A Q-Values Sharing Framework for Multiple Independent Q-Learners

Final Project

Presented by Denaldo Lapi, Hailey S. Miles

Contents

Paper Summary

Setting:

- Cooperative MARL involving multiple IQL sharing the same environment

Develop a *partaker-sharer* advising framework (PSAF) where agents must also learn when to ask for Q-values and when to provide Q-values

Partaker

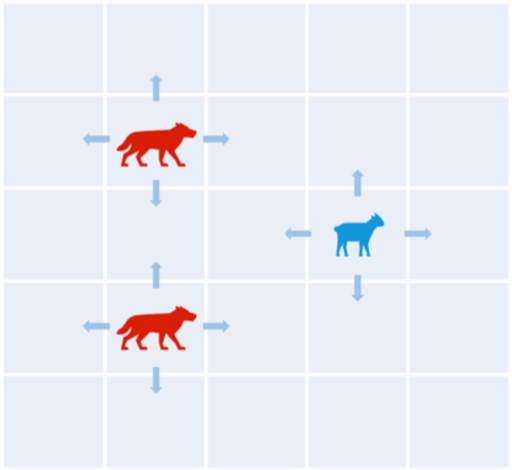
- can request Q-values for insufficiently explored states.

Sharer

- Agents evaluate their confidence of the max Q-value. Sharer shares when its confidence is higher than the partaker.

Simulation Procedure

Predator-Prey Scenario



Environment

- Grid world
- Predators are agents
- Prey takes evasive action 80% of the time, random action 20% of the time.
- *Goal state*: One predator is in the same square as the prey, the other is in an adjacent square.
- *Time to goal*: Number of steps necessary to reach prey.

Agent

Actions

Up	Moves 1 grid space in y direction
Down	Moves 1 grid space in -y direction
Left	Moves 1 grid space -x direction
Right	Moves 1 grid space in x direction
Nothing	Does not move

State
<x1, y1, x2 y2>
Relative distances to predator and prey.

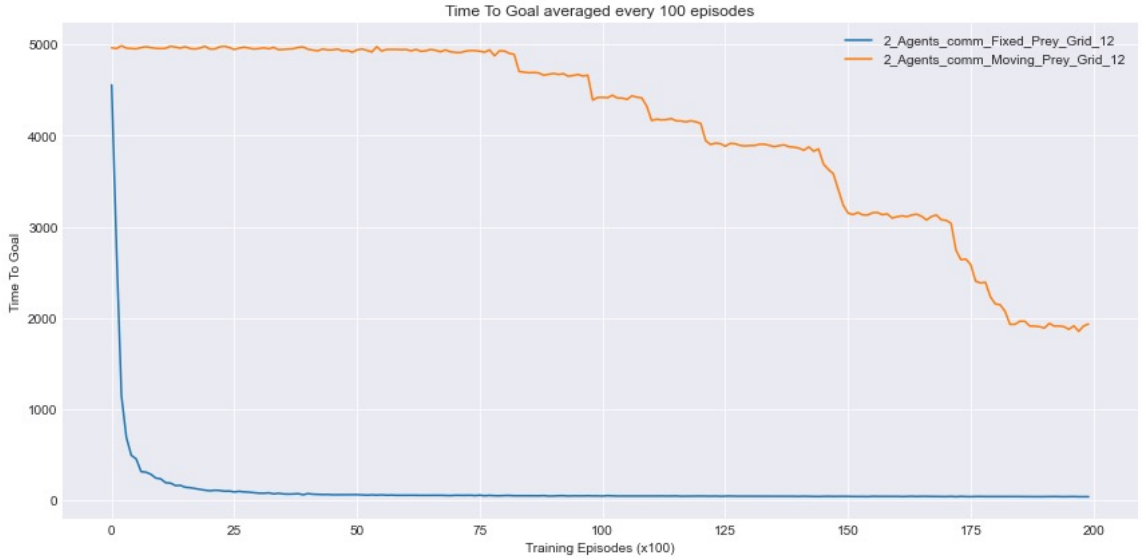
$$\textit{Reward} = \begin{cases} 1, if\ predator\ is\ in\ the\ same\ grid\ space\ as\ the\ prey \\ and\ the\ other\ predator\ is\ in\ an\ adjacent\ grid\ space \\ 0, \quad otherwise \end{cases}$$

