Structured Cooperative Multi-agent Coordination

Dissertation Defense

Sheng Li

May 2, 2022

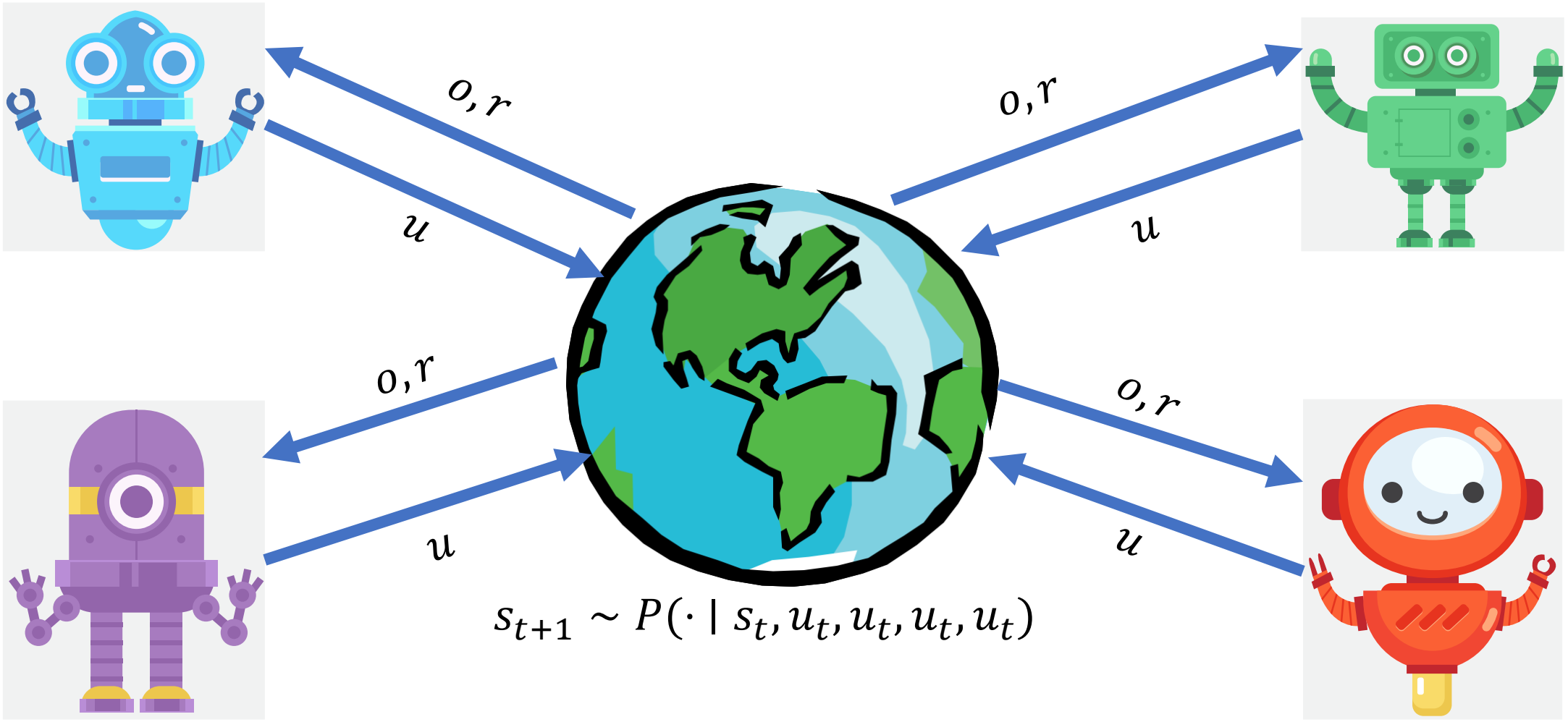
Outline

- Background and Introduction
- Contribution 1: Improving Utility Fusion with Deep Correction
- Contribution 2: Deep Implicit Coordination Graphs
- Contribution 3: Learning Emergent Discrete Communication
- Summary and Future Work

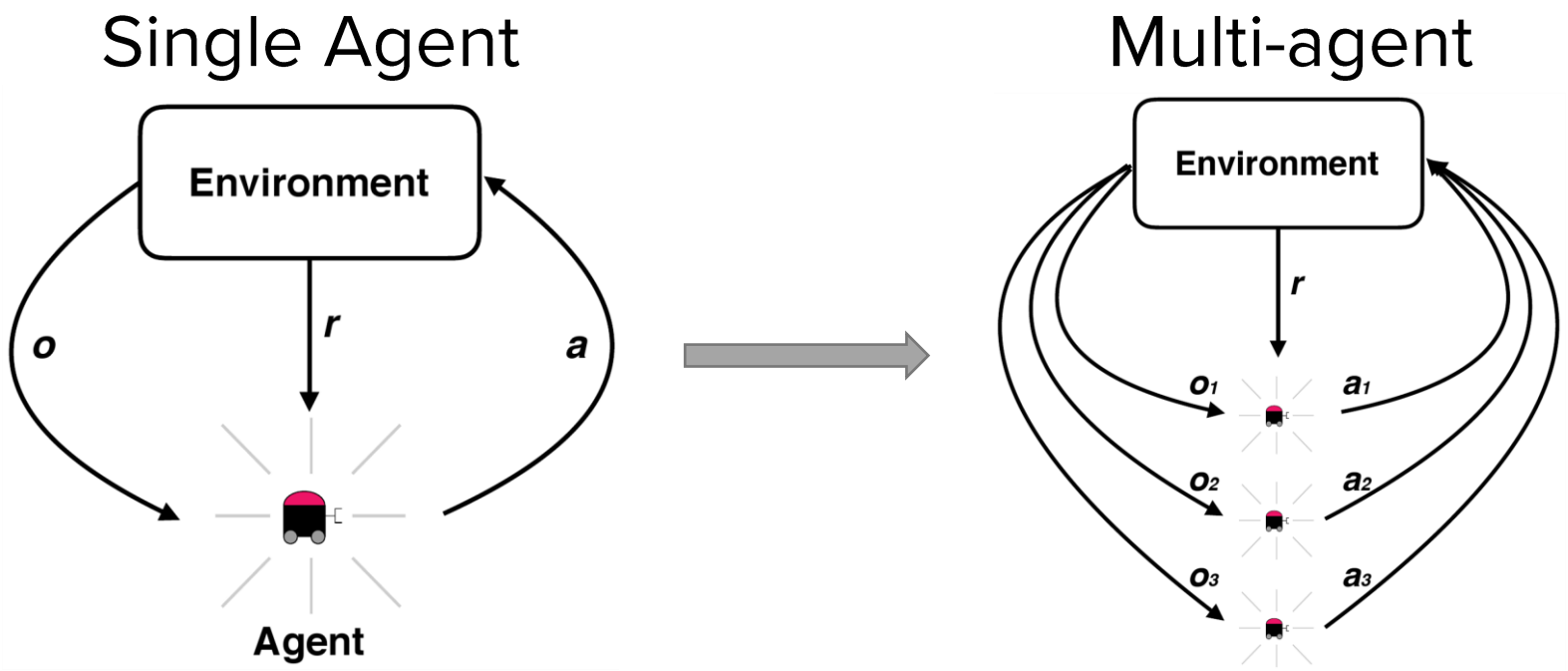
Outline

- Background and Introduction
- Contribution 1: Improving Utility Fusion with Deep Correction
- Contribution 2: Deep Implicit Coordination Graphs
- Contribution 3: Learning Emergent Discrete Communication
- Summary and Future Work

Multi-agent Systems

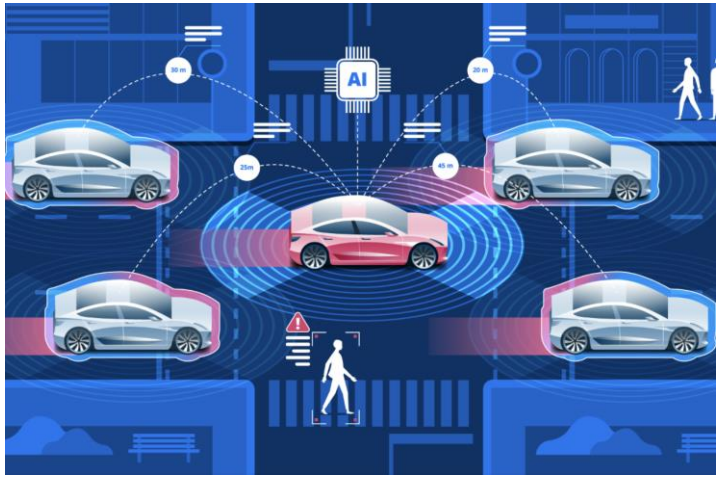


Multi-agent Reinforcement Learning (MARL)



Multiple agents coordinate to achieve a shared objective
Formalized as a Dec-POMDP

Applications of Multi-agent RL



Autonomous driving

Game play

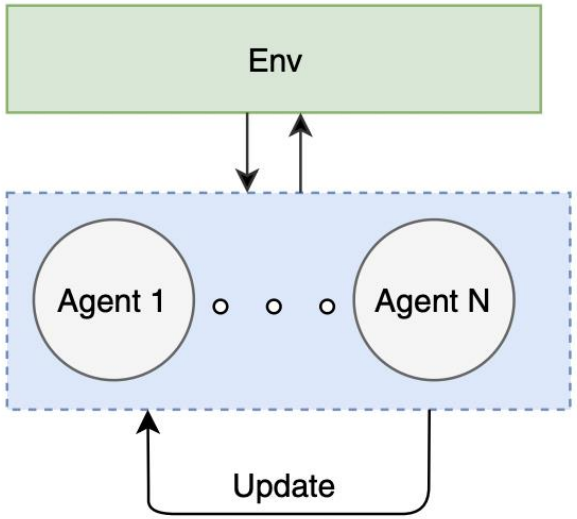
Drone swarm / Collision avoidance

Teamwork that requires coordination

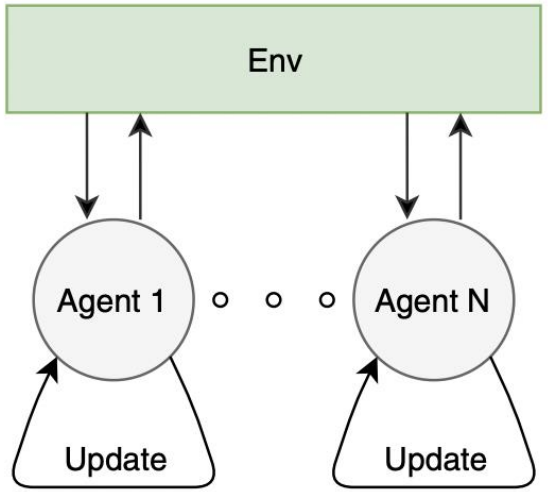
Challenges in Multi-agent RL (MARL)

1. Scalability
   - Train multiple agents concurrently
   - The curse of dimensionality
2. Partial Observability
   - In real world, full observability is rarely available for individual agents
   - Information insufficiency
3. Non-stationarity
   - Other agents are a part of environment for each agent
   - The evolvement of agents causes non-stationary environment dynamics
   - “Relative overgeneralization”

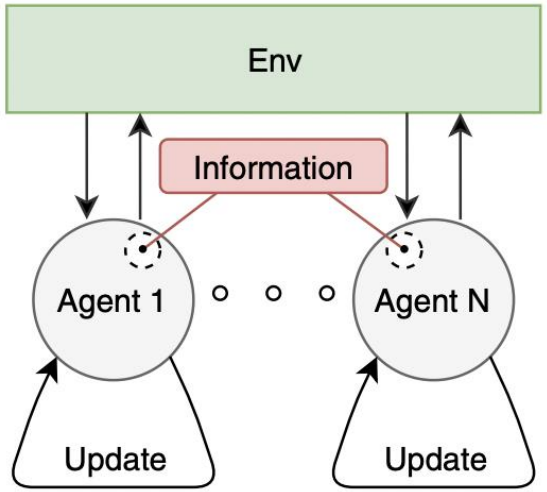
Common MARL Solution Approaches



CTCE: a joint policy for all agents



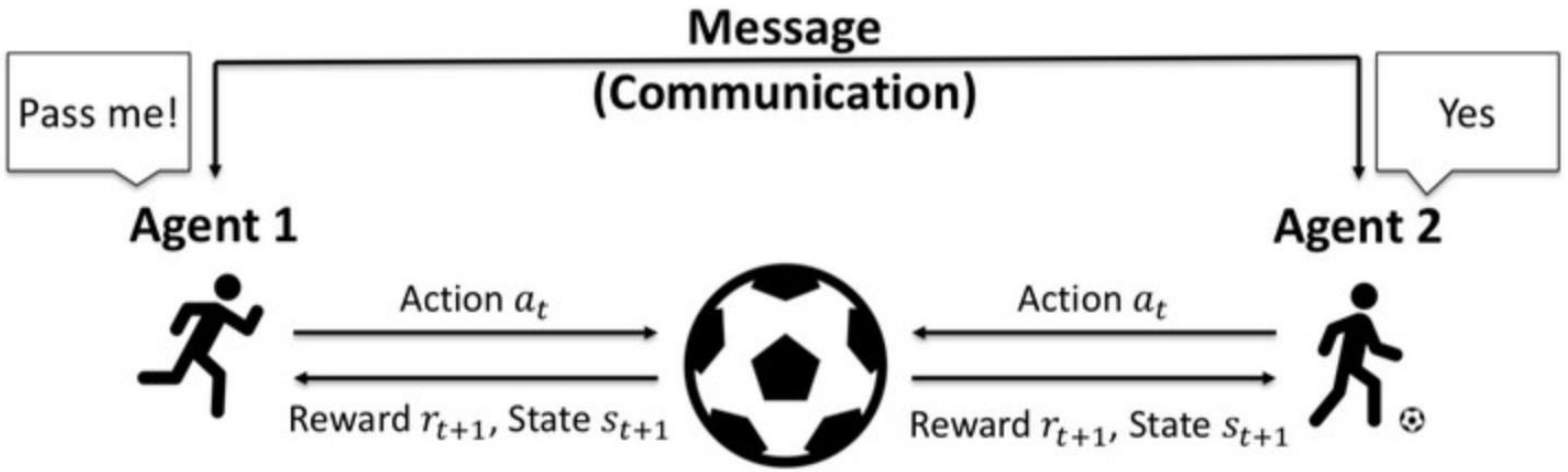
DTDE: Each agent updates its own individual policy



CTDE: agents to exchange information during training which is then discarded at execution time

Gronauer, Sven, and Klaus Diepold. "Multi-agent deep reinforcement learning: a survey." *Artificial Intelligence Review* 55.2 (2022): 895-943.

