HAMMER: Multi-Level Coordination of Reinforcement Learning Agents via Learned Messaging

Nikunj Gupta, G Srinivasararaghavan, Swarup Kumar Mohalik, Nishant Kumar, Matthew E. Taylor

Nikunj Gupta

IIIT Bangalore
Ericsson Research
IRL Lab, University of Alberta

April 25, 2021

Agenda

- Introduction
- Setting and Goal
- Proposed Framework — HAMMER
- Key Results
- Conclusion
- Future Work

Agenda

- Introduction
- Setting and Goal
- Proposed Framework — HAMMER
- Key Results
- Conclusion
- Future Work

Agenda

- Introduction
- Setting and Goal
- Proposed Framework — HAMMER
- Key Results
- Conclusion
- Future Work

Agenda

- Introduction
- Setting and Goal
- Proposed Framework — HAMMER
- Key Results
- Conclusion
- Future Work

Agenda

- Introduction
- Setting and Goal
- Proposed Framework — HAMMER
- Key Results
- Conclusion
- Future Work

Agenda

- Introduction
- Setting and Goal
- Proposed Framework — HAMMER
- Key Results
- Conclusion
- Future Work

Introduction

Cooperative Multi-agent Reinforcement Learning

Communication in MARL

Hierarchical approach to MARL

Introduction

Cooperative Multi-agent Reinforcement Learning

- Simultaneous learning and interaction of multiple agents in the same environment to ***achieve shared goals***.
- Some naturally-fitting applications include: distributed logistics, packet delivery and disaster rescue.

Communication in MARL

Hierarchical approach to MARL

Introduction

Cooperative Multi-agent Reinforcement Learning

Communication in MARL

In ***coordination-intensive*** tasks, communication has been shown to be an important aspect. It could be in the form of:

- sharing experiences among the agents
- sharing low-level information like gradient updates via communication channels, or
- sometimes even directly advising appropriate actions using a pre-trained agent (teacher)

Hierarchical approach to MARL

Introduction

Cooperative Multi-agent Reinforcement Learning

Communication in MARL

Hierarchical approach to MARL

- We propose ***multi-level coordination*** among intelligent agents via messages learned by a separate agent to ease the localized learning of task-related policies.
- The main insight of our algorithm is to ***learn to communicate relevant pieces of information*** from a global perspective to help agents with limited capabilities improve their performance.

Setting and Goal

- Consider a warehouse setting with lots of small, simple, robots fetching and stocking items on shelves.
- These local agents are not capable of communicating among themselves.
- Also, one very powerful centralized agent to determine the joint action of all the local agents would *not* scale well with an exponential growth in the observation and actions spaces with the number of agents.

Goal of HAMMER

We propose learning and communicating ***high-level messages*** based on complete knowledge of all the local agents in the environment and *"facilitating"* overall team learning.

Setting and Goal

- Consider a warehouse setting with lots of small, simple, robots fetching and stocking items on shelves.
- These local agents are not capable of communicating among themselves.
- Also, one very powerful centralized agent to determine the joint action of all the local agents would *not* scale well with an exponential growth in the observation and actions spaces with the number of agents.

Goal of HAMMER

We propose learning and communicating ***high-level messages*** based on complete knowledge of all the local agents in the environment and *"facilitating"* overall team learning.

Setting and Goal

- Consider a warehouse setting with lots of small, simple, robots fetching and stocking items on shelves.
- These local agents are not capable of communicating among themselves.
- Also, one very powerful centralized agent to determine the joint action of all the local agents would *not* scale well with an exponential growth in the observation and actions spaces with the number of agents.

Goal of HAMMER

We propose learning and communicating ***high-level messages*** based on complete knowledge of all the local agents in the environment and *"facilitating"* overall team learning.

