Многоагентное обучение с подкреплением с непрерывным пространством действий

Пшеницын А.С.
pshenitsyn.as@phystech.edu

Центр когнитивного моделирования МФТИ

2025 – Статус НИР 11.05





Мотивация



Многоагентное планирование — одна из наиболее сложных задач для навигации роботов. Существующие методы направлены на решение задачи в дискретных средах (NP-hard), потому для перехода в непрерывные состояния часто используют дискретизацию пространства, и построенные пути не являются оптимальными.

Цель работы

Разработка методов обучения с подкреплением для многоагентного планирования в непрерывных средах, способных к масштабированию и генерализации по числу агентов и сложности среды.



Постановка задачи

Децентрализованный мультиагентный частично наблюдаемый процесс принятия решений - это $M = \langle S, I, (A_i), P, (\Omega_i), (O_i), (r_i), \gamma \rangle$, где:

- S - пространство состояний,
- $I = \{1, \ldots, n\}$ - агенты,
- A_i - пространство действий для агента i
- $A = \times_i A_i$ - пространство объединенных действий
- $P = P(s'|s, \vec{a})$ - функция переходов,
- Ω_i - пространство наблюдений для агента i,
- $O_i = O_i(s, a)$ - функция наблюдения для агента i,
- $r_i = r_i(s, \vec{a})$ - функция награды для агента i,
- γ - фактор дисконтирования

Постановка задачи

В задаче планирования:

$$x_i(t) = \int_0^t f(x_1(\xi), x_2(\xi), ...a_1(\xi), a_2(\xi), ...)d\xi$$

- f - функция переходов (динамика) среды,
- $x_i(t)$ - положение агента i в момент времени t,
- $x_i(T) = \mathtt{goal_i}$ - цель агента i,
- $\forall i \neq j \; \forall 0 < t < T \quad \|x_i(t) - x_j(t)\| > r_i + r_j$ - избегание столкновений,
- $\forall 0 < t < T \quad dist(x_i(t), Obstacle) > r_i$ - избегание препятствий

Метрики качества алгоритмов:

- $T_{avg} = \frac{1}{N}\sum_{i=0}^{N} T_i$ - среднее время достижения всех целей агентами
- $success_rate = \frac{\sum |x_i(T) = goal_i|}{N}$ - доля корректных путей
- Время построения плана*

Задачи

Center for Cognitive Modeling

- Проанализировать существующие среды с непрерывными пространствами состояний и действий для создания сценариев многоагентного планирования с большим числом агентов
- Разработать эффективную среду для задач многоагентного планирования с непрерывным пространством состояний и действий
- Реализовать MARL бейзлайны (MAPPO) и классические алгоритмы на полученных сценариях

