Learning Multiagent Communication with Backpropagation (CommNet)

Contents

Abstract

- Many tasks in AI require the collaboration of multiple agents
- The communication protocol between agents is manually specified and not altered during training

We explore a simple nueral model, called ***CommNet***, that uses continuous communication for fully cooperative tasks

The model consists of multiple agents and the communication between them is learned alongside their policy.

Introduction

- Communication is a fundamental aspect of intelligence, enabling agents to behave as a group, rather than a collection of individuals
    - It is vital for performing complex tasks in real-world environments where each actor has limited capabilities and/or visibility of the world
    - Ex) Elevator control, sensor network, robot soccer
- In any partially observed environment, the communication between agents is vital to coordinate the behavior of each individual
- While the model controlling each agent is typically learned via reinforcement learning, the specification and format of the communication is usually pre-determined

We propose a model where cooperating agents learn to communicate ***amongst themselves*** before taking actions

Advantages: ***Simple and versatile***

1. A wide range of problems involving partial visibility of the environment
2. The model allows dynamic variation at run time in both the number and type of agents

Introduction

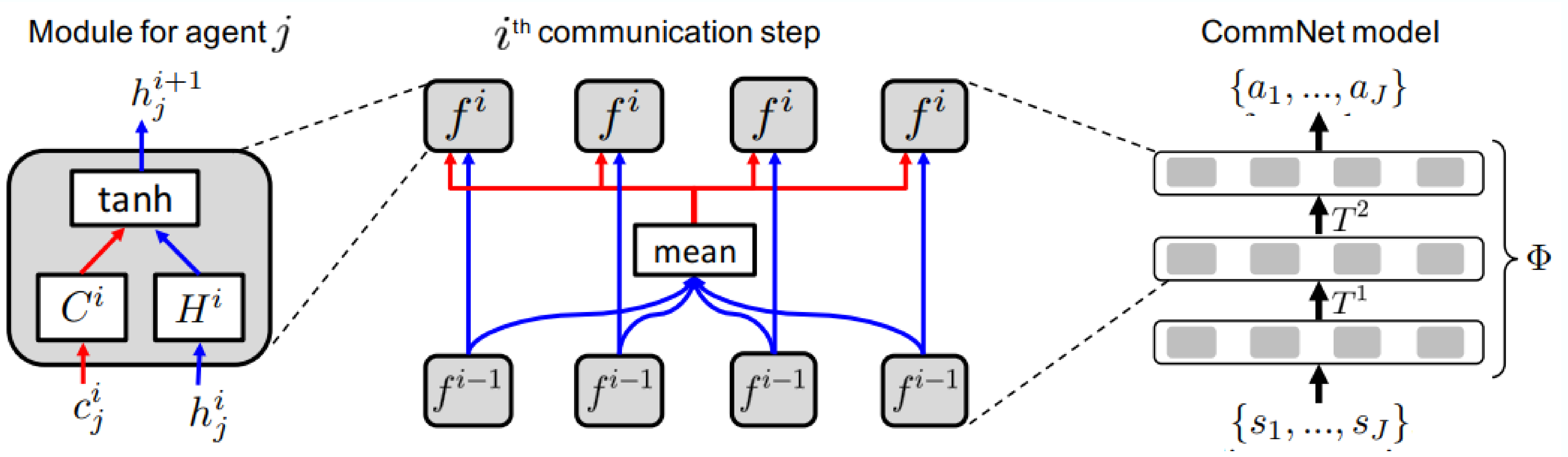
- Setting
  - J agents
  - All cooperating to maximize reward R
  - Each agent receives R independent of their contribution
- In this setting, there is no difference ①between each agent having its own controller, or ②viewing them as pieces of a larger model controlling all agents

<u>Our controller is a large feed-forward neural network</u> that maps inputs for all agents to their actions, each agent occupying a subset of units

Communication Model (CommNet)