Setting and Goal

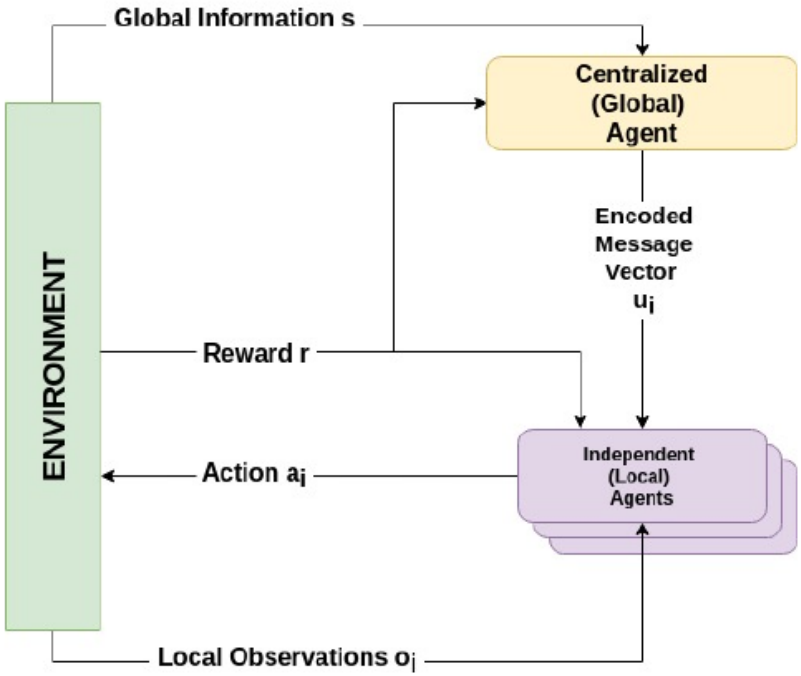
- Consider a warehouse setting with lots of small, simple, robots fetching and stocking items on shelves.
- These local agents are not capable of communicating among themselves.
- Also, one very powerful centralized agent to determine the joint action of all the local agents would *not* scale well with an exponential growth in the observation and actions spaces with the number of agents.

Goal of HAMMER

We propose learning and communicating ***high-level messages*** based on complete knowledge of all the local agents in the environment and *"facilitating"* overall team learning.

Proposed Framework — HAMMER

Heterogeneous **A**gents **M**astering **M**essaging to **E**nhance **RL**



Our cooperative MARL setting: a supplemental global agent sends messages to help multiple independent local agents act in an environment.

HAMMER — Strategies to Communicate

Multiple techniques were used for training the central agent to learn how to communicate.

- **HAMMERv1:** Employed PPO (HAMMER can use any other RL algorithm too) to learn directly using the local rewards (weak signal / difficult training)
- **HAMMERv2:** Pushed gradients from the local agents' network to HAMMER (better gradient signal)
- **HAMMERv3:** Pre-processed central agent's messages using a regularisation unit before transmitting (message discretization)

HAMMER — Strategies to Communicate

Multiple techniques were used for training the central agent to learn how to communicate.

- **HAMMERv1:** Employed PPO (HAMMER can use any other RL algorithm too) to learn directly using the local rewards (weak signal / difficult training)
- **HAMMERv2:** Pushed gradients from the local agents' network to HAMMER (better gradient signal)
- **HAMMERv3:** Pre-processed central agent's messages using a regularisation unit before transmitting (message discretization)

HAMMER — Strategies to Communicate

Multiple techniques were used for training the central agent to learn how to communicate.

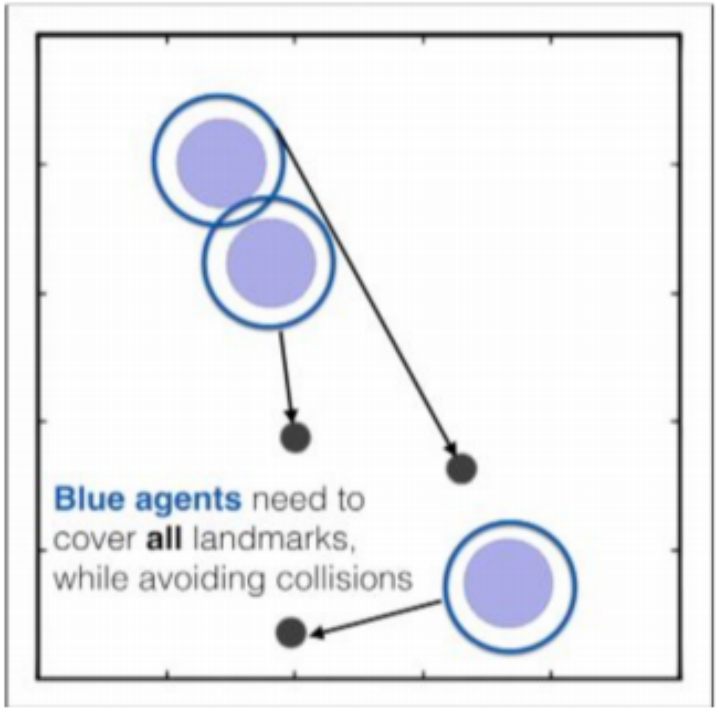
- **HAMMERv1:** Employed PPO (HAMMER can use any other RL algorithm too) to learn directly using the local rewards (weak signal / difficult training)
- **HAMMERv2:** Pushed gradients from the local agents' network to HAMMER (better gradient signal)
- **HAMMERv3:** Pre-processed central agent's messages using a regularisation unit before transmitting (message discretization)