Network & Communication-based Approaches



- Coordination through communication is the behavioral pattern of humans
- Learning emergent communication protocols for information sharing
  - Observation augmentation
- **Partially** centralized execution

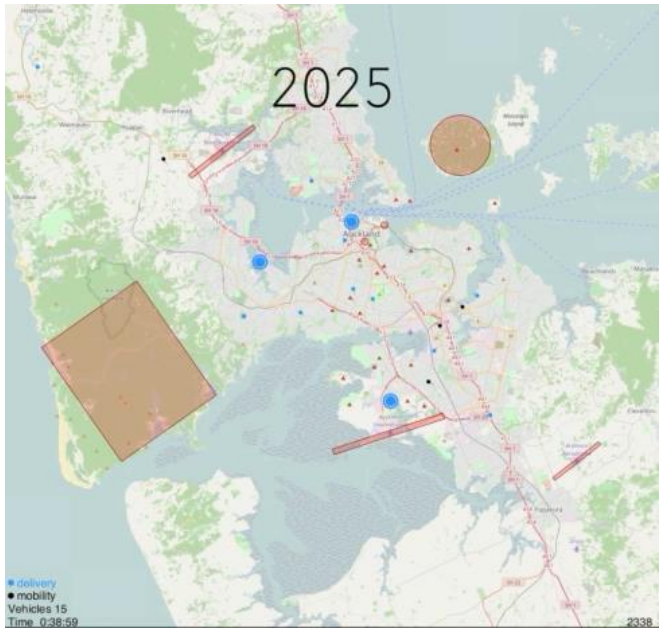
Motivations

- **The Research Question**: Can we find effective and efficient coordination for all the agents in cooperative multi-agent reinforcement learning (MARL)?
- **Centralized** approaches can be optimal but hard to solve, **decentralized** approaches are feasible but often suboptimal, we want to find **balanced** methods
- Exploiting the **physical structures** of MARL problems can give us better solutions
  - Multi-agent value decomposition
  - Graph structure
  - Communication model

Outline

- Background and Introduction
- Contribution 1: Improving Utility Fusion with Deep Correction
- Contribution 2: Deep Implicit Coordination Graphs
- Contribution 3: Learning Emergent Discrete Communication
- Summary and Future Work

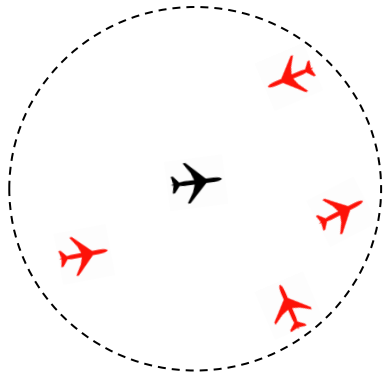
Contribution 1: Optimizing Collision Avoidance in Dense Airspace



- **Problems**
  - Future airspace might become very crowded
  - Existing collision avoidance methods often consider single intruder
  - We need a reliable solution for collision avoidance facing **multiple** reactive intruders
- **Contributions**
  - Solving multi-agent problem from single agent's perspective
  - Using deep correction to improve conventional method

S. Li, M. Egorov, M. J. Kochenderfer, "Optimizing collision avoidance in dense airspace using deep reinforcement learning", in Air Traffic Management Research and Development Seminar, 2019.

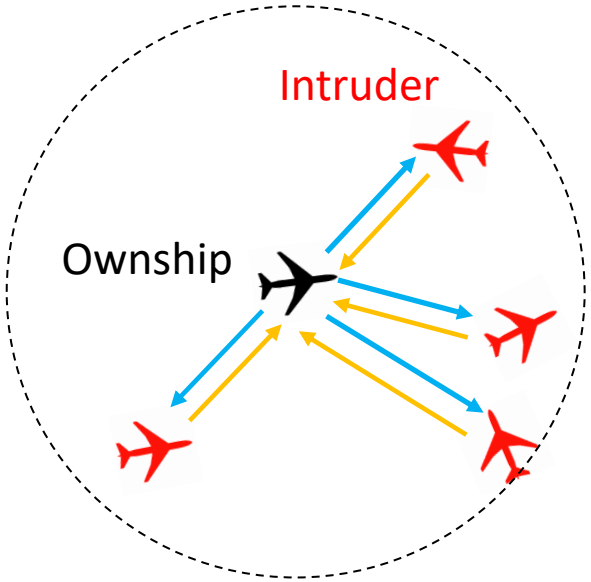
Background & Existing Work

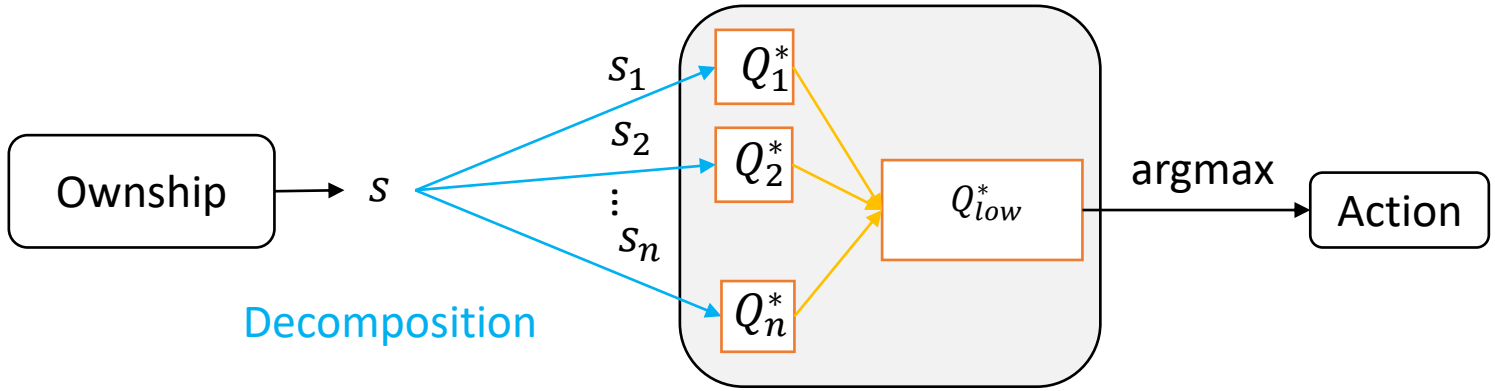


A multi-agent encounter

	ORCA	ACAS X
Principle	Geometry based	Value table based
Advantages	• Fast • Smooth	• Robust • Safe • Can handle uncertainties
Disadvantages	• Hard to tune • Sensitive to uncertainties • Sometimes infeasible	• Optimized for pairwise encounters • Possibly over-conservative
	Not optimized for dense airspace	

Utility Decomposition and Fusion



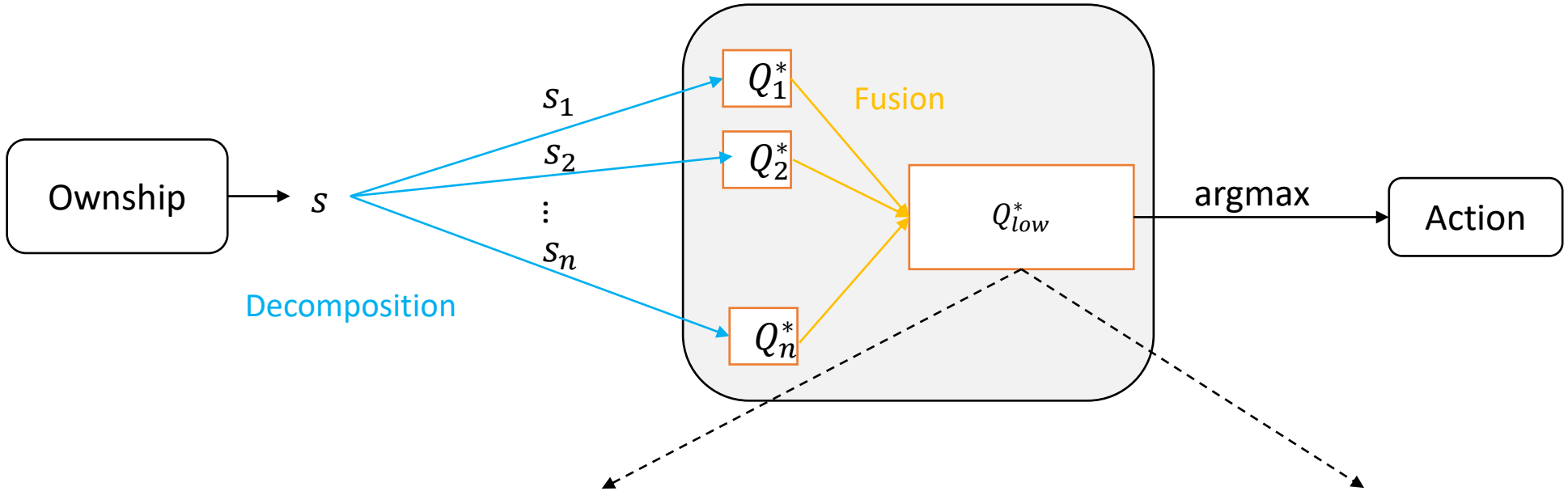


Utility Decomposition: dividing the encounter into pairwise encounters

Utility Fusion: "adding-up" pairwise utilities to decide on safe actions

Each pairwise Collision Avoidance policy is solved by Value Iteration => **VICAS**

VICAS for Multi-intruders



VICASClosest: using the closest intruder

$$Q_{low}^*(s,a) \approx Q_i^*(s_i,a)$$
$$i = \arg\min_{j\in\{1,\dots,n\}} dist_j$$

- A very rough approximation

VICASMulti: using all the n intruders

$$Q_{low}^*(s,a) \approx \min_{i\in\{1,\dots,n\}} Q_i^*(s_i,a)$$

- Considers the most dangerous intruder for each action, risk averse

Deep Correction and Why

Deep Correction:

- Multi-fidelity optimization
- The high-fidelity model (f_{hi}) is expensive to evaluate
- Use a surrogate model :

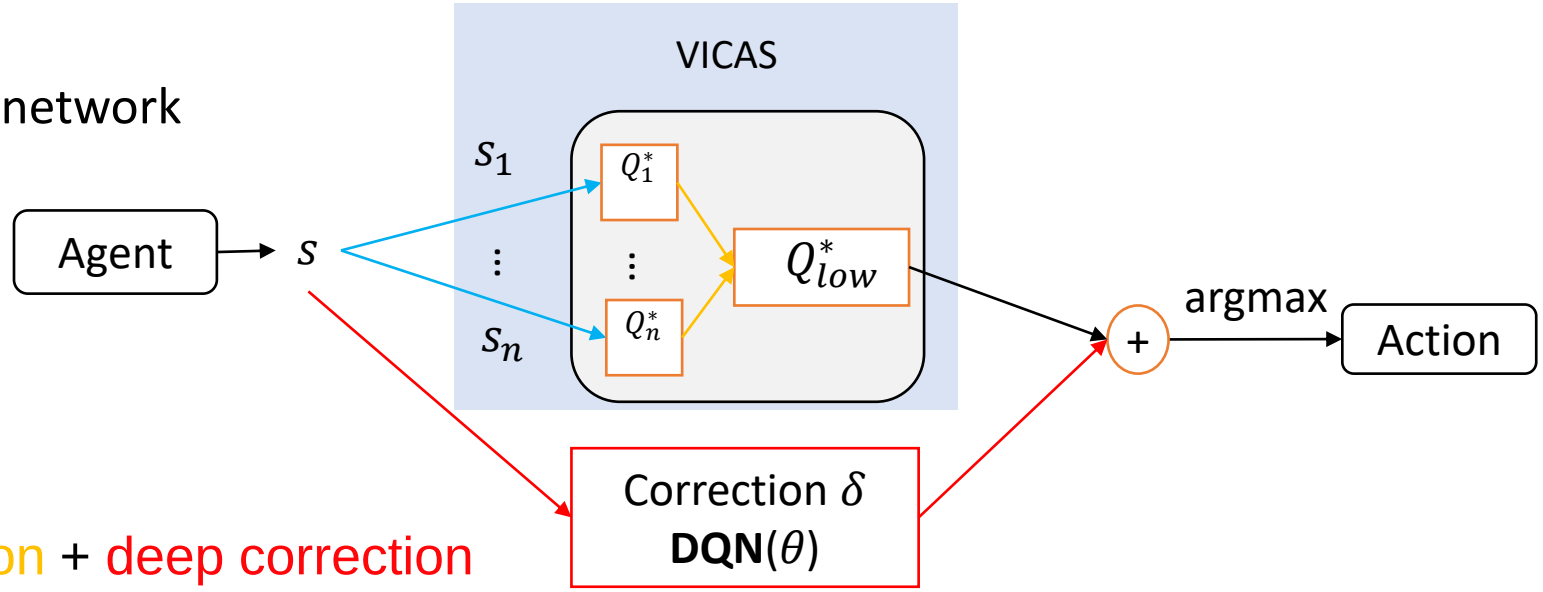
$$f_{hi} \approx f_{low} + \delta$$

$$Q^* \approx Q^*_{low} + \delta$$

- Correction δ is a deep Q-network

Why?