Model = Controller = $\boldsymbol{\Phi}$

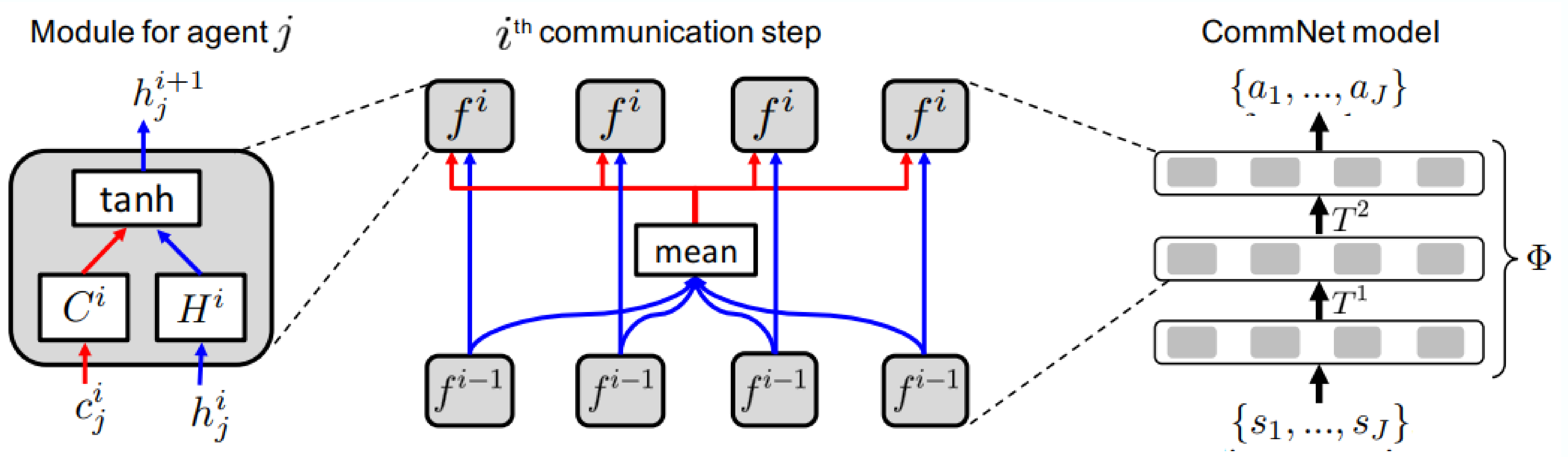
- Controller: individual controllers & communication between agents
  - $\Phi(\mathbf{s}) = \boldsymbol{a}\ (\boldsymbol{s} = \{s_1, \ldots, s_J\}, \mathbf{a} = \{a_1, \ldots, a_J\}$, J agents)
  - $\Phi(\{s_1, \ldots, s_J\}) = \{a_1, \ldots, a_J\}$
- f^i module: two input vectors for each agent j
  - $h_j^{i+1} = f^i(h_j^i, c_j^i)$
  - $c_j^{i+1} = \frac{1}{J-1}\sum_{j' \neq j} h_{j'}^{i+1}$
  - $i \in \{0, \ldots, K\}$, K is the number of communication step
- A single layer
  - $\boldsymbol{h}^{i+1} = \sigma(T^i \boldsymbol{h}^i)$
  - $T^i = \begin{pmatrix} H^i & \bar{C}^i & \bar{C}^i & \cdots & \bar{C}^i \\ \bar{C}^i & H^i & \bar{C}^i & \cdots & \bar{C}^i \\ \bar{C}^i & \bar{C}^i & H^i & \cdots & \bar{C}^i \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \bar{C}^i & \bar{C}^i & \bar{C}^i & \cdots & H^i \end{pmatrix}$
  - $\bar{C}^i = \frac{C^i}{J-1}$
- T is *dynamically sized*



Communication Model (CommNet)

Model = Controller = *Φ*

- Local connectivity
  - Agents to communicate to others within a certain range
  - N(j) : the set of agents present within communication range of agent j
  - $c_j^{i+1} = \frac{1}{J-1}\sum_{j'\neq j} h_{j'}^{i+1} \Rightarrow c_j^{i+1} = \frac{1}{|N(j)|}\sum_{j'\in N(j)} h_{j'}^{i+1}$
- Skip connections
  - $h_j^{i+1} = f^i(h_j^i, c_j^i) \Rightarrow h_j^{i+1} = f^i(h_j^i, c_j^i, h_j^0)$
- Temporal recurrence
  - Using the same module f^t for all t



Related Work

- Go, Atari games
    - Multi-agent domains
    - Full visibility of the environment and lack communication
- Many approaches avoid the need for communication by making strong assumptions about visibility
- Others use communication but with a pre-determined protocol
- Foerster *et al.* : the closest approach

Our model combines ***a deep network*** with ***reinforcement learning***.

The communication is learned rather than being pre-determined

Multiple ***continuous*** communication cycles are used.