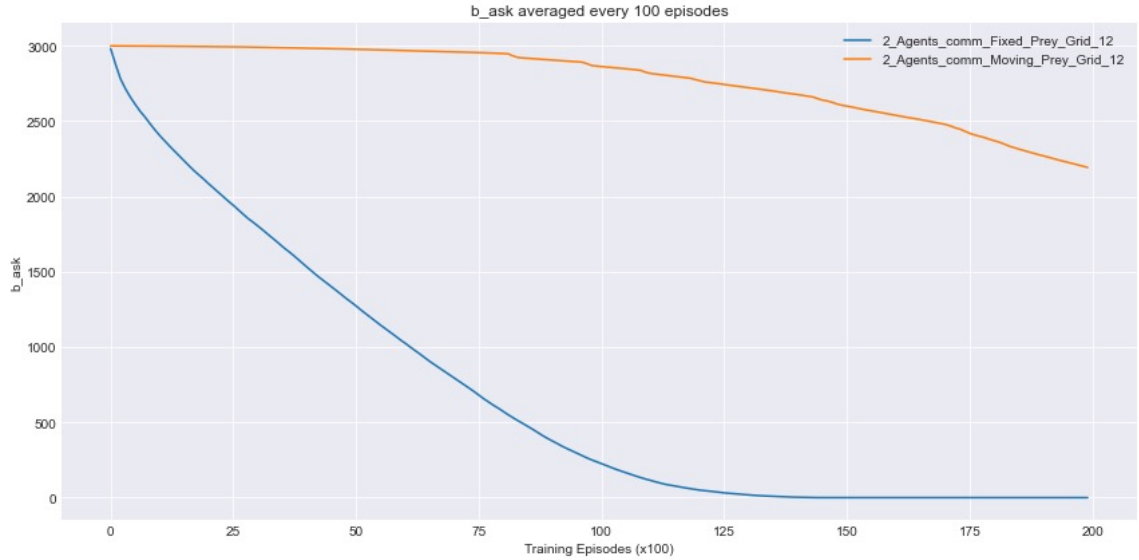
Demo

SOAS Demo

Results

Test 1: Moving prey vs. Stationary prey

