^{(i)}]$ is fed into the actor network to get the action, and X_{agg} is fed into the critic network to get the state-action values.

Task Environments

We perform experiments in 4 different environments: target, coverage, polygon-formation and line-formation environments.



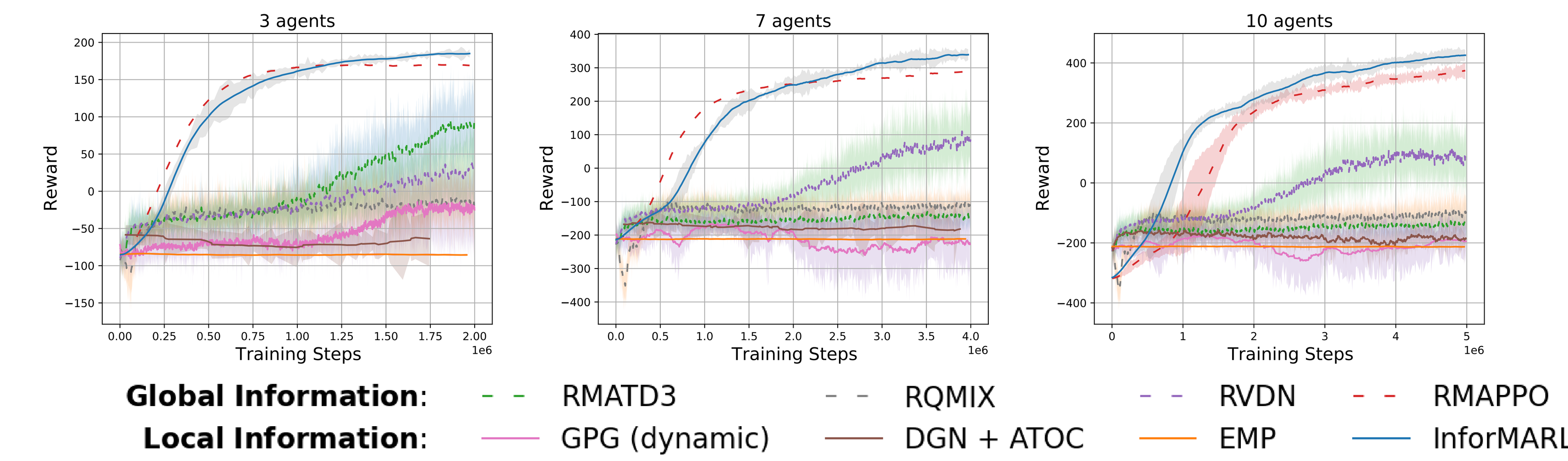
(c) Target (d) Coverage (e) Polygon Formation (f) Line Formation

Conclusions

- InforMARL uses a graph neural network (GNN)-based architecture for scalable multi-agent RL in a decentralized fashion.
- InforMARL is transferable to scenarios with a different number of entities in the environment than what it was trained on.
- InforMARL has better sample complexity than most other standard MARL algorithms with global observations.

Results

Comparison to Baselines



Algorithm	Information mode	$N=3$				$N=10$			
		R	T	# col	$S\%$	R	T	# col	$S\%$
RMATD3	Global	105.49	0.51	3.07	67	-131.72	0.99	11.14	1
RQMIX	Global	19.21	0.77	1.42	28	-76.98	0.96	17.04	2
RVDN	Global	64.04	0.62	1.05	45	157.63	0.64	10.00	43
GPG (dynamic)	Local	−46.27	0.87	0.43	8	-173.53	1.00	4.68	0
DGN + ATOC	Local	67.70	0.66	1.49	35	-201.01	1.00	4.06	0
RMAPPO	Global	173.13	0.41	1.47	96	366.81	0.44	13.21	79
InforMARL	Local	**205.24**	**0.38**	**1.45**	**100**	**429.14**	**0.39**	**10.50**	**100**

InforMARL significantly outperforms most baseline algorithms. Although RMAPPO has similar performance, it requires global information.

Scalability and Performance in different task environments

Test \ Train		$n=3$	$n=7$	$n=10$
$m=7$	Reward/m	61.16	62.23	61.32
	T	0.38	0.40	0.40
	(# col)/m	0.74	0.66	0.70
	$S\%$	100	100	100
$m=10$	Reward/m	58.59	58.23	58.67
	T	0.38	0.40	0.39
	(# col)/m	0.95	0.88	0.87
	$S\%$	100	99	100
$m=15$	Reward/m	53.19	53.46	54.21
	T	0.39	0.40	0.40
	(# col)/m	1.28	1.21	1.20
	$S\%$	100	99	99

Table 1. InforMARL trained and tested in the target environment when then no. of agents is varied.

Environment	m	Metric	Algorithm	
			RMAPPO	InforMARL
Coverage	3	T	0.34	0.36
		$S\%$	100	100
	7	T	0.42	0.43
		$S\%$	100	99
Formation	3	T	0.31	0.30
		$S\%$	100	100
	7	T	0.47	0.43
		$S\%$	100	100
Line	3	T	0.24	0.21
		$S\%$	100	100
	7	T	0.38	0.36
		$S\%$	100	100

Table 2. InforMARL was trained with 3 agents in the environment whereas MAPPO was trained in environments with 3 and 7 agents.

InforMARL is able to achieve a success rate of almost 100% across all scenarios in different environments whilst also being transferable.

This research was sponsored in part by the NASA University Leadership initiative (grant #80NSSC20M0163) and the U.S. AFRL/U.S. Air Force Artificial Intelligence Accelerator (Cooperative Agreement Number FA8750-19-2-1000).
The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of NASA, the U.S. Air Force or the U.S. Government.