Lazy Agents: A New Perspective on Solving Sparse Reward Problem in Multi-agent Reinforcement Learning

Boyin Liu, Zhiqiang Pu[1,2], Yi Pan[2], Jianqiang Yi[1,2], Yanyan Liang[3], Du Zhang[3]

[1] School of Artificial Intelligence, University of Chinese Academy of Sciences

[2] Institute of Automation, Chinese Academy of Sciences

[3] Macau University of Science and Technology

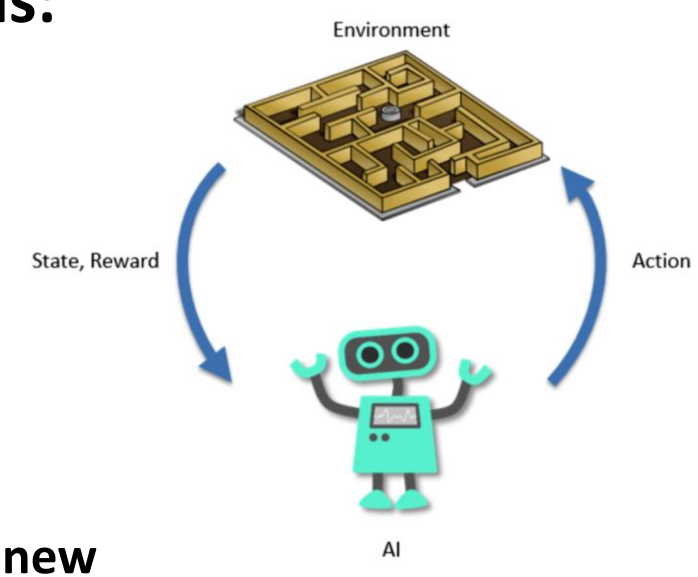
E-mail: boyinliu2019@ia.ac.cn

CONTENT

01 INTRODUCTION

Multi agent reinforcement learning with sparse rewards:

- *Joint action space grows exponentially with the number of agents.*
- *Typically,agents receive a global reward.*
- *Reward comes only under certain circumstances,e.g.,success/failure.*



- **We approach multi-agent sparse RL scenarios from a brand new perspective, i.e., lazy agents, and propose a method to address this problem.**

01 CONTRIBUTIONS

Our method has four contributions :

- This article presents a novel approach to analyzing the low training efficiency in sparse reward scenarios from the perspective of lazy agents. We define lazy agents mathematically, which distinguishes it from prior research in MARL.