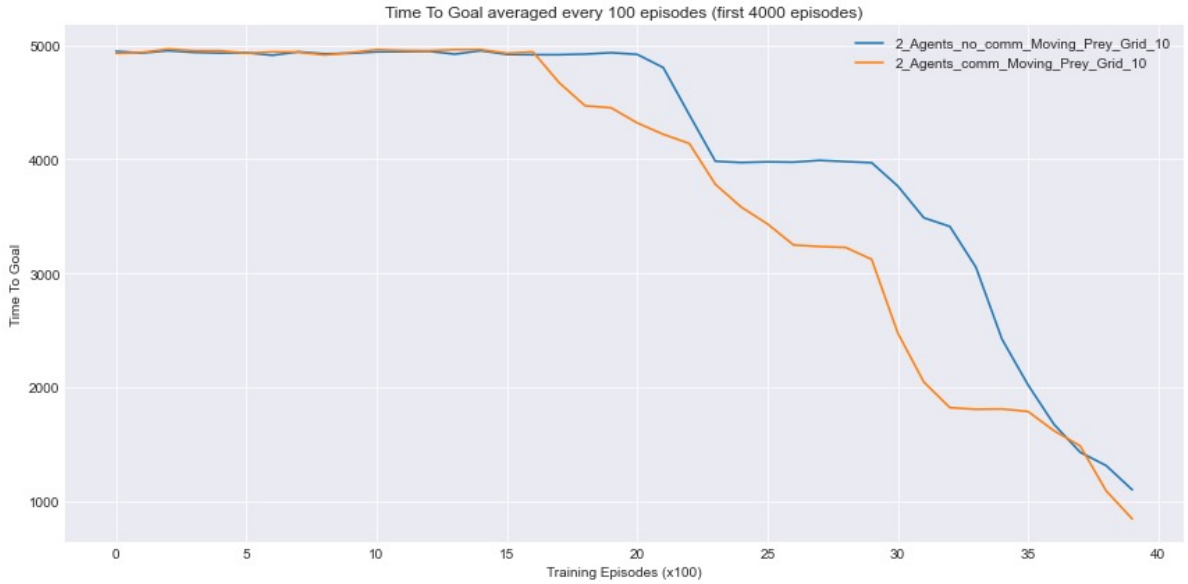


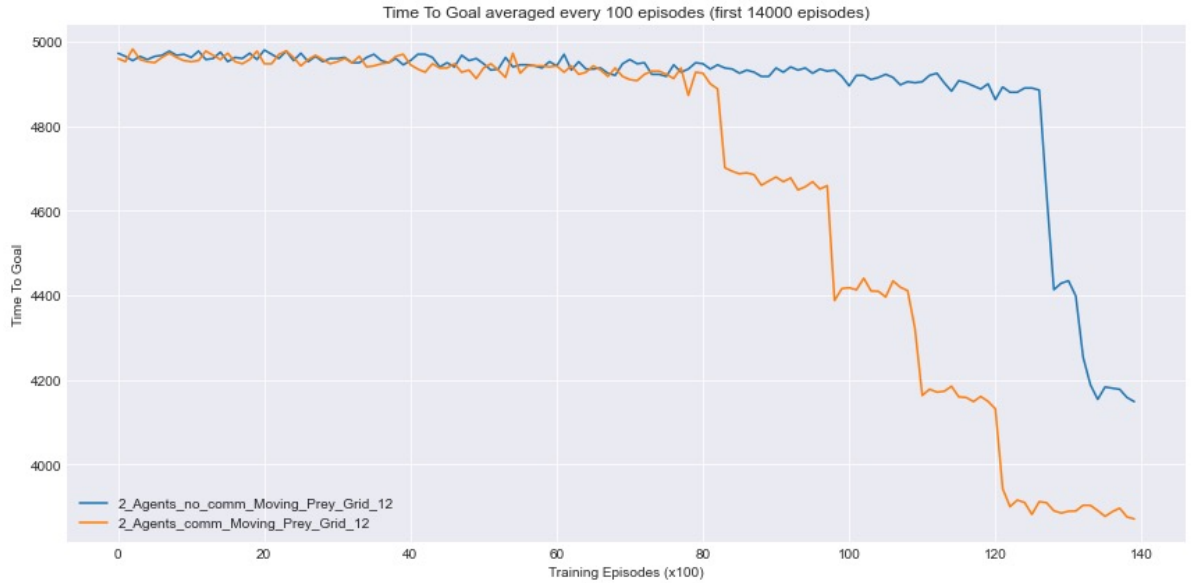


How does the behavior of the prey affect learning?