HAMMER — Strategies to Communicate

Multiple techniques were used for training the central agent to learn how to communicate.

- **HAMMERv1:** Employed PPO (HAMMER can use any other RL algorithm too) to learn directly using the local rewards (weak signal / difficult training)
- **HAMMERv2:** Pushed gradients from the local agents' network to HAMMER (better gradient signal)
- **HAMMERv3:** Pre-processed central agent's messages using a regularisation unit before transmitting (message discretization)

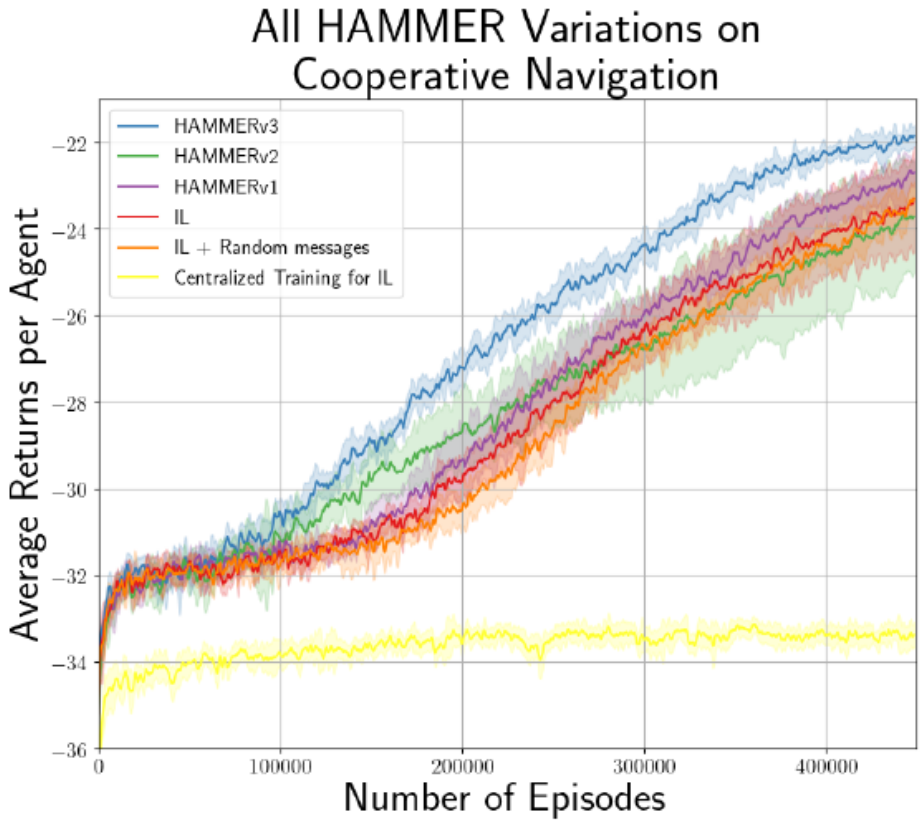
Key Environment for our Study



We used the **cooperative navigation** environment, composing blue agents and black (stationary) landmarks the agents must cover, while avoiding collisions.