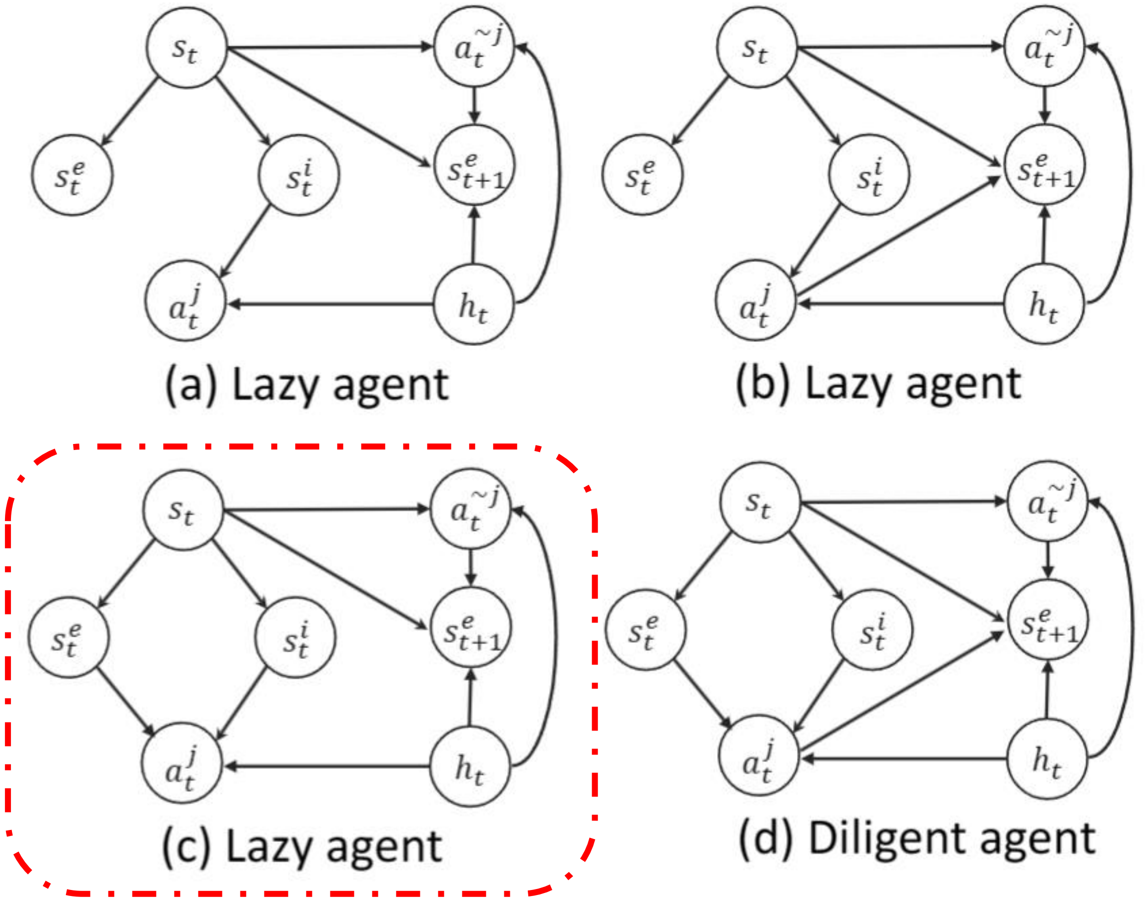
- We propose a new framework called LAIES to tackle the issue of lazy agents in sparse reward scenarios. LAIES has the potential to greatly enhance the training performance of value-based and policy gradient-based MARL algorithms.

- We propose a novel application of causal inference to address MARL problems. Our approach involves analyzing the causes of lazy agents using causal graphs and using counterfactual reasoning to resolve this issue.

- Our method outperforms the current state-of-the-art (SOTA) method on both sparse reward versions of SMAC and GRF [1].

03 METHOD

Definition of lazy agents in multi-agent systems



(a) Lazy agent

(b) Lazy agent

(c) Lazy agent

(d) Diligent agent

In MARL, lazy agents typically refer to the third type of lazy agent (shown in c)

Definition 1: Global states $s_t = (s_t^i, s_t^e)$ can be divided into two parts internal state s_t^i and external state s_t^e .

In Figure d, the causal effect of a_t^j on s_{t+1}^e can be calculated as:

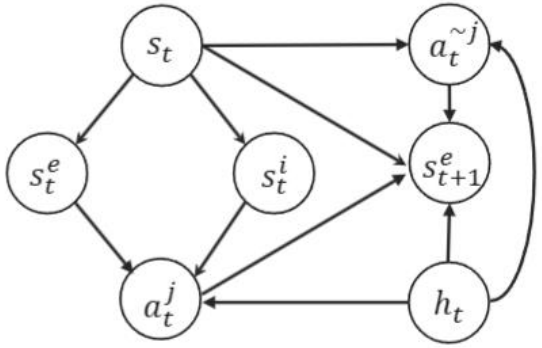
$$
\begin{aligned}
Y_{s_{t+1}^e}\left(A_t^j = a_t^j\right) &= P\left(s_{t+1}^e \mid do\left(A_t^j = a_t^j\right)\right) \\
&= \sum_{w_t} P\left(s_{t+1}^e \mid a_t^j, w_t\right) P\left(w_t \mid do\left(A_t^j = a_t^j\right)\right) \\
&= \sum_{w_t} P\left(s_{t+1}^e \mid a_t^j, w_t\right) P(w_t) \\
&= \sum_{w_t} P\left(s_{t+1}^e \mid a_t^j, w_t\right)
\end{aligned}
$$