Results

Test 2: PSAF vs Multiple IQL

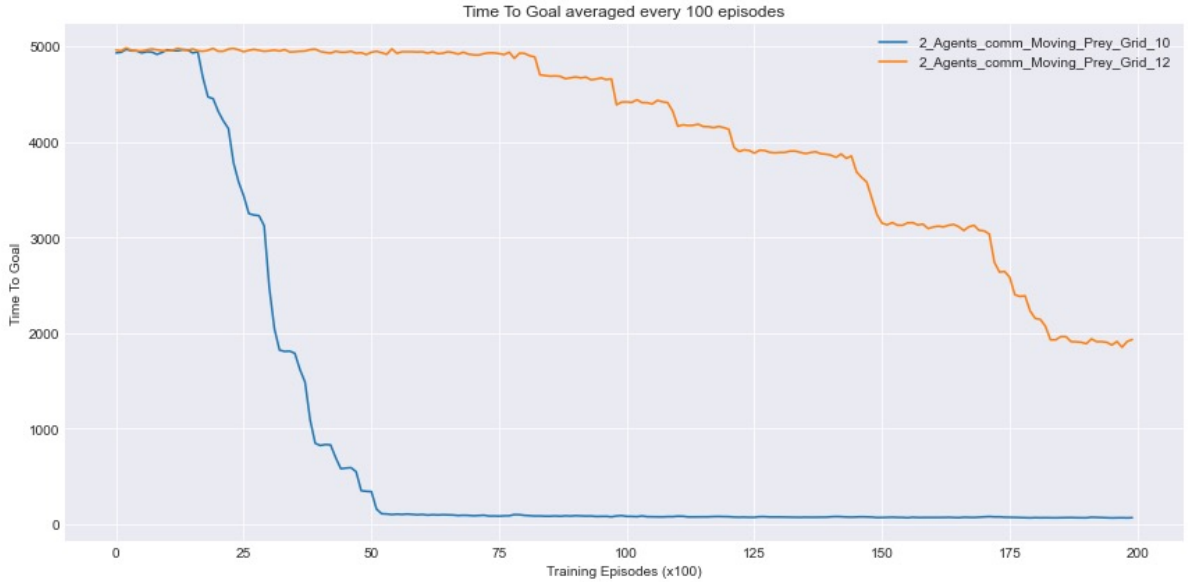


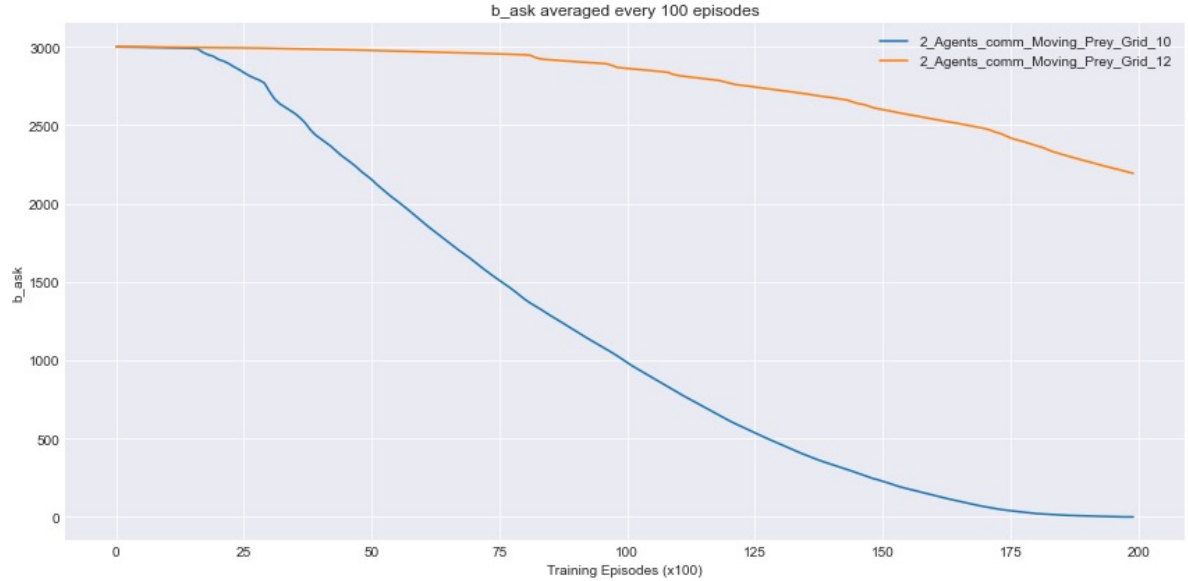


Does communication between agents accelerate learning?