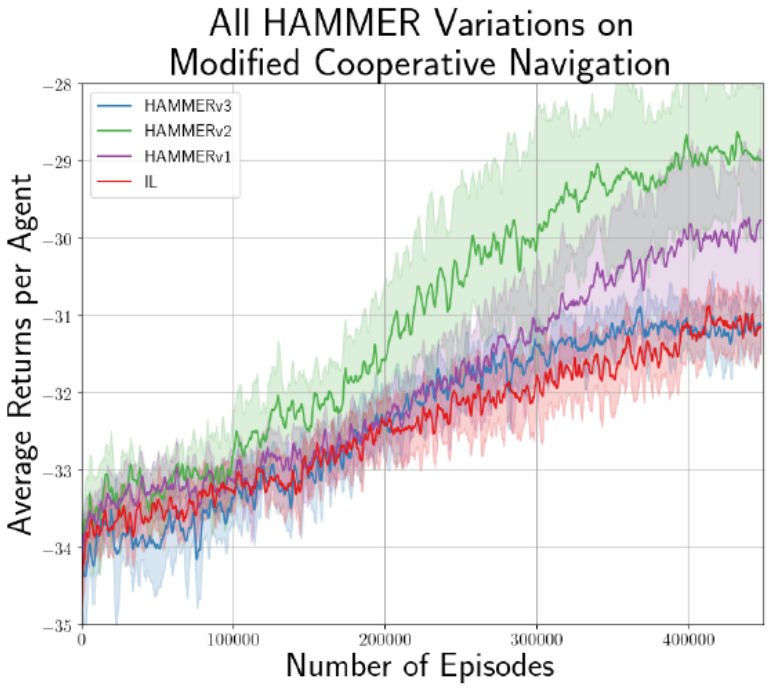
Key Results

HAMMER agents outperform independent PPO learners in cooperative navigation.



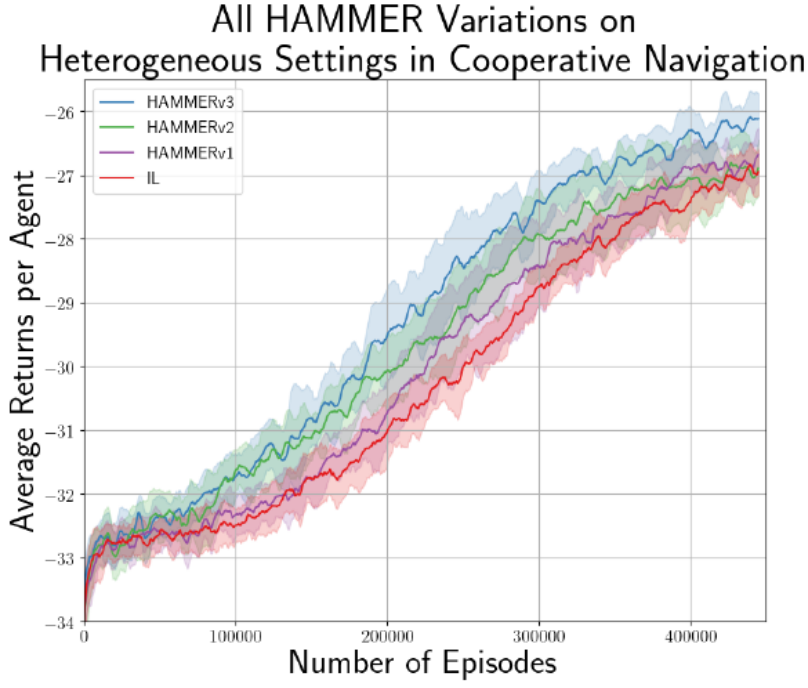
Providing the local agents with random messages causes degraded performance (as expected). Plus, Hammer also significantly outperforms centrally learned policy for independent agents.

Additional Studies



Modified Cooperative Navigation Environment

Disallowed local agents to observe each other, hence necessitating the need for communication via the central agent. HAMMER's performance in this setting shows that effective communication is indeed being learned to help the local agents coordinate.



Heterogeneous Cooperative Navigation Environment

Here, the local agents in the environment were heterogeneous in nature – one of the local agents was unable to observe the other agents, whereas the other two agents could. Results show that HAMMER was able to generalize to these settings too.

Conclusion

In this work we present our MARL algorithm ***HAMMER***, describe where it would be most applicable, and implement it in domains like ***cooperative navigation***.

Conclusion

In this work we present our MARL algorithm ***HAMMER***, describe where it would be most applicable, and implement it in domains like ***cooperative navigation***.