$P(w_t) = 1$

where $w_t = \{s_t, h_t, a_t^{\sim j}\}$ is a node set;

03 METHOD

Definition of lazy agents in multi-agent systems



(d) Diligent agent

The treatment effect of a_t^j to s_{t+1}^e can be mathematically calculated as:

$$\tau_{a_t^j} = D_{KL}\left[Y_{s_{t+1}^e}(A_t^j = a_t^j) \| Y_{s_{t+1}^e}(A_t^j \neq a_t^j)\right] = D_{KL}\left[P(s_{t+1}^e | a_t^j, w_t) \| P(s_{t+1}^e | w_t)\right],$$

Similarly, we can also calculate the treatment effect of joint action $\vec{a}_t$ to s_{t+1}^e:

$$\tau_{\vec{a}_t} = D_{KL}\left[Y_{s_{t+1}^e}(\vec{A}_t = \vec{a}_t) \| Y_{s_{t+1}^e}(\vec{A}_t \neq \vec{a}_t)\right] = D_{KL}\left[P(s_{t+1}^e | \vec{a}_t, s_t, h_t) \| P(s_{t+1}^e | s_t, h_t)\right],$$

Definition 2: The agent j is a fully lazy agent iff:

$$\sum_{t=0}^{T} \tau_{a_t^j} = 0.$$

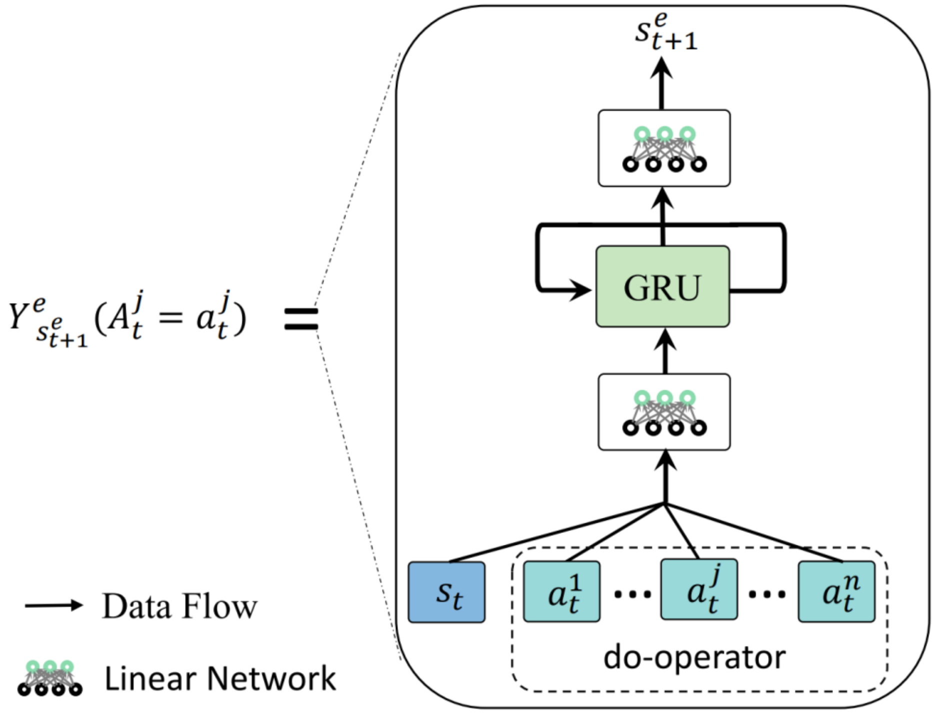
Definition 3: The team is fully lazy iff

$$\sum_{t=0}^{T} \tau_{\vec{a}_t} = 0.$$

03 METHOD

➢ External States Transition Model (ESTM):