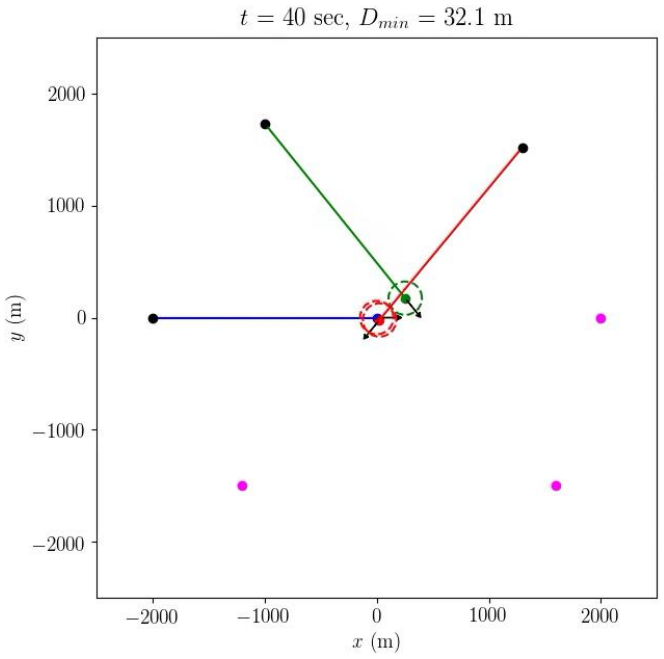
- Q^* is hard to optimize
- Q^*_{low} is easy to obtain
- Deep Q-network is powerful



Utility decomposition / fusion + deep correction

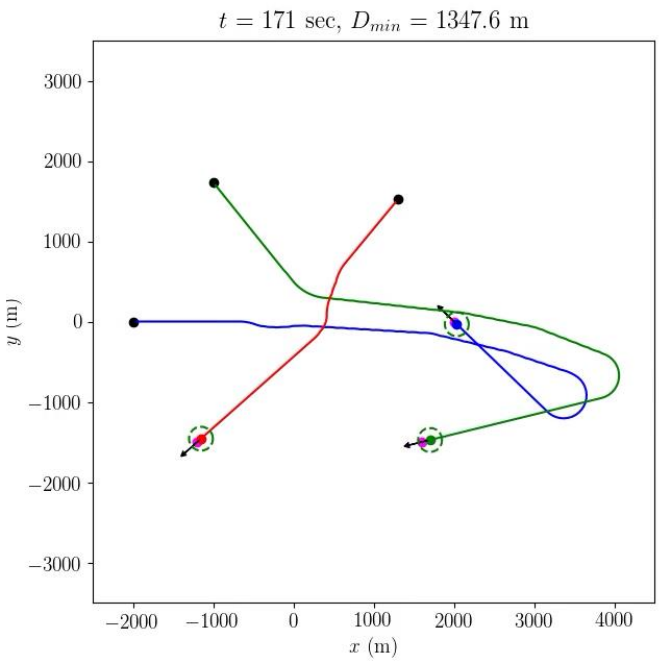
Results of Conventional VICAS

No Collision Avoidance



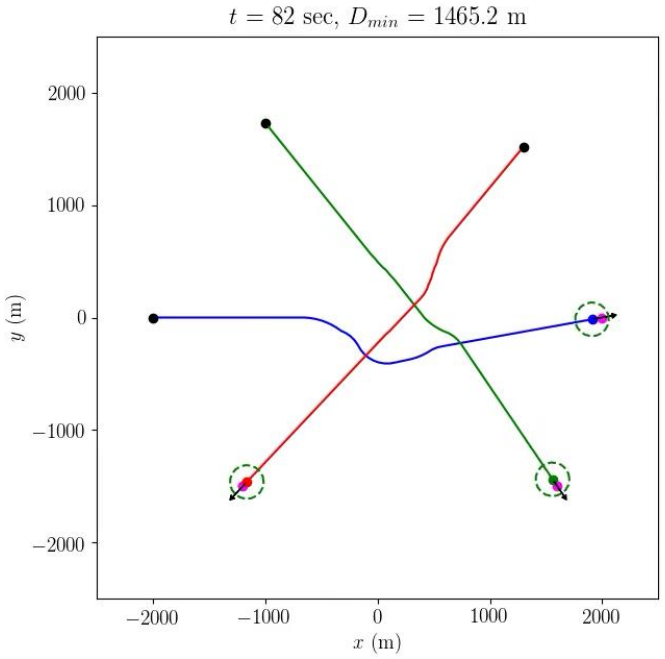
An encounter with collision

VICASMulti



Winding routes and oscillations in actions

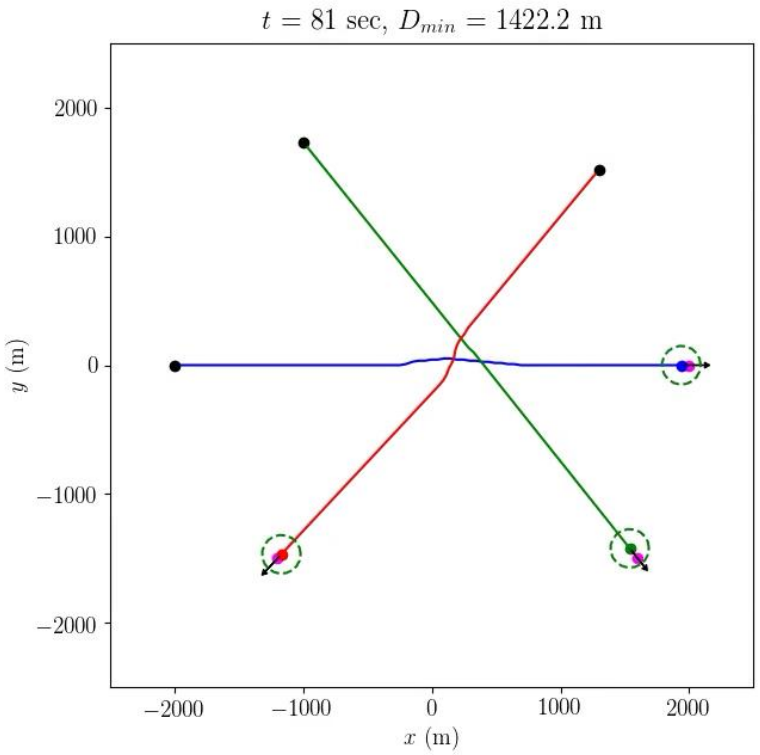
VICASClosest



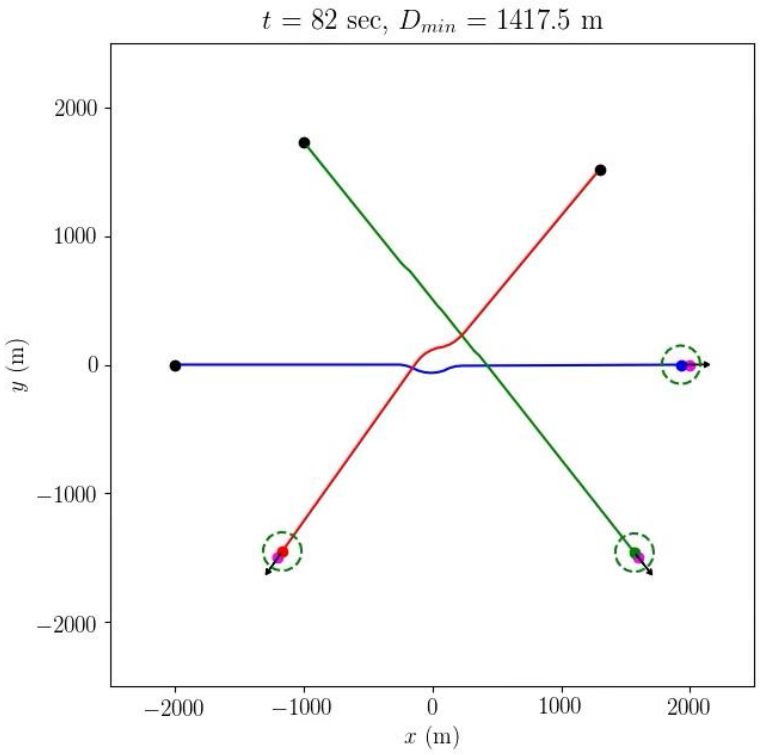
Less winding routes

Results of Deep Correction

Corrected

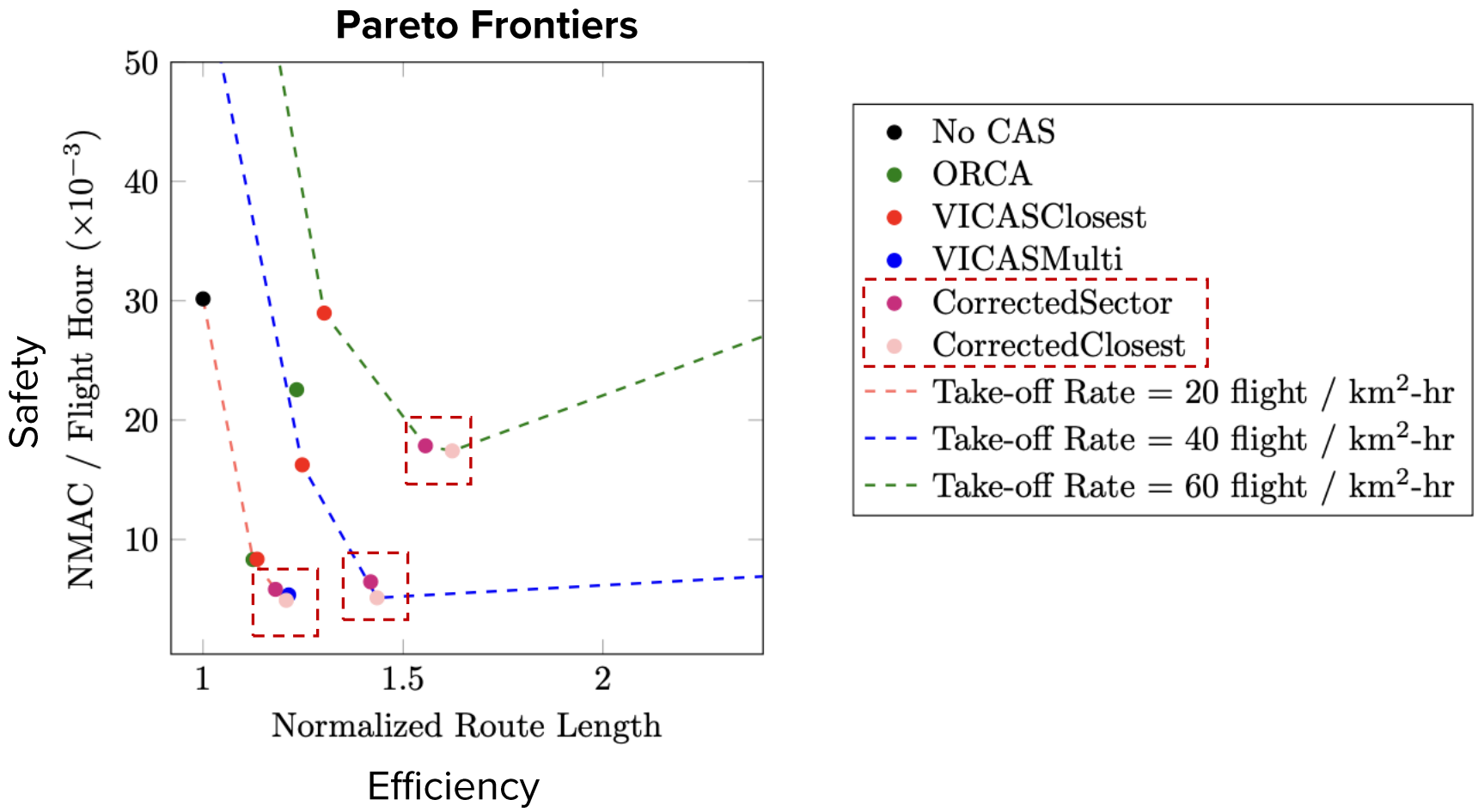


More Conservative Corrected



Avoiding collision with minimal maneuvers and straightforward paths

Safety and Efficiency Trade-off



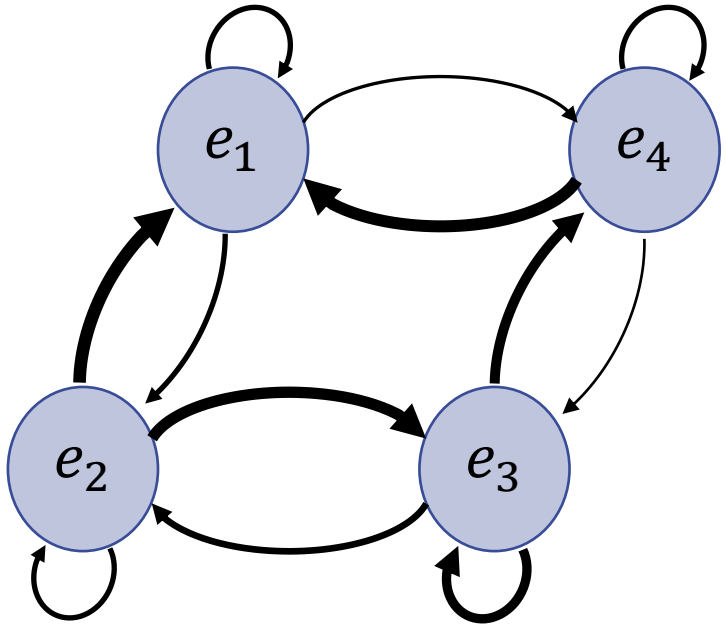
Discussion

- Using utility decomposition and fusion to solve a multi-agent problem from single agent's perspective
- Using a deep Q-network as correction term
- Trained by deep Q-learning
- Both safety and efficiency are improved in dense airspace

Outline

- Background and Introduction
- Contribution 1: Improving Utility Fusion with Deep Correction
- Contribution 2: Deep Implicit Coordination Graphs
- Contribution 3: Learning Emergent Discrete Communication
- Summary and Future Work

Contribution 2: Deep Implicit Coordination Graphs (DICG)



- **Problems**
  - Agents need communication to coordinate
  - Existing approaches use hard-coded ways or distance thresholding
- **Contributions**
  - Using self-attention to dynamically build soft-edged implicit coordination graphs
  - Using graph convolution networks to pass information between agents

S. Li, J. K. Gupta, P. Morales, R. Allen, and M. J. Kochenderfer, “Deep implicit coordination graphs for multi-agent reinforcement learning,” in International Conference on Autonomous Agents and Multiagent Systems (AAMAS), 2021.