Results

Test 3: PSAF with varying grid sizes

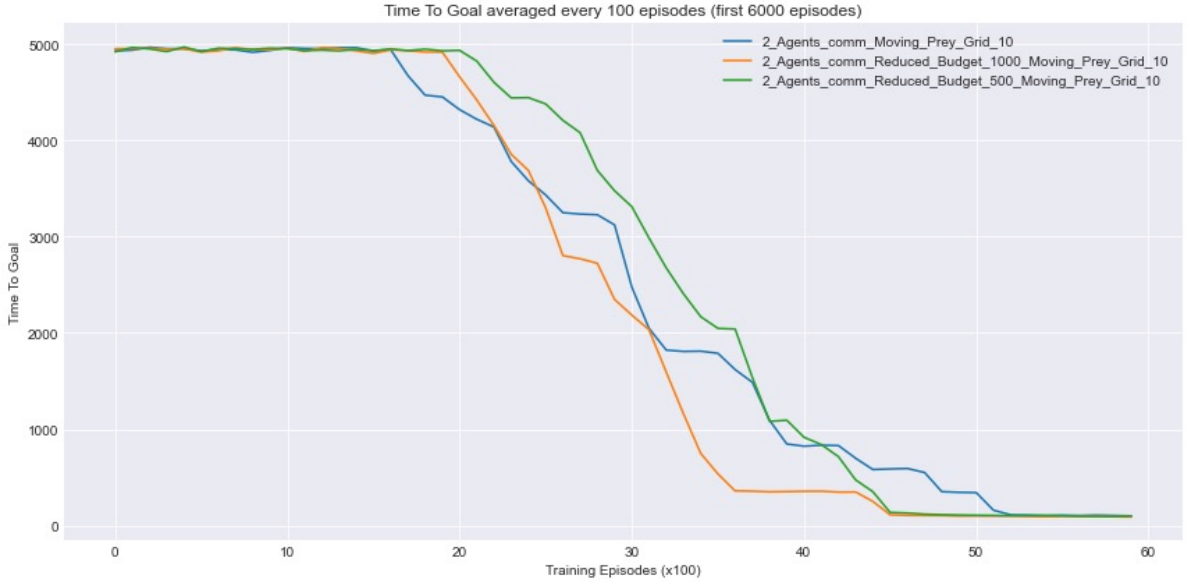


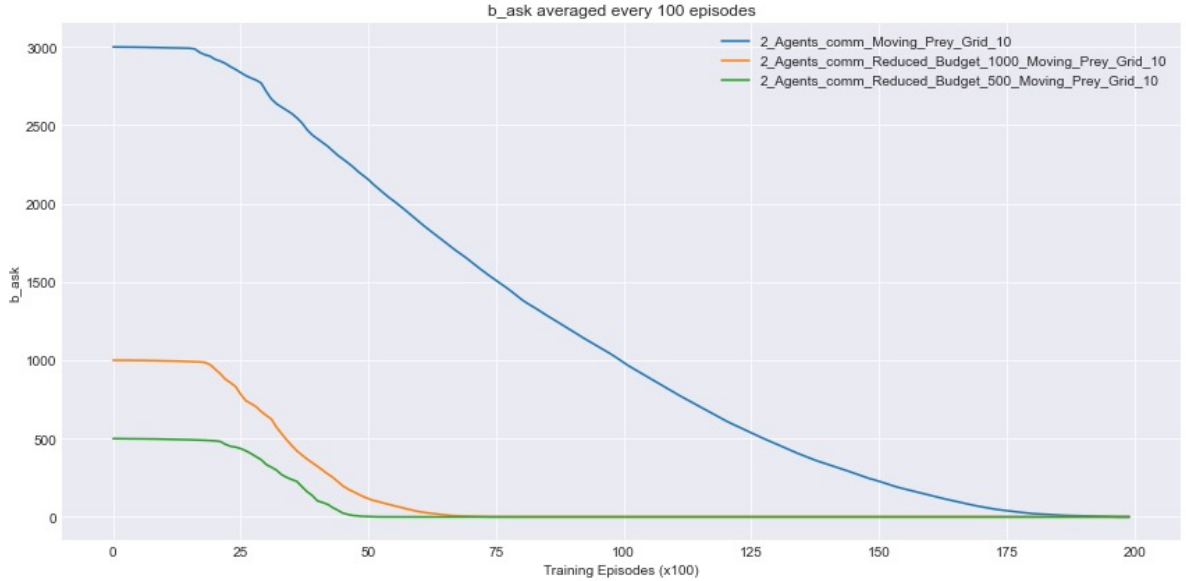


What affect does a larger grid size have on learning?

Results

Test 4: Various budget sizes





Does reducing the budget affect learning?

Conclusions

- PSAF is better suited to complex scenarios
- PSAF accelerates learning compared to multiple IQL
- Larger grid sizes are much more complex as more states have to be explored and thus learning is slower
- The ask and give budgets are most critical at the beginning if the learning stages.

Questions

Thank you for listening