- With ESTM, we can conveniently conduct do-operator on agent's action and calculate a potential outcome.
  - Since the state transition is always fixed given s_t, h_t and $\vec{a}_t$, $P(s^e_{t+1}|do(\vec{a}_t), s_t, h_t)$and $P(s^e_{t+1}|s_t, h_t)$ obey one-point distribution.
  - Therefore we replace the KL-divergence of distribution with MSE loss of external states to calculate the treatment effect in practice: $\tau_{a_t^j} = MSE(Y_{s^e_{t+1}}(A_t^j = a_t^j), Y_{s^e_{t+1}}(A_t^j \neq a_t^j))$.



We can enumerate all other available actions of the agent and calculate the mean value of these potential counterfactual outcomes as $Y^e_{s_{t+1}}(A_t^j \neq a_t^j)$.

$$\tau_{a_t^j} = MSE(Y_{s^e_{t+1}}(A_t^j = a_t^j), \frac{1}{|A|-1} \sum_{k=1, a_k \neq a_t^j}^{|A|} Y_{s^e_{t+1}}(A_t^j = a_k))$$

where $|A|$ is the size of action set A.

03 METHOD

➢ Intrinsic motivation:

- Idividual Diligence Intrinsic motivation (IDI) can be calculated as:

$$r_t^{IDI} = \sum_{j=1}^{|N|} \tau_{a_t^j}.$$

- Collaborative Diligence Intrinsic motivation (CDI) can be calculated as:

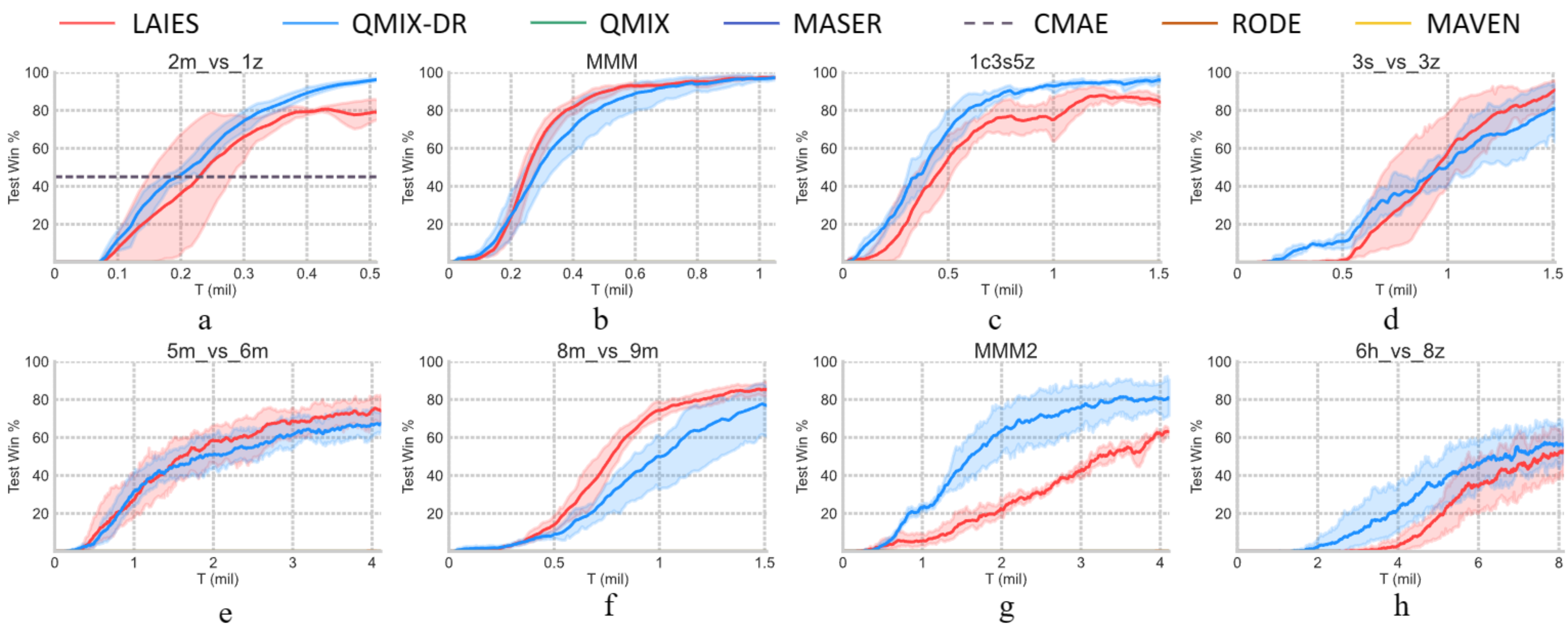
$$r_t^{CDI} = MSE(Y_{s_{t+1}^e}(\vec{A}_t = \vec{a}_t), \frac{1}{|A|^n - 1} \sum_{k=1, \vec{a}_k \neq \vec{a}_t}^{|A|^n} Y_{s_{t+1}^e}(\vec{A}_t = \vec{a}_k)).$$