Empirical results show that

- learned communication does indeed improve system performance,
- results generalize to heterogeneous local agents, and
- results generalize to different reward structures (shown via results on another domain — Multi-agent Walker)

Conclusion

In this work we present our MARL algorithm ***HAMMER***, describe where it would be most applicable, and implement it in domains like ***cooperative navigation***.

Empirical results show that

- learned communication does indeed improve system performance,
- results generalize to heterogeneous local agents, and
- results generalize to different reward structures (shown via results on another domain — Multi-agent Walker)

Conclusion

In this work we present our MARL algorithm ***HAMMER***, describe where it would be most applicable, and implement it in domains like ***cooperative navigation***.

Empirical results show that

- learned communication does indeed improve system performance,
- results generalize to heterogeneous local agents, and
- results generalize to different reward structures (shown via results on another domain — Multi-agent Walker)

Conclusion

In this work we present our MARL algorithm ***HAMMER***, describe where it would be most applicable, and implement it in domains like ***cooperative navigation***.

Empirical results show that

- learned communication does indeed improve system performance,
- results generalize to heterogeneous local agents, and
- results generalize to different reward structures (shown via results on another domain — Multi-agent Walker)

Future Work

Further complex Multi-agent Environments

Promoting further hierarchies in HAMMER

Understanding HAMMER's learned messages

Allowing Limited Communication in HAMMER

Future Work

Further complex Multi-agent Environments

Experiments in multi-agent settings with tighter coupling and further complex interactions among agents such as in ***autonomous driving*** in SMARTS or heterogeneous multi-agent battles in ***StarCraft*** could help better appreciate the significance of the method.

Promoting further hierarchies in HAMMER

Understanding HAMMER's learned messages

Allowing Limited Communication in HAMMER

Future Work

Further complex Multi-agent Environments

Promoting further hierarchies in HAMMER

More ***complex hierarchies*** could be used, such as by making several central agents available in the system.

Understanding HAMMER's learned messages

Allowing Limited Communication in HAMMER

Future Work

Further complex Multi-agent Environments

Promoting further hierarchies in HAMMER

Understanding HAMMER's learned messages

Additional work remains to better understand if and ***how HAMMER tailors messages*** for the local agents using its global observation.

Allowing Limited Communication in HAMMER

Future Work

Further complex Multi-agent Environments

Promoting further hierarchies in HAMMER

Understanding HAMMER's learned messages

Allowing Limited Communication in HAMMER

In our setting, communication is free – future work could consider the case where it was costly and attempt to ***trade off the number of messages sent*** with the learning speed of HAMMER.

Acknowledgements

Professors at IIIT-Bangalore

Researchers at Ericsson R&D Bangalore

Team and Collaborators at IRL Lab, University of Alberta

Acknowledgements

Professors at IIIT-Bangalore

Most of this work was done at IIIT Bangalore, as part of Nikunj Gupta's Masters Thesis titled – *"Fully Cooperative Multi-Agent Reinforcement Learning"*. We would like to thank *Prof. Dinesh Babu J*, *Prof. V Ramasubramanian* and *Prof. Shrisha Rao*, for useful inputs during the early stages of this work.

Researchers at Ericsson R&D Bangalore

Team and Collaborators at IRL Lab, University of Alberta

Acknowledgements

Professors at IIIT-Bangalore

Researchers at Ericsson R&D Bangalore

This work commenced at Ericsson Research Lab, Bangalore, and we are grateful to the team for actively discussing real-world use cases to enrich HAMMER as a framework.

Team and Collaborators at IRL Lab, University of Alberta

Acknowledgements

Professors at IIIT-Bangalore

Researchers at Ericsson R&D Bangalore

Team and Collaborators at IRL Lab, University of Alberta

Part of this work has taken place in the ***Intelligent Robot Learning (IRL) Lab*** at the ***University of Alberta***, which is supported in part by research grants from the *Alberta Machine Intelligence Institute (Amii)*, *CIFAR*, and *NSERC*. We would like to thank the entire team, *Laura Petrich*, *Shahil Mawjee* and anonymous reviewers for comments and suggestions on earlier versions of this paper.

Thank You!

Nikunj Gupta
Nikunj.Gupta@iiitb.org | +91-9409607135
Active Collaborator | IRLL, University of Alberta
Data Scientist | Aganitha Cognitive Solutions
IMTech CSE | IIIT Bangalore
LinkedIn | GitHub | Google Scholar | CV | Website