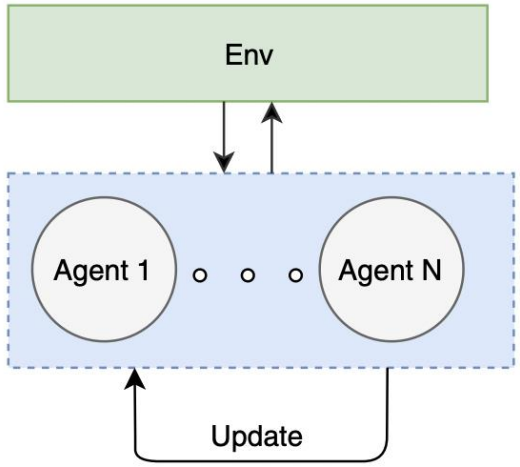
Background & Existing Work

- Parameter-sharing decentralized approaches
  - PS-TRPO: Fast, scalable but suboptimal due to partial observability
- Value function decomposition
  - VDN, Q-MIX
  - Reasoning about joint value function by decomposition
  - No cross-agent information passing
- Distance thresholding or hard-coded graph building
  - DCG: formal reasoning about joint actions given the graph structure
  - DGN, ATOC: using distance-based graph for communication

Recap: Two Extremes of MARL

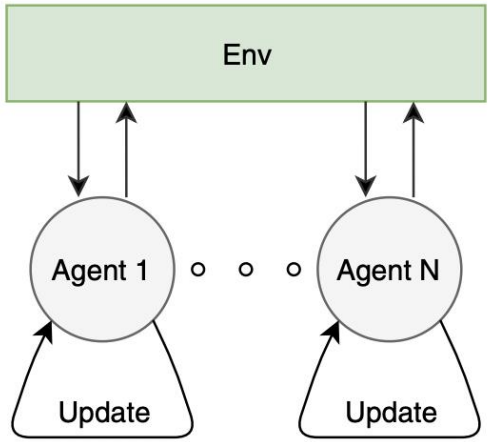
Fully Centralized

- Optimal solutions
- Better coordination
- Action space explodes



Fully Decentralized

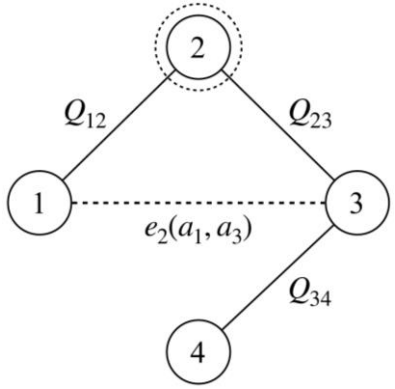
- Action space in check
- Suboptimal solutions
- Non-stationarity during learning



Preliminary: Coordination Graphs

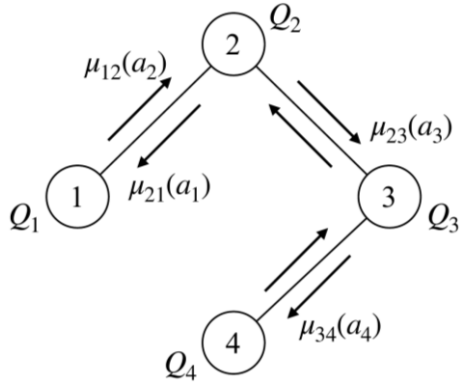
- Interactions between agents usually local and sparse
- Locality of interaction effectively encoded with coordination graphs

$$Q(a) = Q_{12}(a_1, a_2) + Q_{23}(a_2, a_3) + Q_{34}(a_3, a_4)$$



Variable Elimination

$$Q(a) = \sum_{i \in V} Q_i(a_i) + \sum_{i,j \in E} Q_{ij}(a_i, a_j)$$



Max Plus

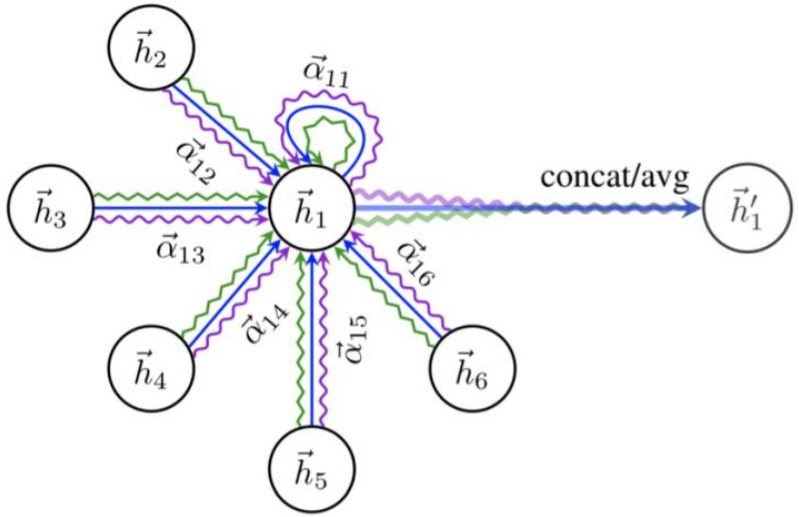
Best action selection from factored representations via Bayesian network inference inspired algorithms

1. Guestrin, Carlos, Daphne Koller, and Ronald Parr. "Multiagent planning with factored MDPs." NeurIPS. 2002.
2. Kok, Jelle R., and Nikos Vlassis. "Using the max-plus algorithm for multiagent decision making in coordination graphs." Robot Soccer World Cup. 2005.

Preliminary: Self-attention for Graph-structured Data

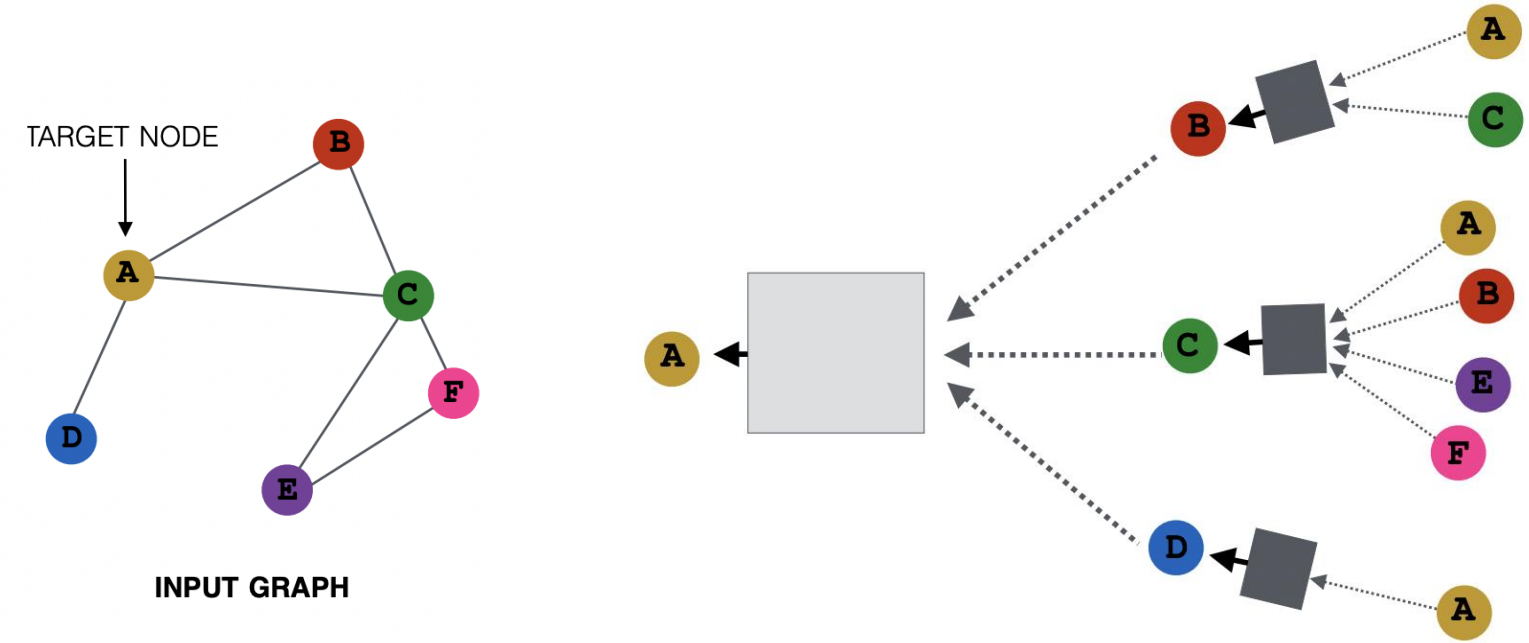
- Attention allows focus on parts of the inputs
- Form of content-based addressing
  - Given queries and a set of keys, find similar keys based on some distance metric

- For graphs, interaction/coordination between nodes can be modeled as attention weights

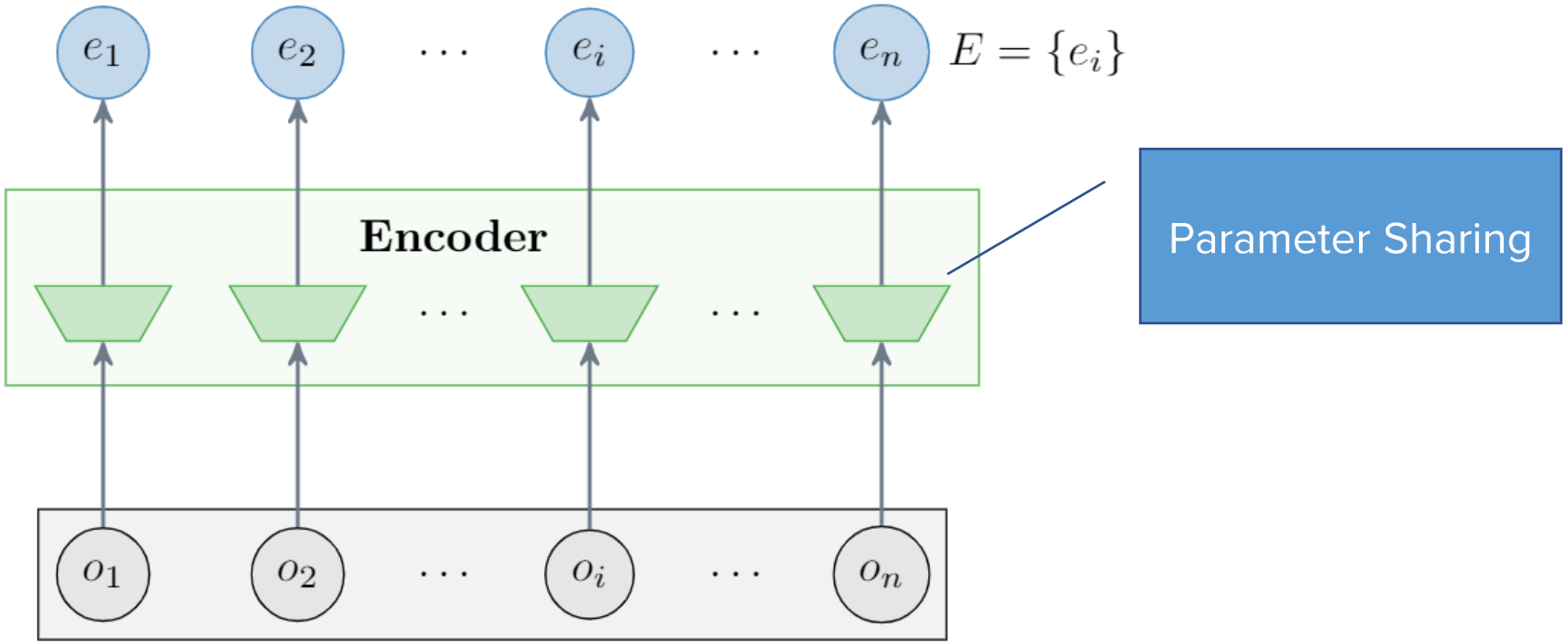


Preliminary: Graph Convolution Networks (GCN)

- **Given:** graph structure as adjacency matrix and features of individual nodes as feature matrix
- **Idea:** Pass messages between pairs of nodes by aggregation and in the process refine node representations