Dynamic variation in the number of agents

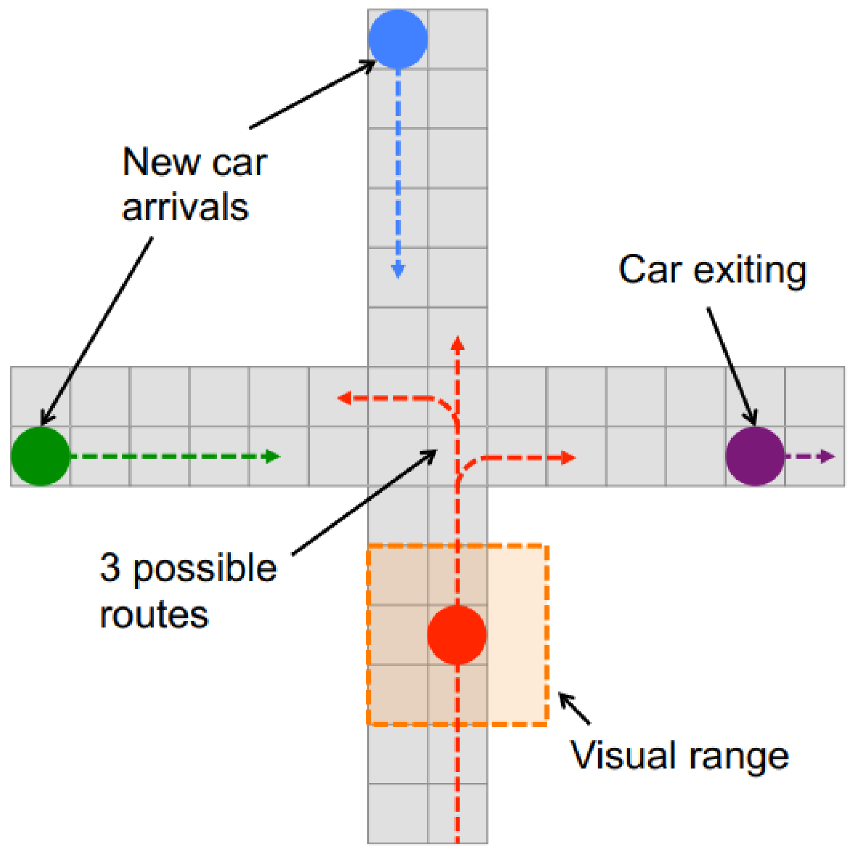
Experiments

- Three baselines models to compare against our model
  - Independent controller
    - Controlled independently & no communication
    - Modularity and flexibility
  - Fully-connected
    - Make Φ a fully-connected multi-layer neural network
    - Not modular, inflexible
  - Discrete communication
    - Agent communicate via discrete symbols
- A lever pulling task (simple game)
  - N = 500
  - M = 5
  - Communication step K = 2
  - Skip connections
  - Training: 50,000 batches of size 64
  - Evaluation: averaged over 500 trials