- Ultimate intrinsic reward:

$$r^I = \beta_1 * r^{IDI} + \beta_2 * r^{CDI},$$

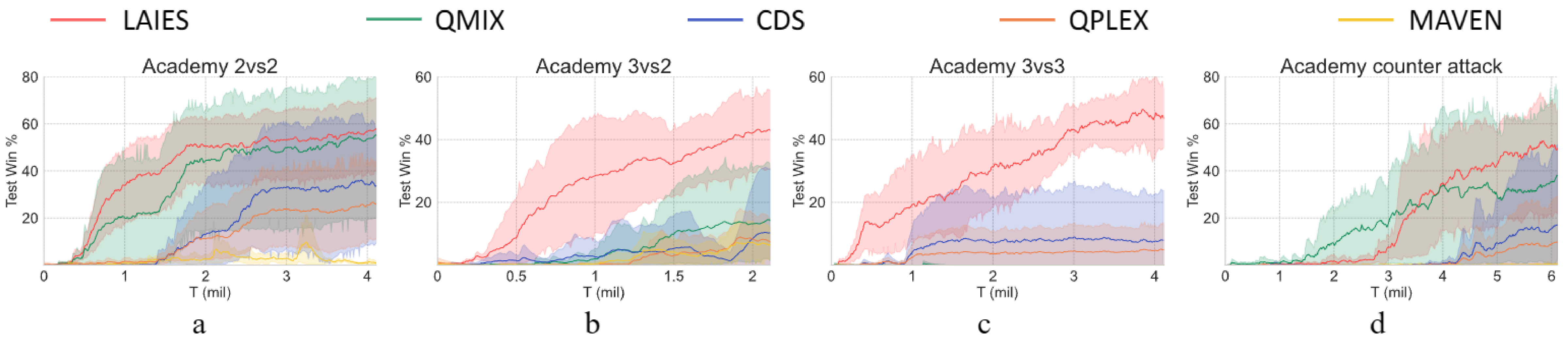
04 EXPERIMENTS

➢ Results (Experiments on StarCraft 2 with Sparse Rewards)



04 EXPERIMENTS

➤ **Results (Experiments on GRF with Sparse Rewards)**



05 CONCLUSION

This paper investigates the sparse reward problem in MARL with a new perspective, i.e., lazy agents. Empirical results demonstrate that our proposed method achieves state-of-the-art performance on various tasks.

However, this paper has two main limitations.

- Firstly, it is possible to categorize states as internal and external.
- Secondly, this paper defines 'lazy agents' regarding their influence on external states, but their laziness may not be limited to not impacting external states.

Therefore, both the definition and solution of lazy agents are worthy of further studies.

THANKS

E-mail: zhiqiang.pu@ia.ac.cn

E-mail: liuboyin2019@ia.ac.cn