VMAS

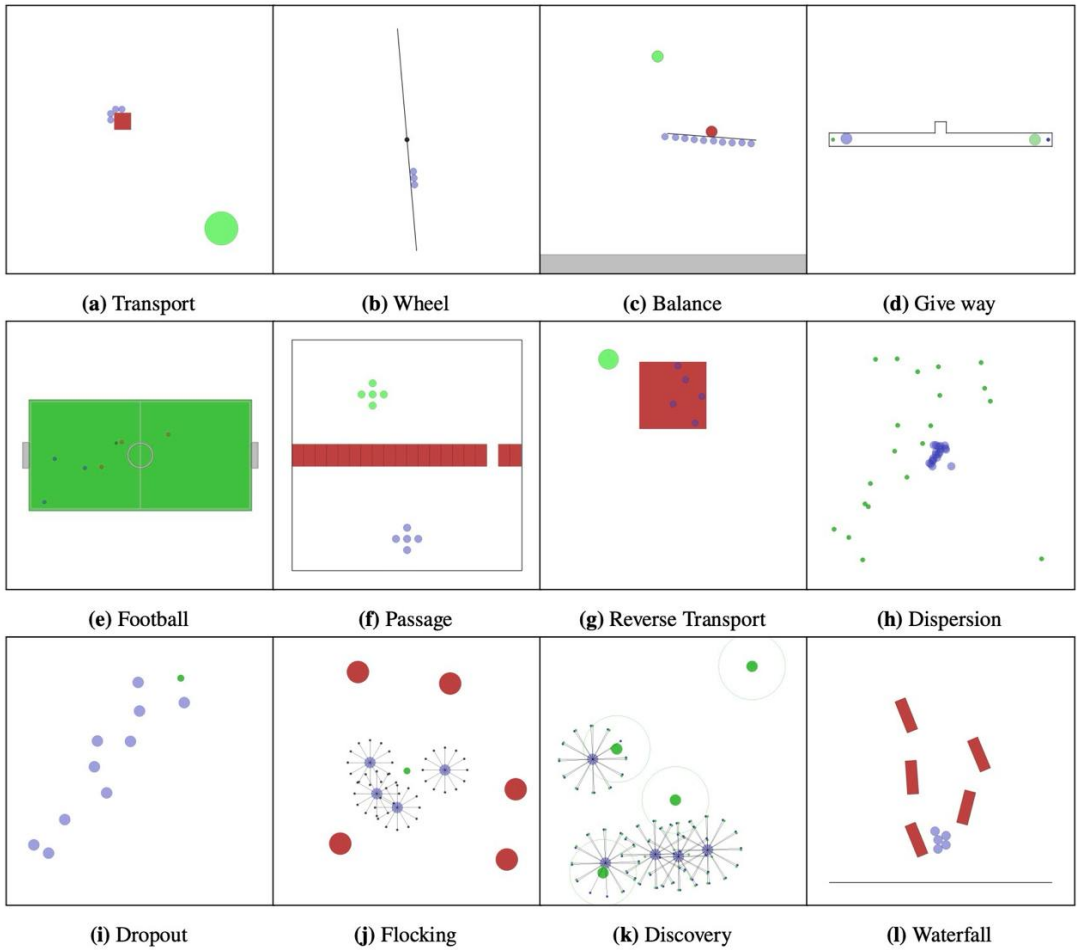
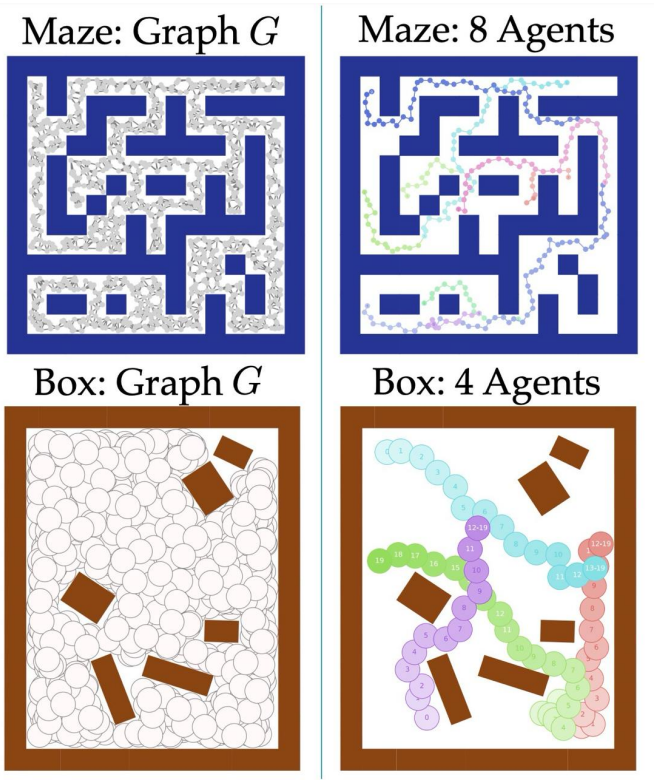
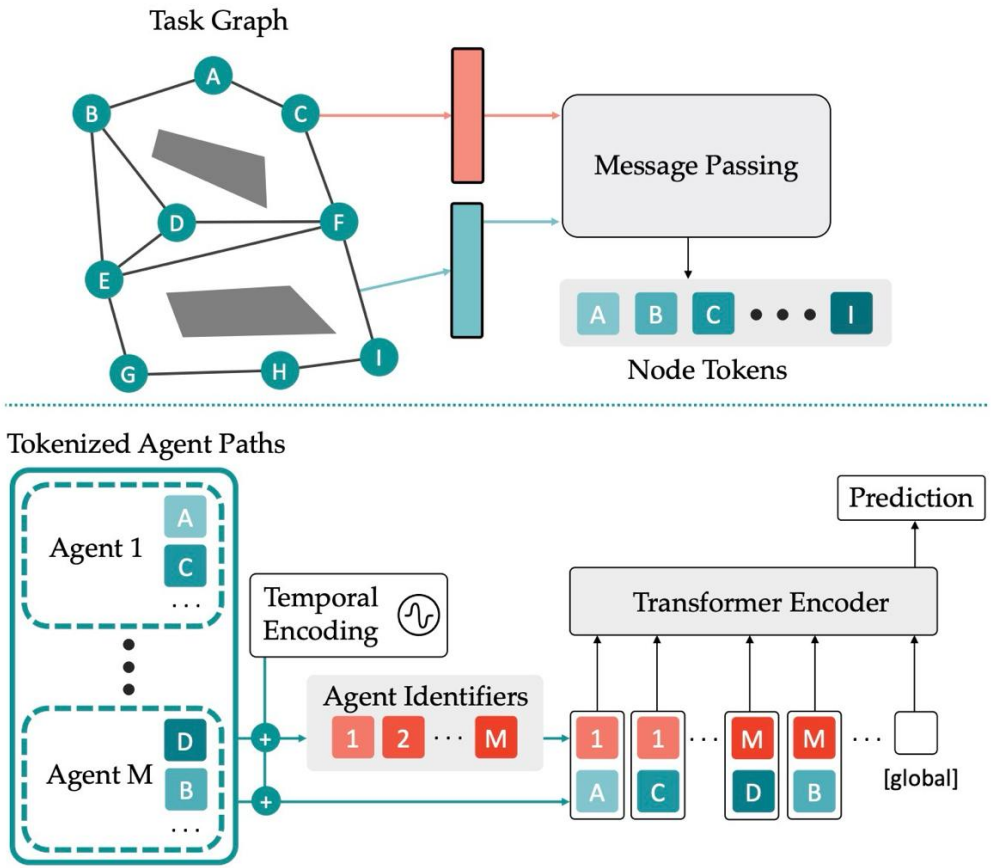