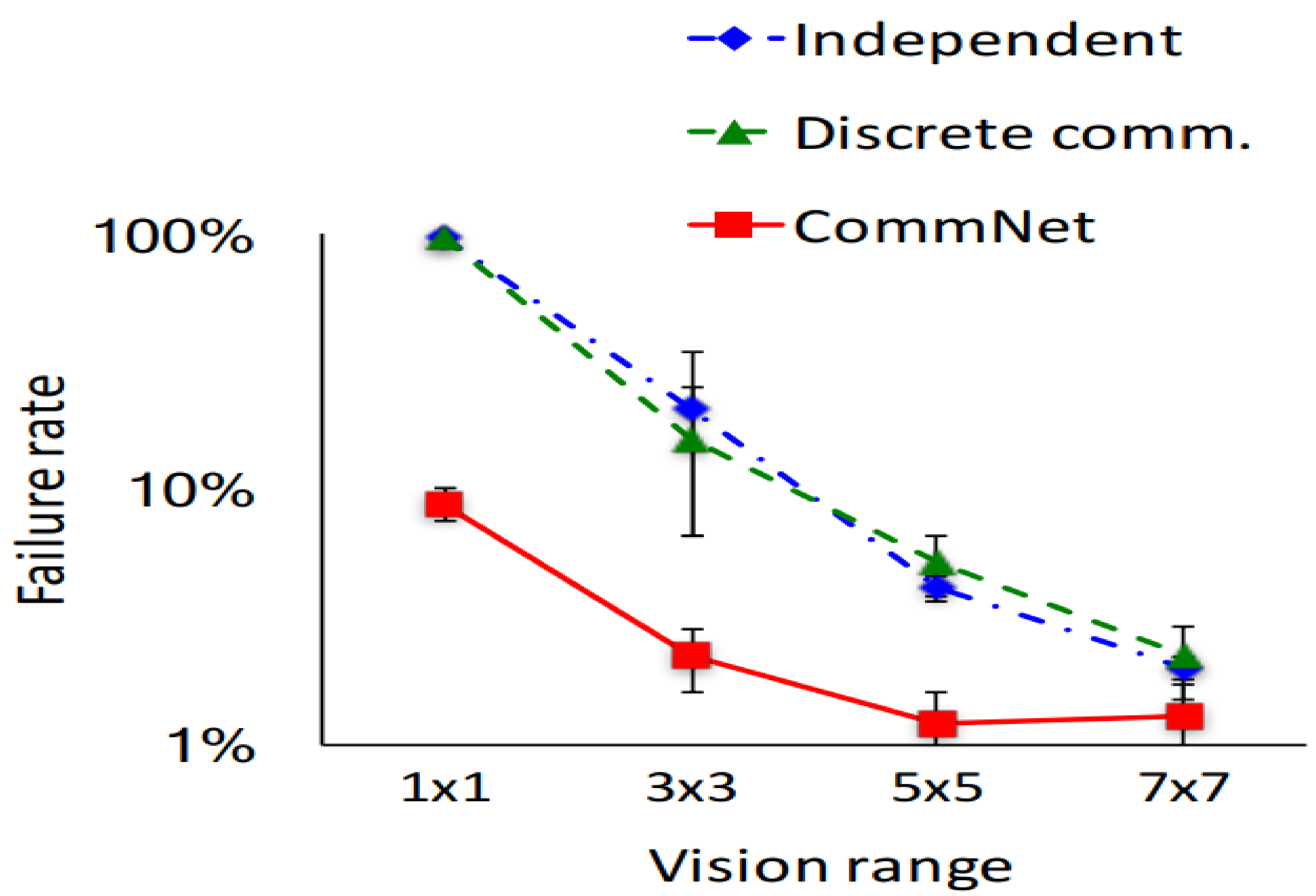
	Training method	
Model Φ	Supervised	Reinforcement
Independent	0.59	0.59
CommNet	**0.99**	**0.94**

Experiments

- Two multi-agent tasks using the MazeBase environment
- Traffic Junction
  - Control cars passing through a traffic junction to maximize the flow while minimizing collisions
  - State for agent j: {n, j, r} & visual range 3*3
  - Action: gas, brake
  - Reward: collision & traffic
    - $r(t) = C^t r_{coll} + \sum_{i=1}^{N^t} \tau_i r_{time}$ (r_{coll}=-10, $\tau_i r_{time}$=-0.01τ)
  - Results: the probability of failure
  - How partial visibility within the environment effects the advantage given by communication