Step 1: Encode observations

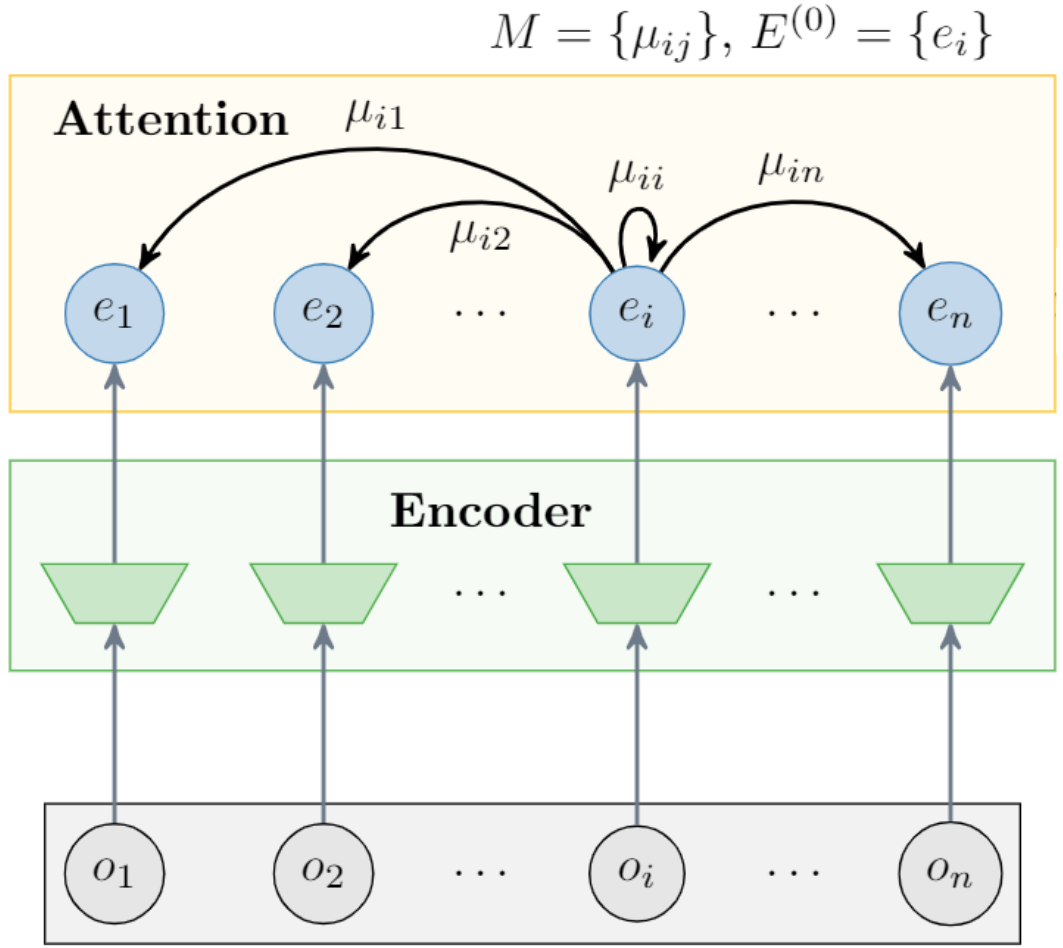


Step 2: Obtain Graph weights via Attention

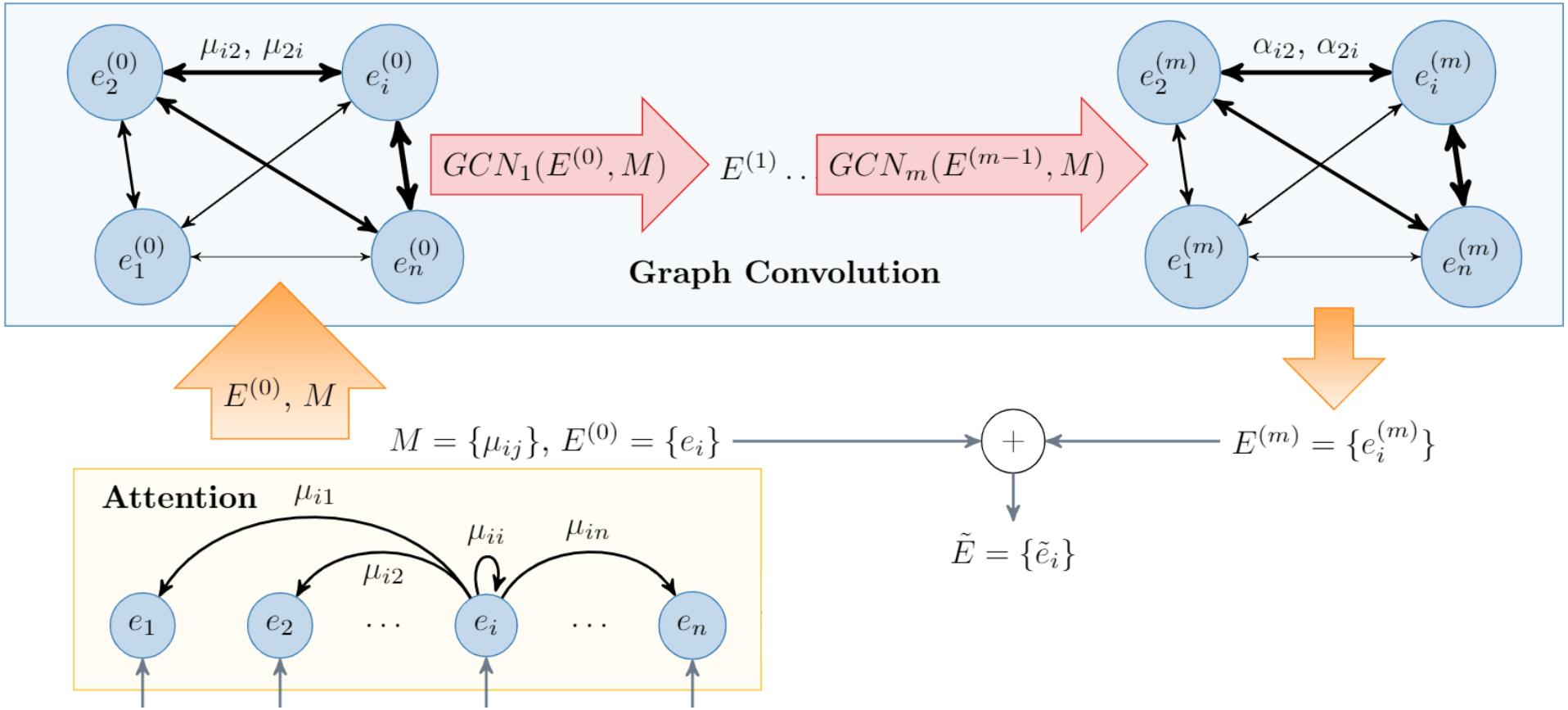
"soft"-edged Coordination Graph

$$\text{Attention}(e_i, e_j, W_a) = e_j^\top W_a e_i$$

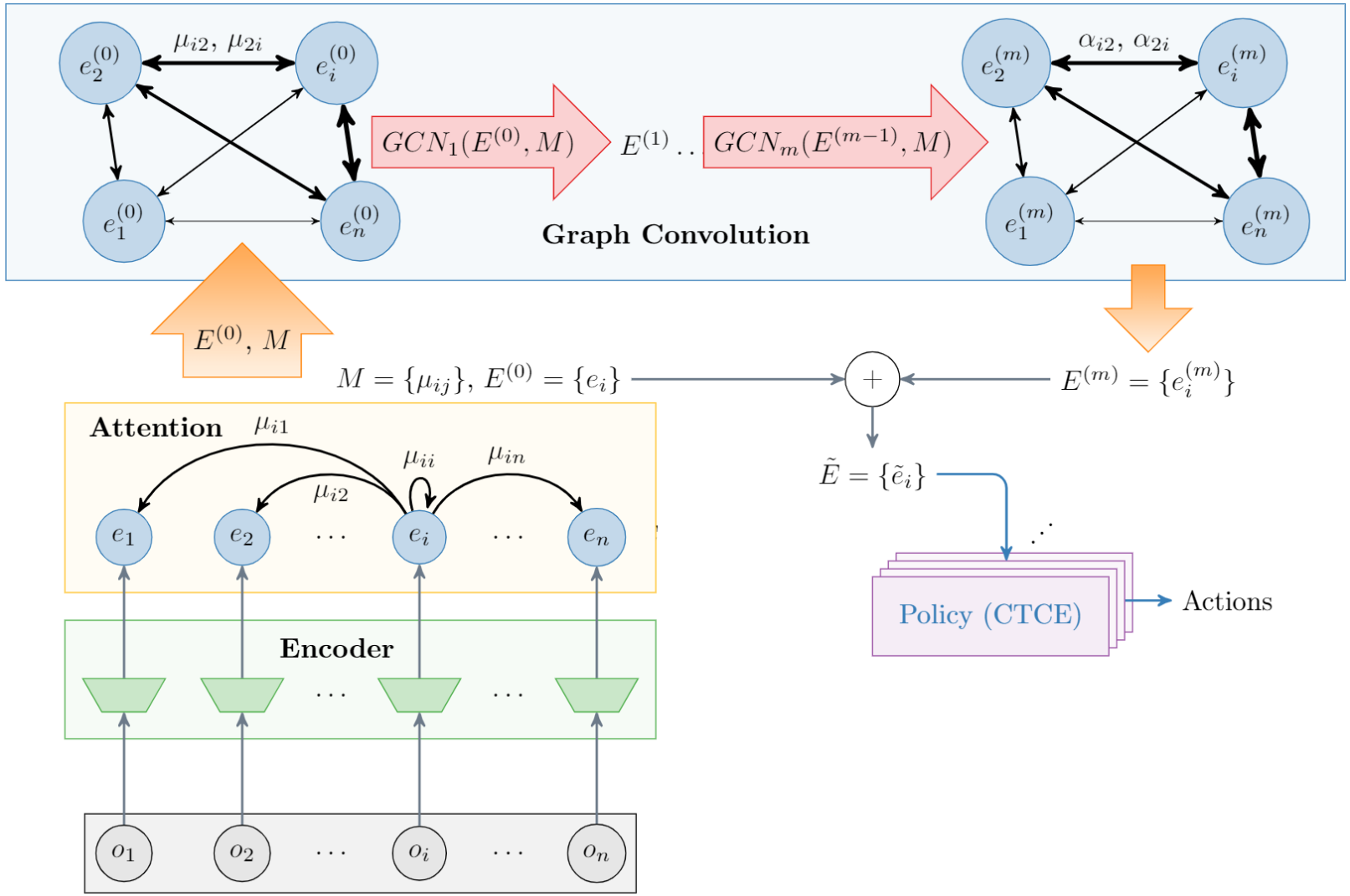
$$\mu_{ij} = \frac{\exp(\text{Attention}(e_i, e_j, W_a))}{\sum_{k=1}^{n} \exp(\text{Attention}(e_i, e_k, W_a))}$$



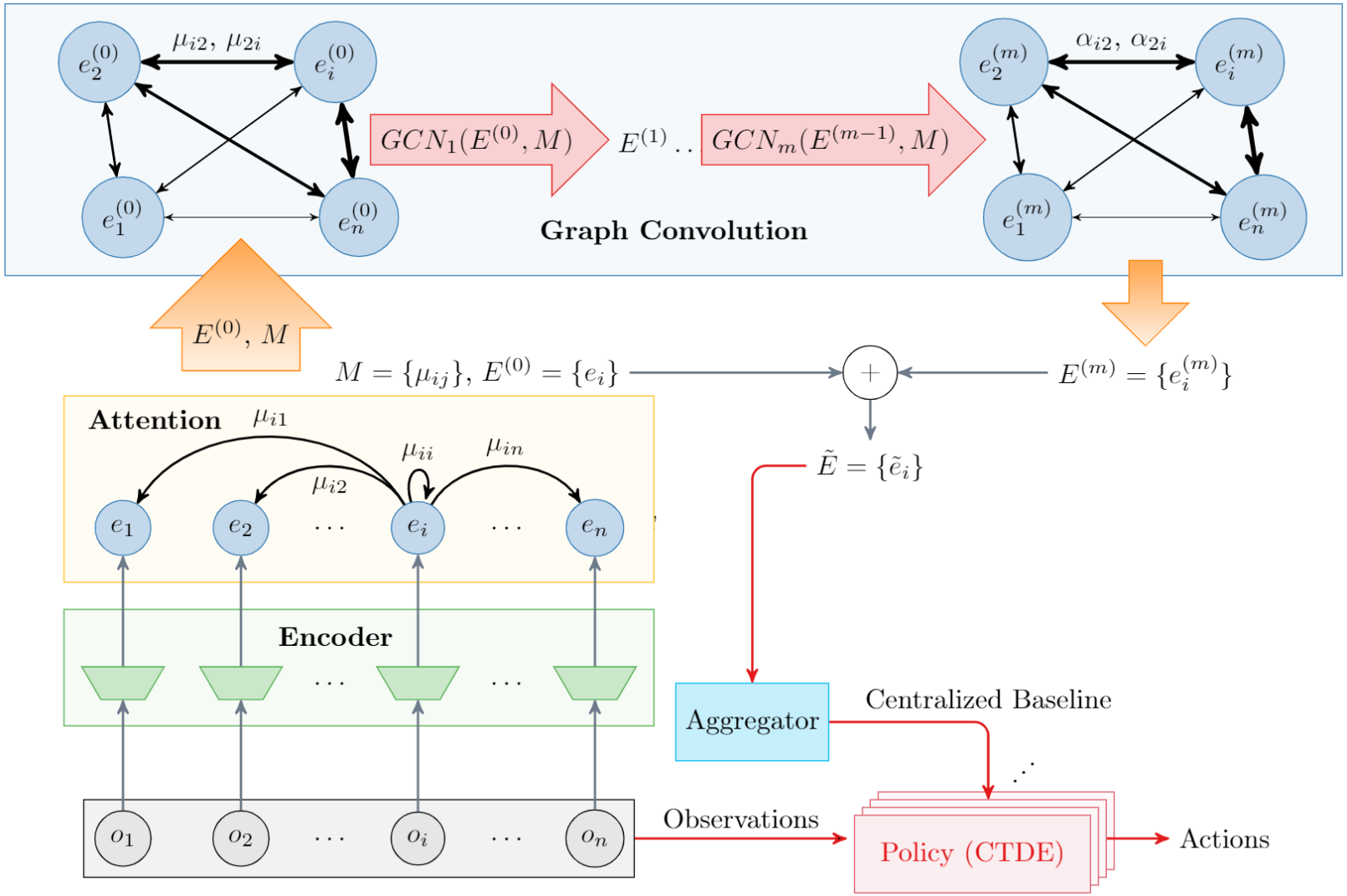
Step 3: Process observation embeddings with Graph Neural Net