Рис.: VMAS [source]

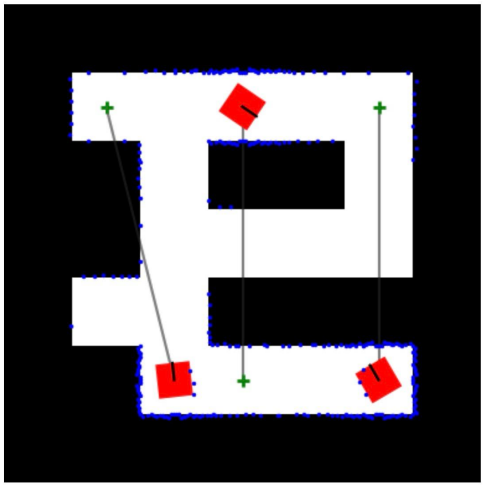
Multi-Agent Planning Using Graph Transformers



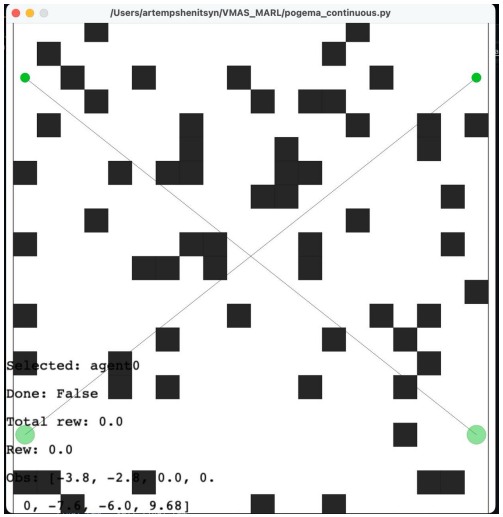


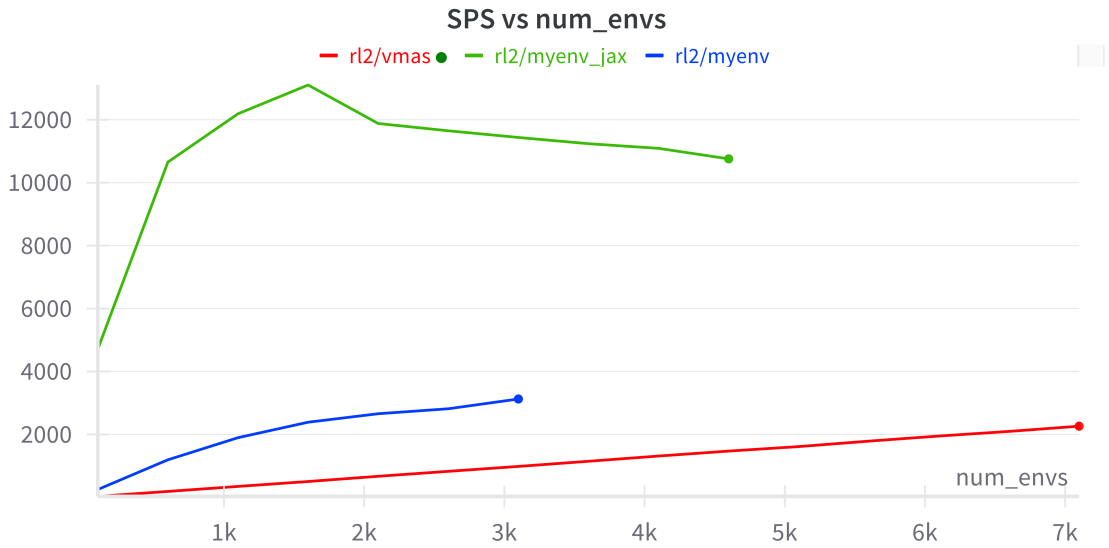
source

Анализ сред



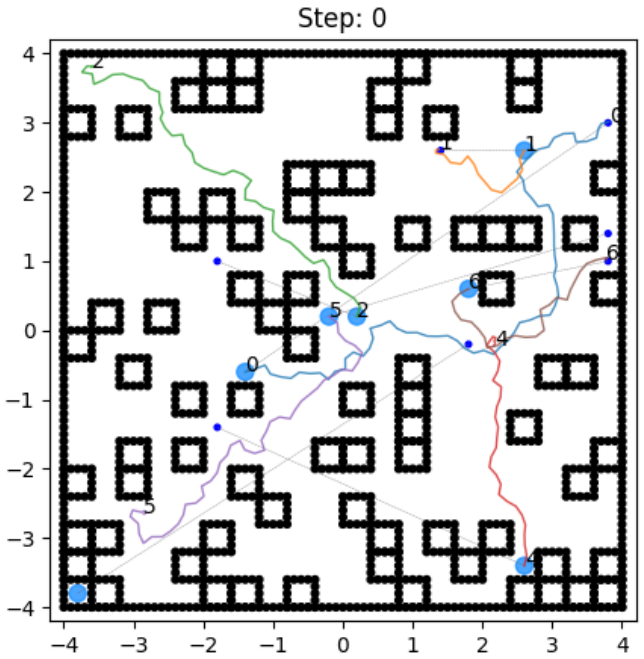
(b) JaxNav

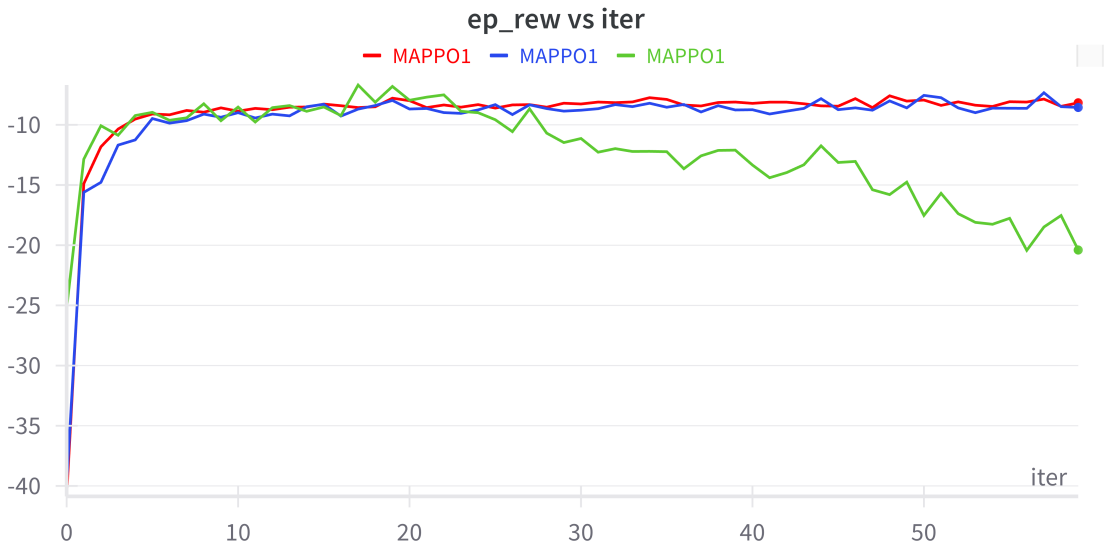




link: env.mp4

Бейзлайны





link: pd_rrt.mp4

Дальнейшая работа

Center for Cognitive Modeling

- Интегрировать movingai бенчмарк для тестирования (с генерацией случайных позиций агентов и целей)
- Добавить генератор случайных непрерывных препятствий
- Добавить процедурную генерацию разных карт
- Модифицировать RRT до RRT-star
- Добавить RRT-based наблюдения и награды
- Добавить возможность Lifelong MAPF сценария
- Приступить к непосредственной разработке методов планирования

Спасибо за внимание!

cogmodel.mipt.ru

pshenitsyn.as@phystech.edu