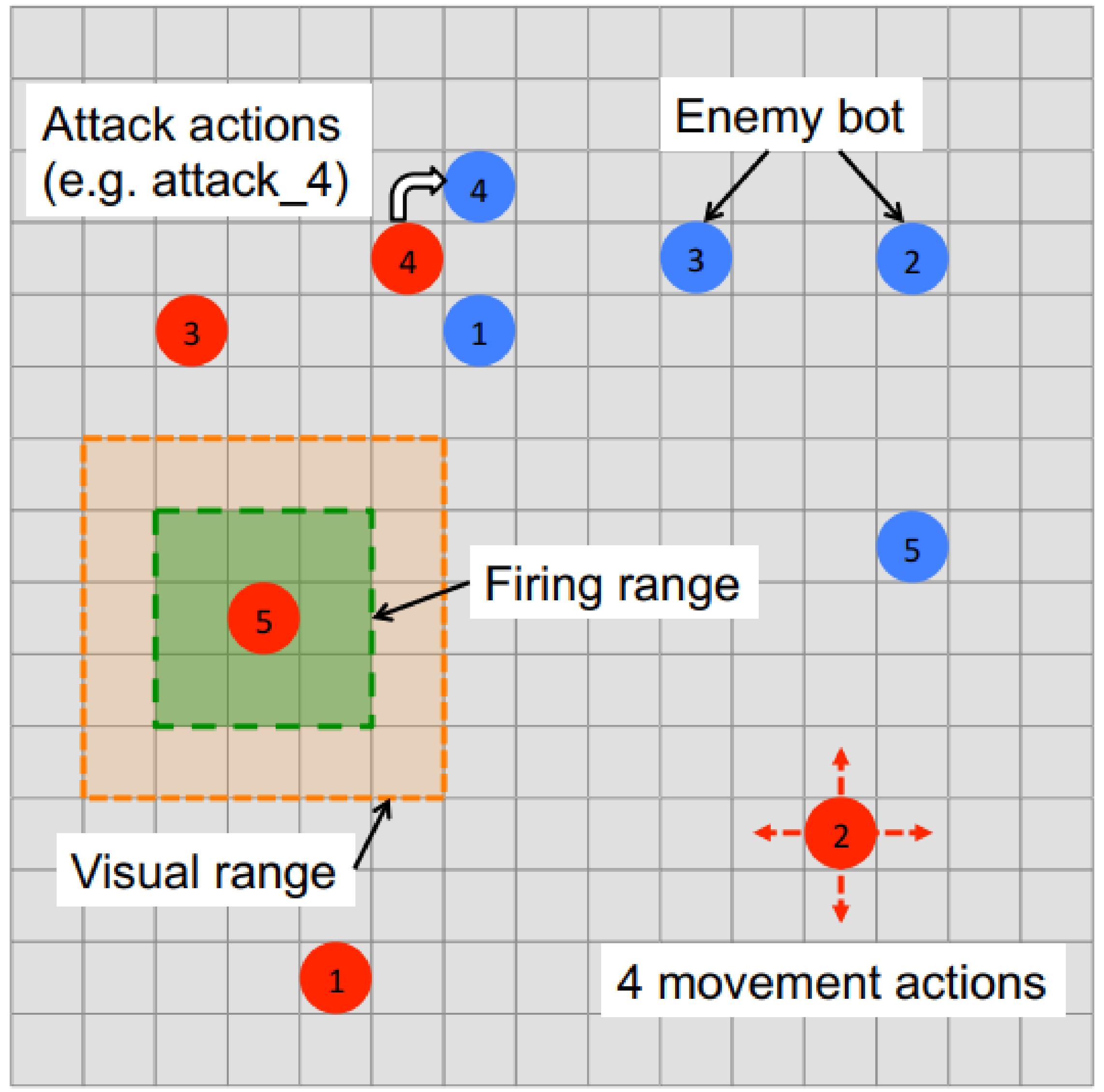


Model Φ	Module $f()$ type MLP	RNN	LSTM
Independent	20.6± 14.1	19.5± 4.5	9.4± 5.6
Fully-connected	12.5± 4.4	34.8± 19.7	4.8± 2.4
Discrete comm.	15.8± 9.3	15.2± 2.1	8.4± 3.4
CommNet	**2.2**± **0.6**	**7.6**± **1.4**	**1.6**± **1.0**



Experiments

- Combat Task
  - Control multiple agents in combat against enemy bots
  - State for agent j: {i, t, l, h, c} & visual range 3*3
  - Action: move, attack, do nothing
  - Reward
    - -1: team loses or draws at the end of the game
    - -0.1 times the total health points of the enemy team
  - Results: the win rate



Model Φ	Module $f()$ type		
	MLP	RNN	LSTM
Independent	34.2± 1.3	37.3± 4.6	44.3± 0.4
Fully-connected	17.7± 7.1	2.9± 1.8	19.6± 4.2
Discrete comm.	29.1± 6.7	33.4± 9.4	46.4± 0.7
CommNet	**44.5**± **13.4**	**44.4**± **11.9**	**49.5**± **12.6**

Model Φ	Other game variations (MLP)		
	$m = 3$	$m = 10$	5×5 vision
Independent	29.2± 5.9	30.5± 8.7	60.5± 2.1
CommNet	**51.0**± **14.1**	**45.4**± **12.4**	**73.0**± **0.7**

Discussion and Future Work

- Evaluation clearly show the model outperforms models without communication, fully-connected models, and models using discrete communication
- Future Work
    - We did not fully exploit ability to handle heterogenous agent types
    - The model will scale gracefully to large number of agents

<u>CommNet, a simple controller for MARL, that is able to learn continuous communication between a dynamically changing set of agents.</u>

Thanks!

ANY QUESTIONS?

You can find me at

tocrit@gmail.com