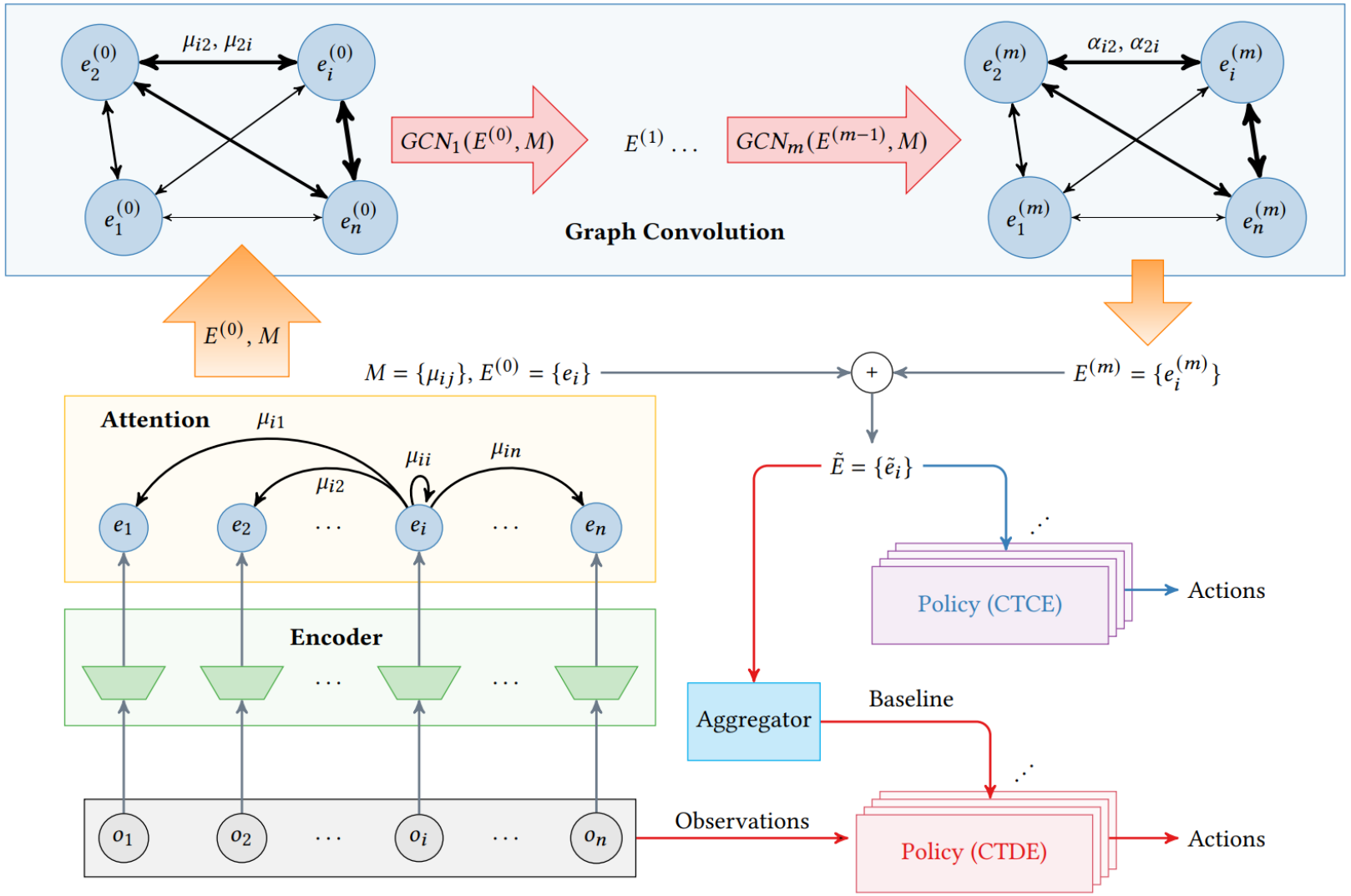
Step 4a: Train policies directly for centralized execution



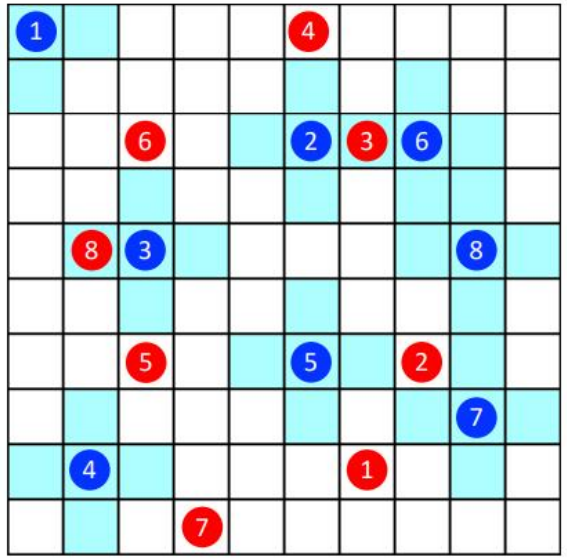
Step 4b: Train Centralized Critic for decentralized execution



Deep Implicit Coordination Graph



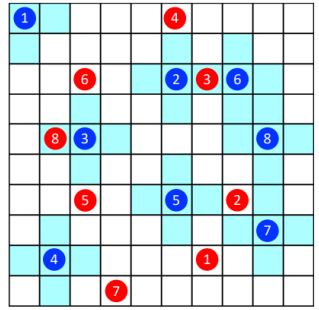
Experiment Domains

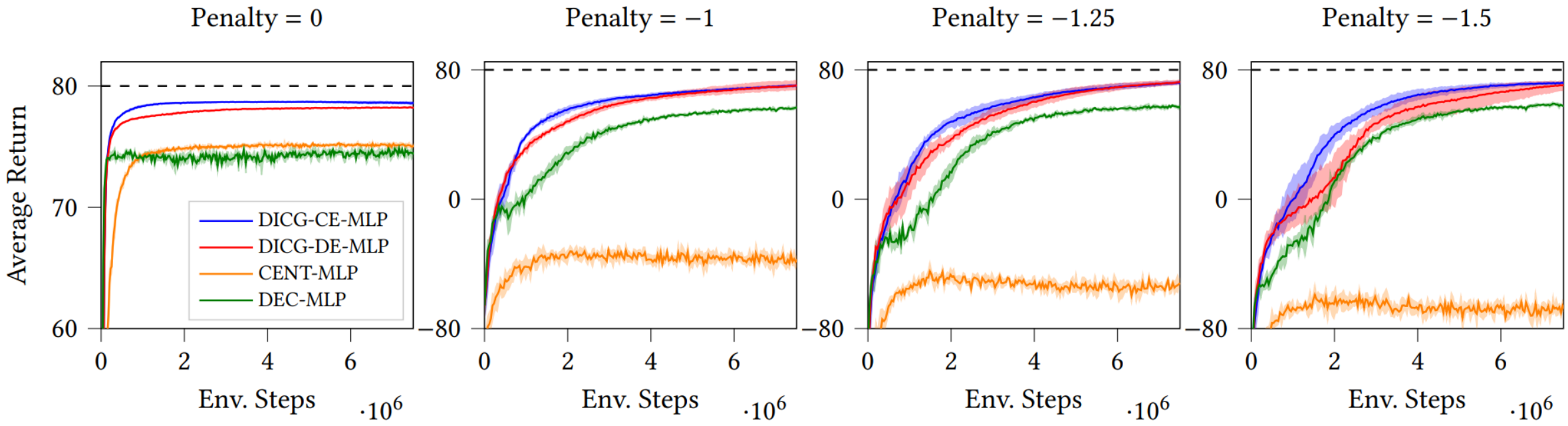


Predator-Prey

StarCraft II Multi-agent Challenges (SMAC)

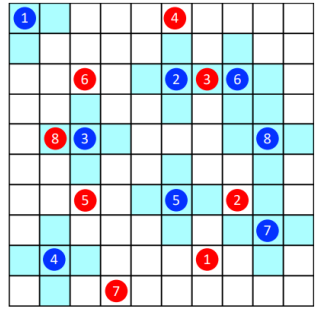
Results: Predator-Prey

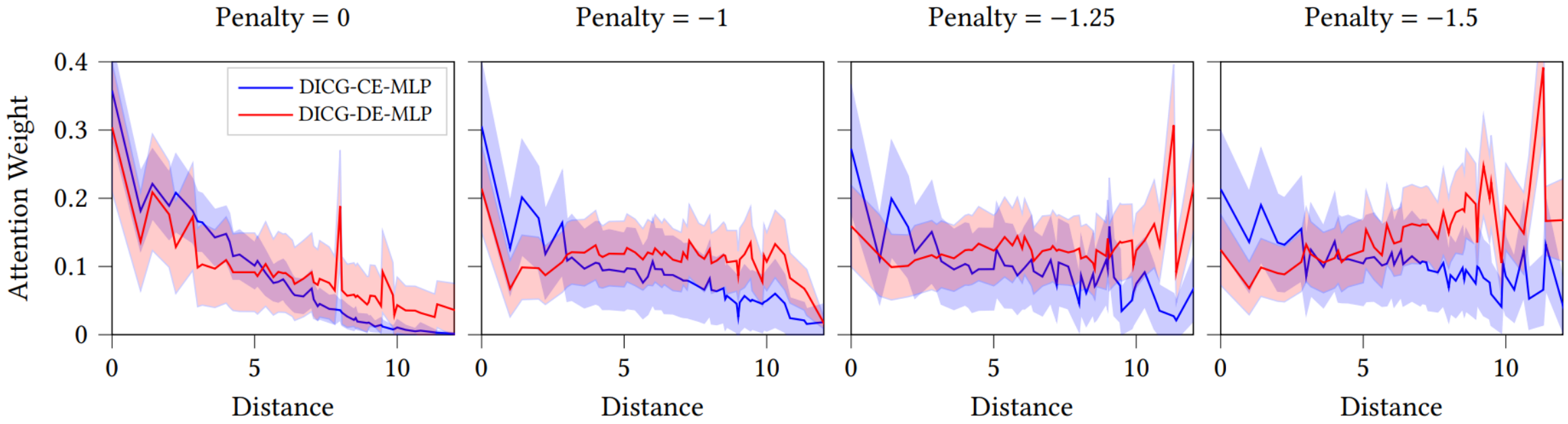




DICG outperforms in both centralized and decentralized execution scenario, under a variety of coordination requirements

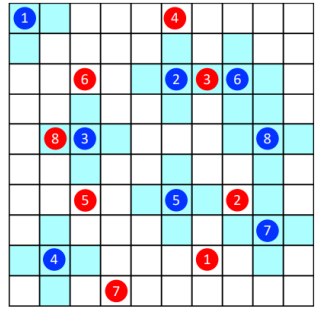
Correlation of attention weight distribution with coordination

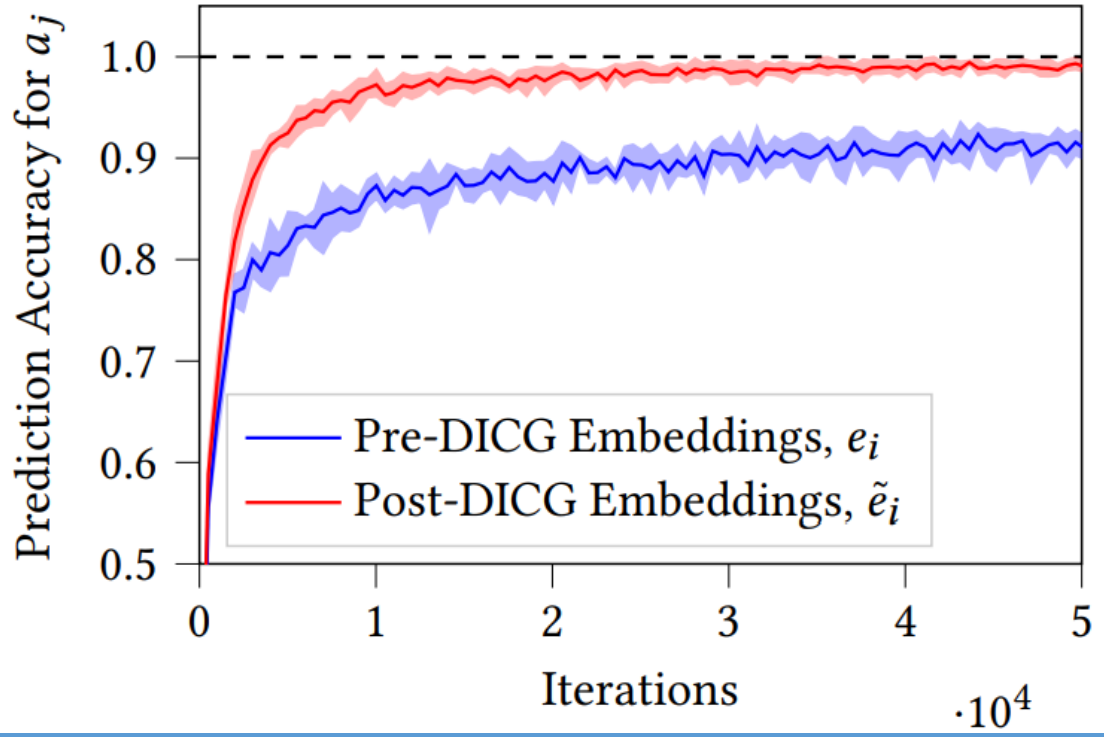




As the coordination requirements get stricter, more attention weight is focused on agents that are further off

Effectiveness of Graph Neural Network for Information Integration





Embeddings obtained after processing with GCN more useful for predicting other agent's actions (intentions)

SMAC Demo: DICG-DE, divide-and-conquer

SMAC Results

Approach	8m_vs_9m	3s_vs_5z	6h_vs_8z
DCG (Böhmer et al., 2020)	$55 \pm 10\%$	$85 \pm 3\%$	$\mathbf{10 \pm 5\%}$
VDN (Sunehag et al., 2018)	$49 \pm 5\%$	$72 \pm 10\%$	0
QMIX (Rashid et al., 2018)	$60 \pm 11\%$	$95 \pm 1\%$	$5 \pm 5\%$
CENT-LSTM	$42 \pm 6\%$	0	0
DEC-LSTM	$65 \pm 16\%$	$94 \pm 5\%$	0
DICG-CE-LSTM	$72 \pm 11\%$	$96 \pm 3\%$	$9 \pm 9\%$
DICG-DE-LSTM	$\mathbf{87 \pm 6\%}$	$\mathbf{99 \pm 1\%}$	0

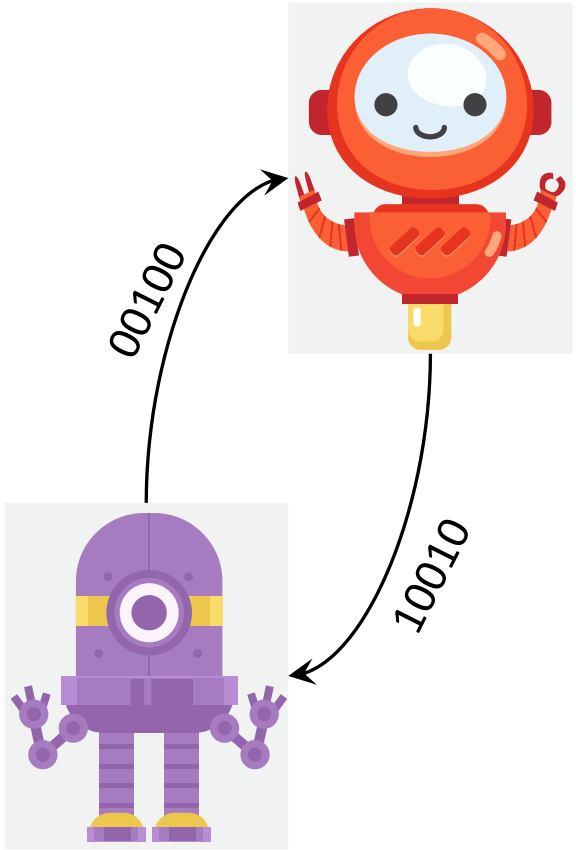
Discussion

- Inspired by coordination graph-based action inference, we designed an effective combination of self-attention and graph neural network modules for multi-agent reinforcement learning
- Proposed approach works for both policy learning as well as centralized critic learning
- State-of-the-art performance on a variety of domains with high coordination

Outline

- Background and Introduction
- Contribution 1: Improving Utility Fusion with Deep Correction
- Contribution 2: Deep Implicit Coordination Graphs
- Contribution 3: Learning Emergent Discrete Communication
- Summary and Future Work

Contribution 3: Learning Emergent Discrete Message Communication



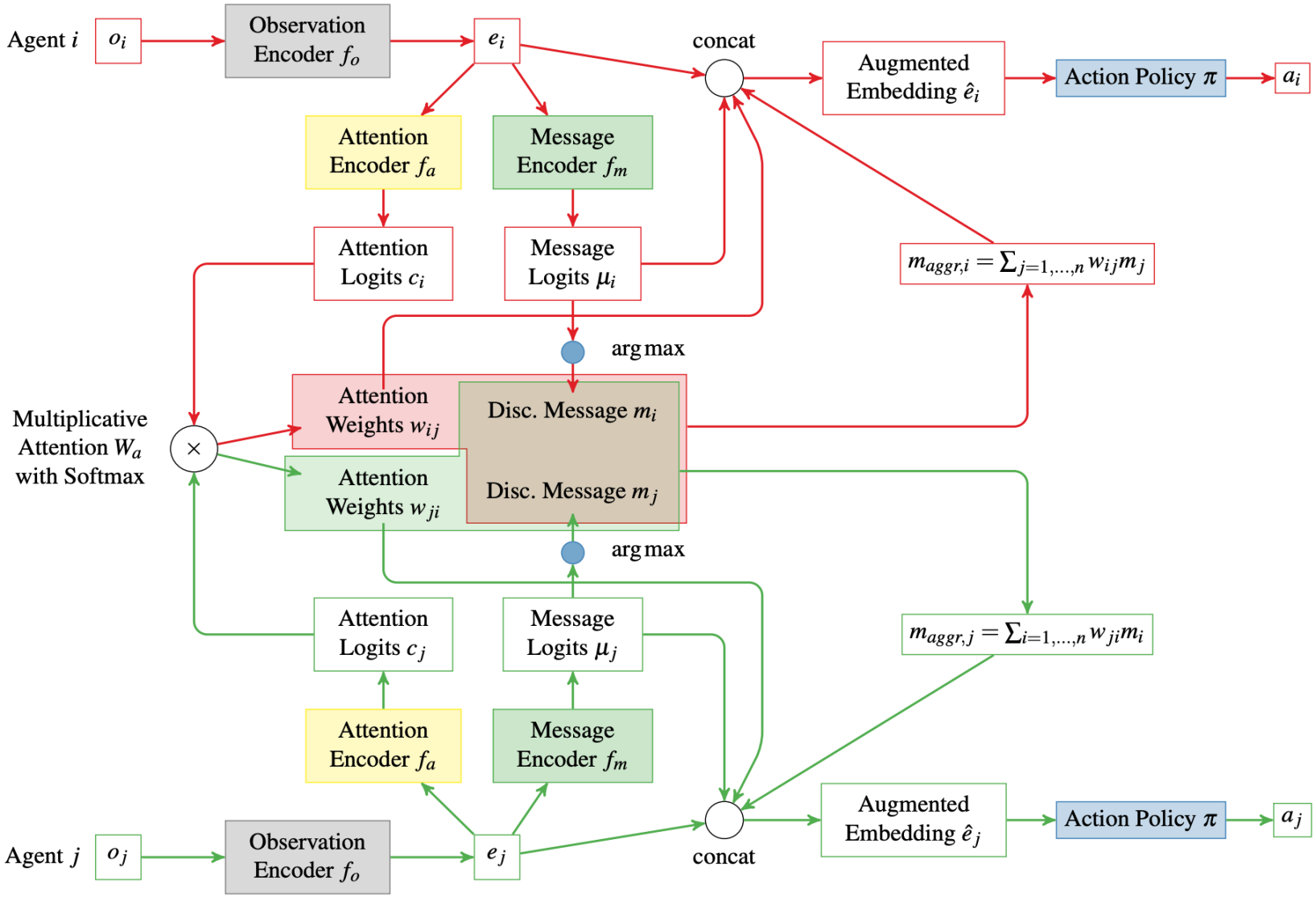
- **Problems**
  - DICG does not have explicit communication channels
  - Implicit emergent messages
  - Interpretability
- **Contributions**
  - A broadcasting-based communication model
  - Discrete token communication
  - Positive signaling and listening analysis
  - Human-agent interaction

S. Li, Y. Zhou, R. Allen, M. J. Kochenderfer, "Learning Emergent Discrete Message Communication for Cooperative Reinforcement Learning", in IEEE International Conference on Robotics and Automation (ICRA), 2022.

Background & Existing Work

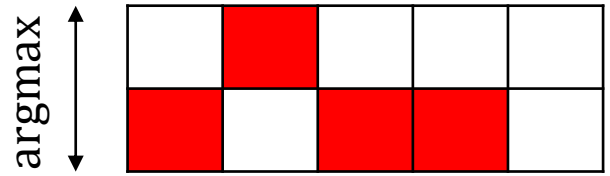
- Simple message aggregation
  - CommNet, IC3Net
- Existing multi-agent communication approaches use continuous messages to communicate
  - TarMAC, ATOC, and all previous work
- Grounding discrete messages
  - Studied in simple two-agent referential games
- Human-agent communication studies are rarely seen in MARL context

Explicit Broadcast Communication Model

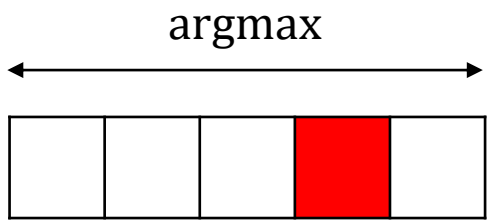


Discrete Messaging by argmax

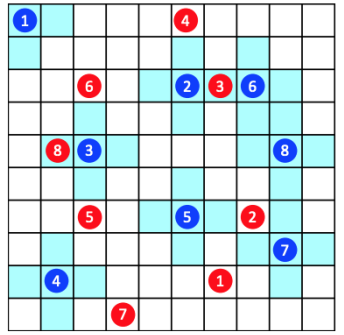
- A lot more ***stable*** than Gumbel softmax
- An extreme way of sampling
- Bit String
  - E.g. $m = 10110$
  - $m_k = \text{argmax}\, logits[k, :]$

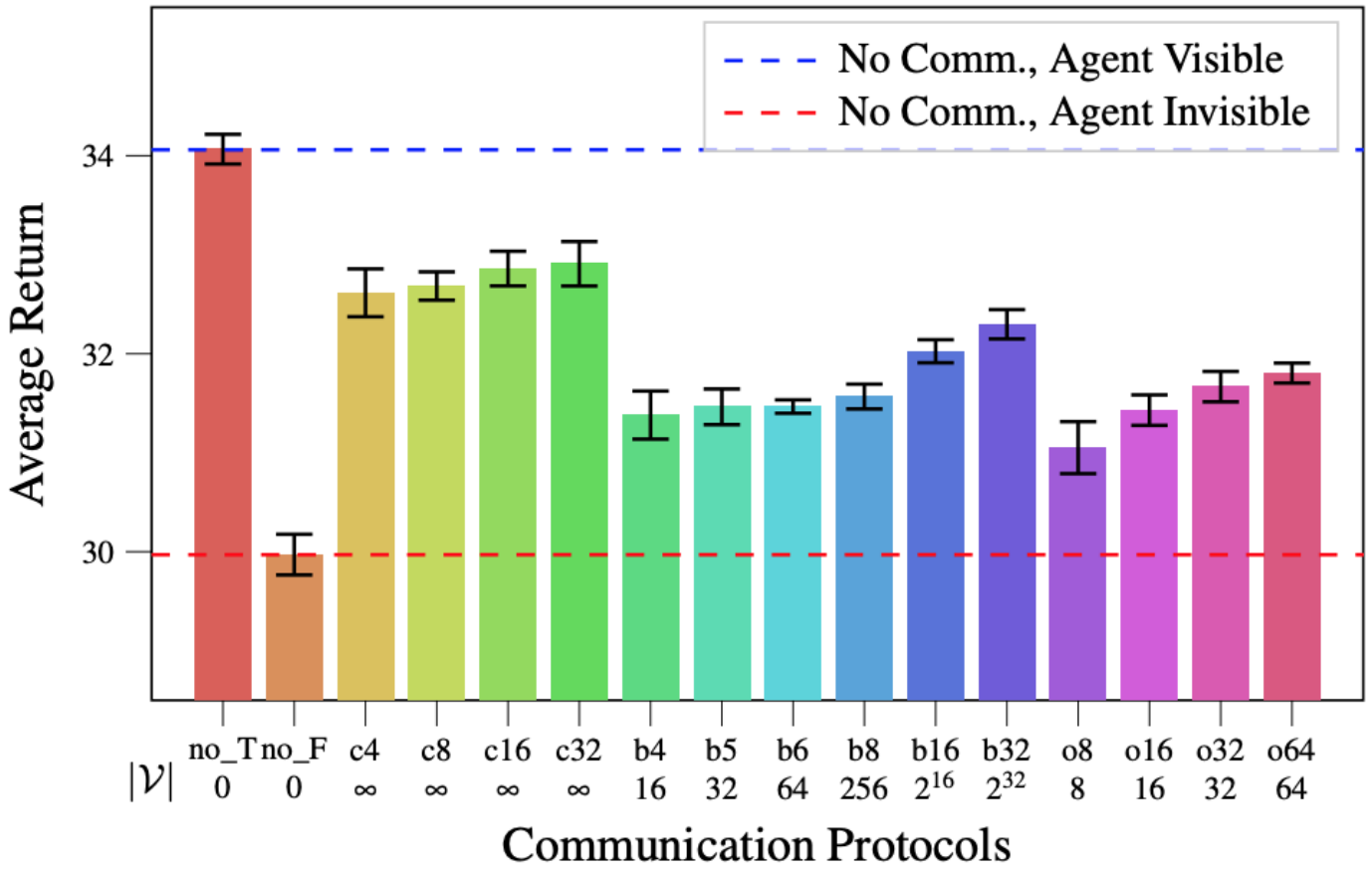


- Onehot
  - E.g. $m = 00010$
  - $m = \text{onehot}(\text{argmax}\, logits)$



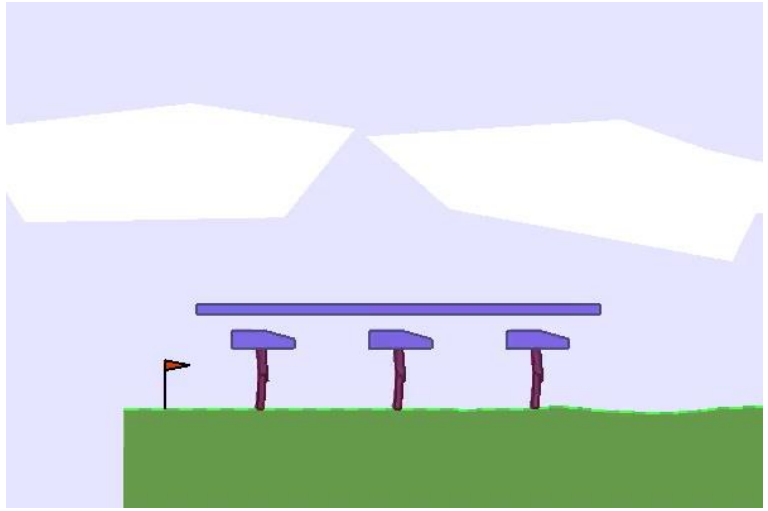
Results: Predator-Prey



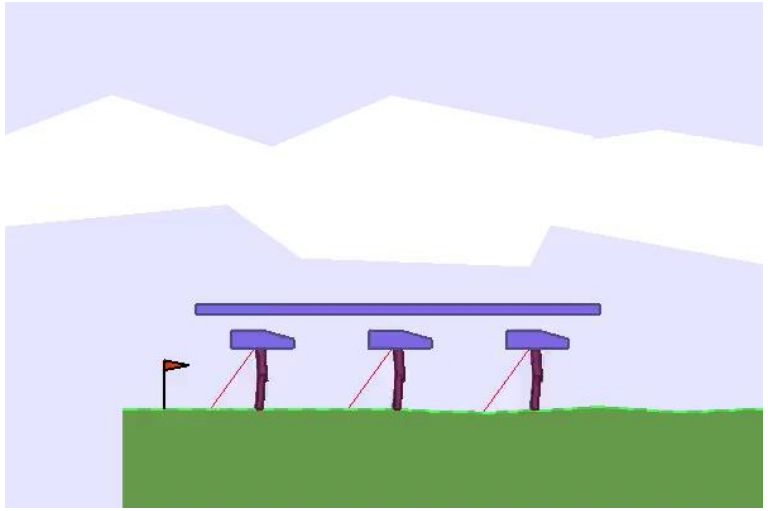


- Single agent capture penalty enabled
- Other agents are invisible to each agent
- Communication is necessary for a high return

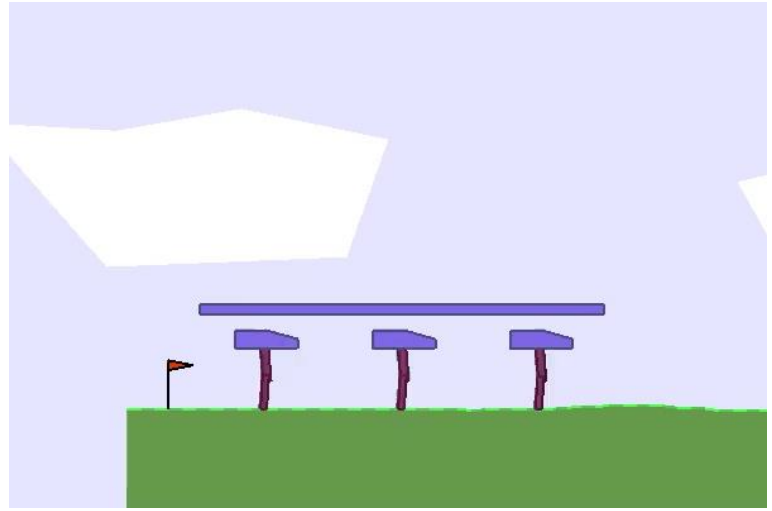
Demos: Multi-Walker-3



- Comm: off
- Neighbor invisible
- Lidar off

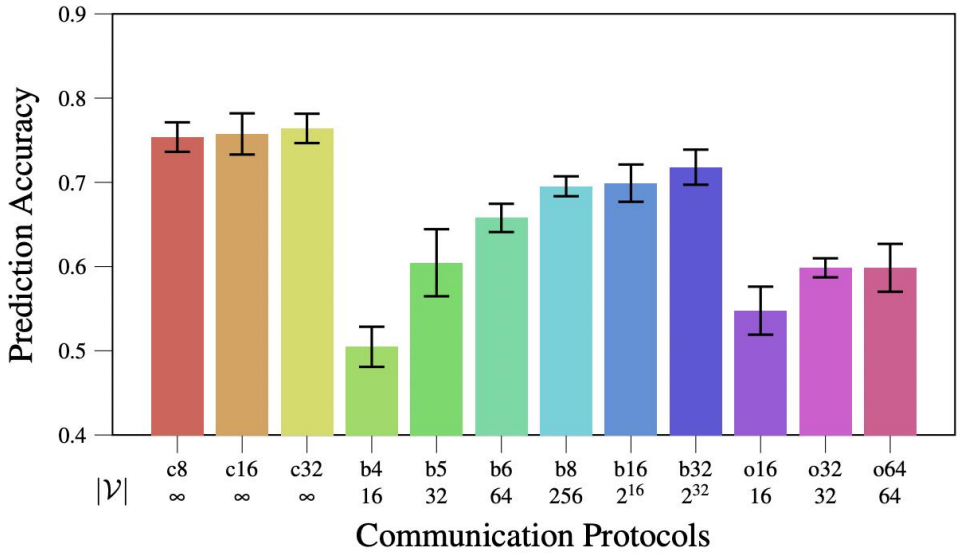


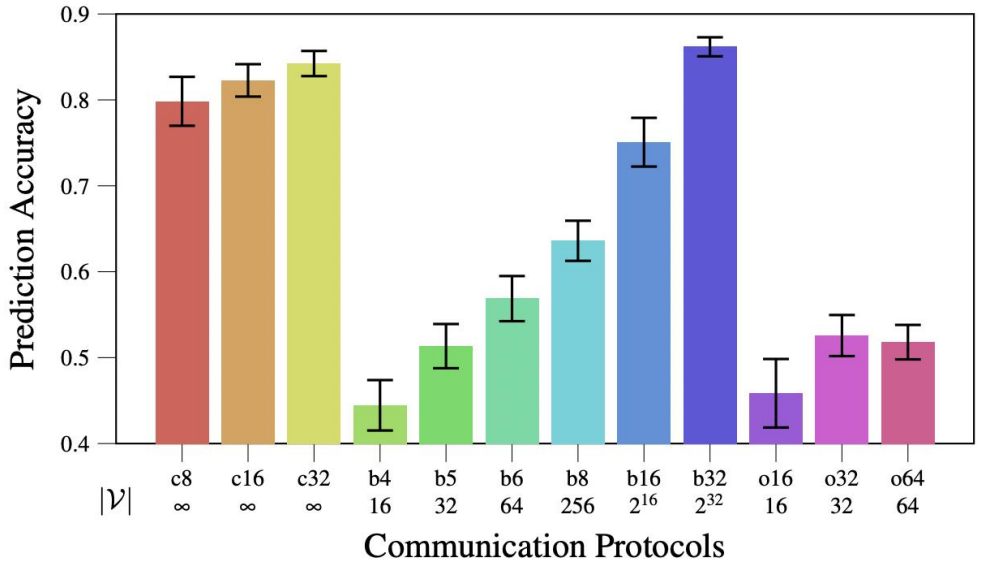
- Comm: off
- Neighbor visible
- Lidar on



- Comm: 8-bit-string ($|V| = 256$)
- Neighbor invisible
- Lidar off

Positive Listening and Signaling





Positive Listening: the degree to which received messages are influencing an agent's behaviors

$$\hat{a}_i = f(\overline{m}_{aggr,i}; \phi)$$

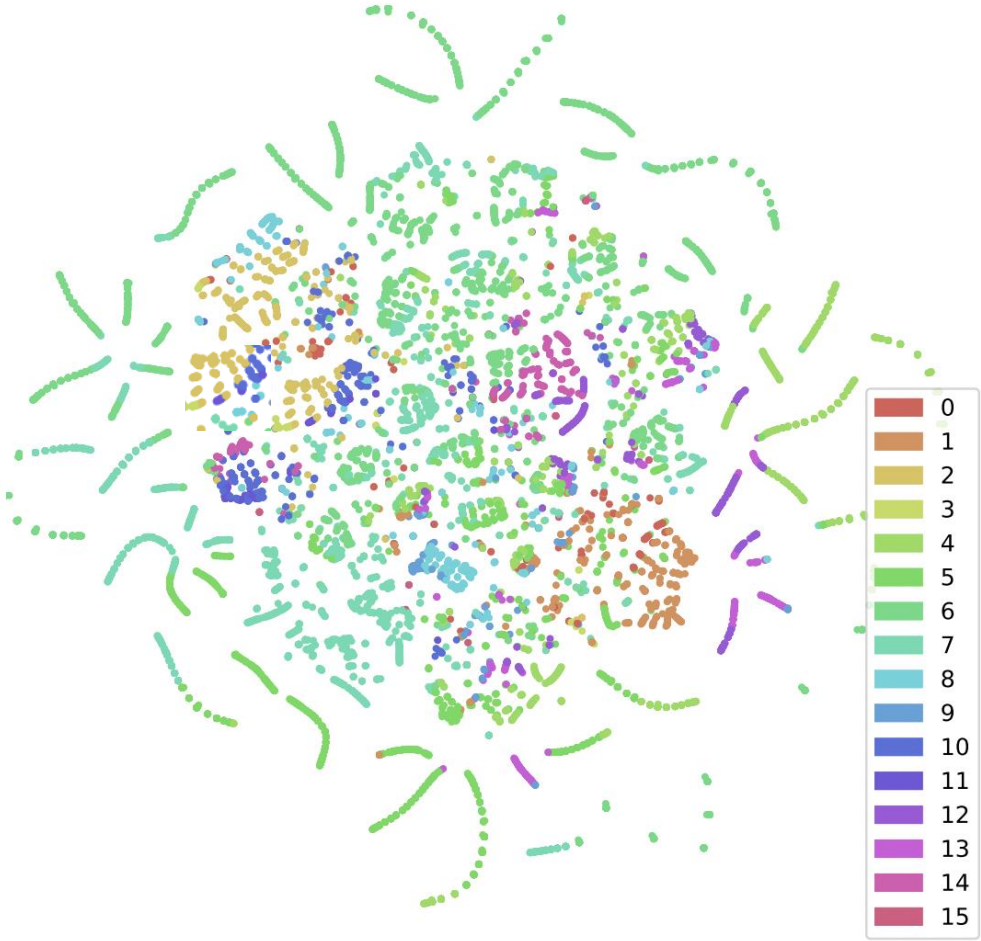
Positive Signaling: the degree to which an agent's sent messages are related to its behaviors

$$\hat{a}_i = f(m_i; \phi)$$

Exploring Human-Agent Interaction

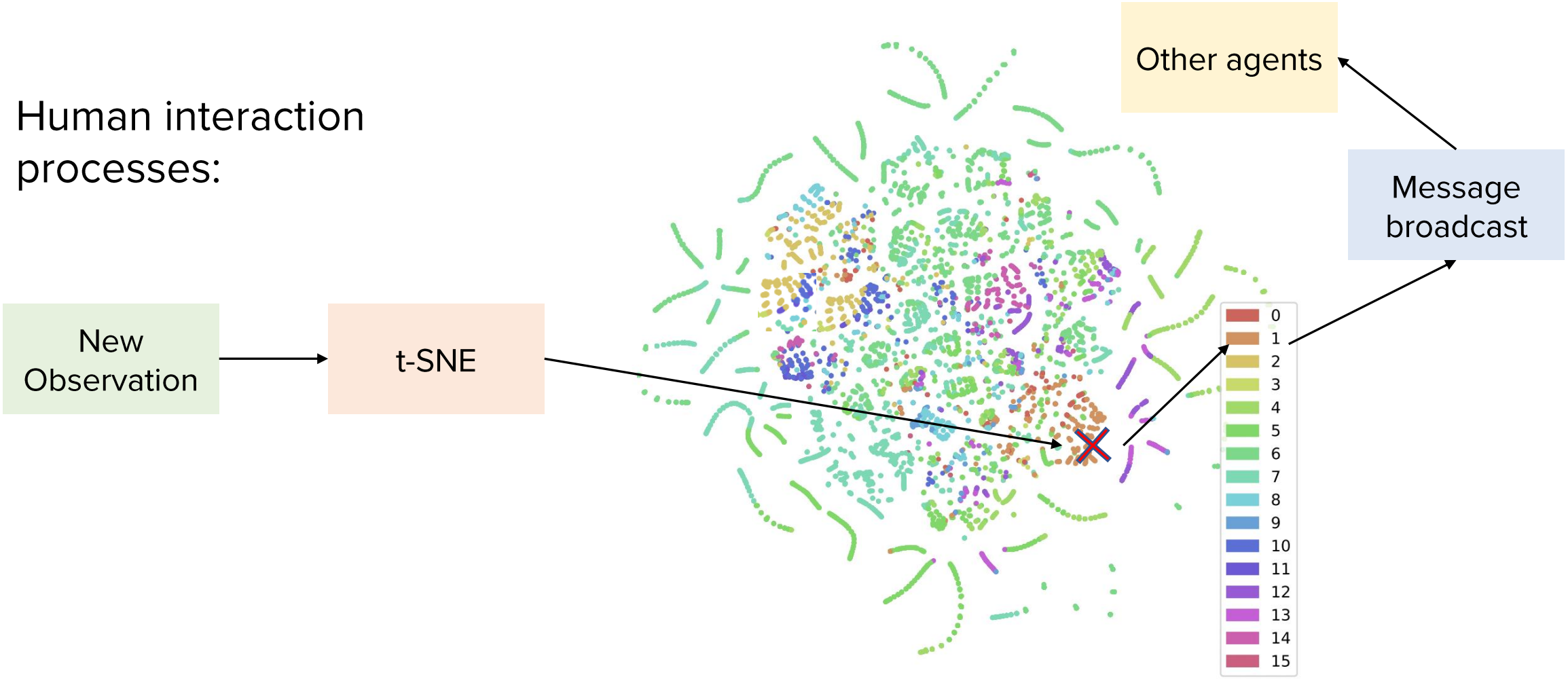
Let human select discrete messages and send them to agents

1. Pre-train the communication model
2. Run t-SNE clustering for raw observations, label them with model generated messages



Exploring Human-Agent Interaction

Human interaction processes:



Discussion

- Proposed an explicit broadcasting-based communication model
- Achieved end-to-end trainable discrete token communication between agents
- Analyzed positive signaling and listening
- Designed a human-agent interaction method

Outline

- Background and Introduction
- Contribution 1: Improving Utility Fusion with Deep Correction
- Contribution 2: Deep Implicit Coordination Graphs
- Contribution 3: Learning Emergent Discrete Communication
- Summary and Future Work

Summary

1. Improving Utility Fusion with Deep Correction
   - Using utility decomposition and fusion to solve MARL from single agent's perspective
   - A deep Q-network as correction term
2. Deep Implicit Coordination Graphs
   - Self-attention and graph neural network
   - Direct policy learning or critic learning
3. Learning Emergent Discrete Communication
   - An explicit broadcasting-based communication model
   - Discrete token communication

Future Research Directions

- **Interpretable Communication**
  - Grounding communication into concrete meanings
  - Natural language
  - Ad-hoc human-agent teaming / zero-shot coordination
- **Explainable Measurements of Interaction**
  - Distance and attention are good but not principled enough
  - Reason about the behavioral history or causal influence
- **Model-based MARL**
  - Maintain a belief about what others are going to do in the future
  - Coordinate actions according to a combination of experience, current observation and the belief of future

Acknowledgements

AIRBUS A3

MIT LINCOLN LABORATORY