Разработка методов навигации для роботехнических систем на основе обучения с подкреплением и оптимального управления

Осколков Н.С.
oskolkov.ns@phystech.edu

Научный руководитель: канд. физ. наук Макаров Д.А

Центр когнитивного моделирования МФТИ

2025 – Научно-исследовательский семинар

Мотивация



Рис.: Колесный робот помощник "Temi Personal Robot"

Цель исследования

Целью исследования ставится используя оптимальное управление обучить с помощью обучение с подкреплением колесного робота доезжать до выбранного текстовой командой объекта из ограниченного множества объектов избегая столкновения с случайно расставленными статическими препятствиями на сцене.

Формальная постановка задачи

Мы моделируем задачу навигации как частично наблюдаемую марковскую задачу принятия решений (POMDP): (Z, S, A, P, R, γ). Z — реальное состояние, S — наблюдаемое состояние, A — действия, R — награды, P — динамика переходов $(S \times A \to S)$, $\gamma \in (0, 1]$ — коэффициент дисконтирования. Цель: найти оптимальную политику π, максимизирующую награду:

$$\eta(\pi) = \mathbb{E}_\pi \left[\sum_{t=0}^{T-1} \gamma^t r(S_t, a_t) \right]. \tag{1}$$

Задача: робот должен достичь цели, избегая препятствий , используя данные сенсоров. Эпизод завершается при достижении цели.

Формальная постановка задачи

Процесс взаимодействия с окружающей средой представляет собой марковский процесс принятия решений (MDP), который требует определения **непрерывных**:

1. **пространства наблюдаемых состояний** $\mathcal{S}$ (State space):
   a) Данные с камеры - тензор формата $(400 \times 400 \times 3)$ - X_{cam},
   b) Текстовая команда X_{goal},
   c) Граф знаний G,
   d) Предсказане динамики D,
   e) Память M,
   f) Скорость базы $V_{current}$(2 dof): скорость состоит из двух компонент: линейной и угловой.
2. **пространства действий** A (Action space): скорость $V_{control} \in \overline{V}$, где $\overline{V}$ - класс допустимого управления,

$$V_{control} = f(X_{cam}, X_{goal}, D, G, MV_{current}).$$

Обзор используемых методов

Using deep reinforcement learning with automatic curriculum learning for mapless navigation in intralogistics:
Применение в задачи навигации обучения с расписанием
Reinforced imitation: Sample efficient deep reinforcement learning for mapless navigation by leveraging prior demonstrations:
Применение в задачи навигации имитационного обучения
Learning object relation graph and tentative policy for visual navigation. Применили граф объектных отношений, который моделирует вероятные пространственные связи между объектами и имитационное обучение.

Описание используемого метода

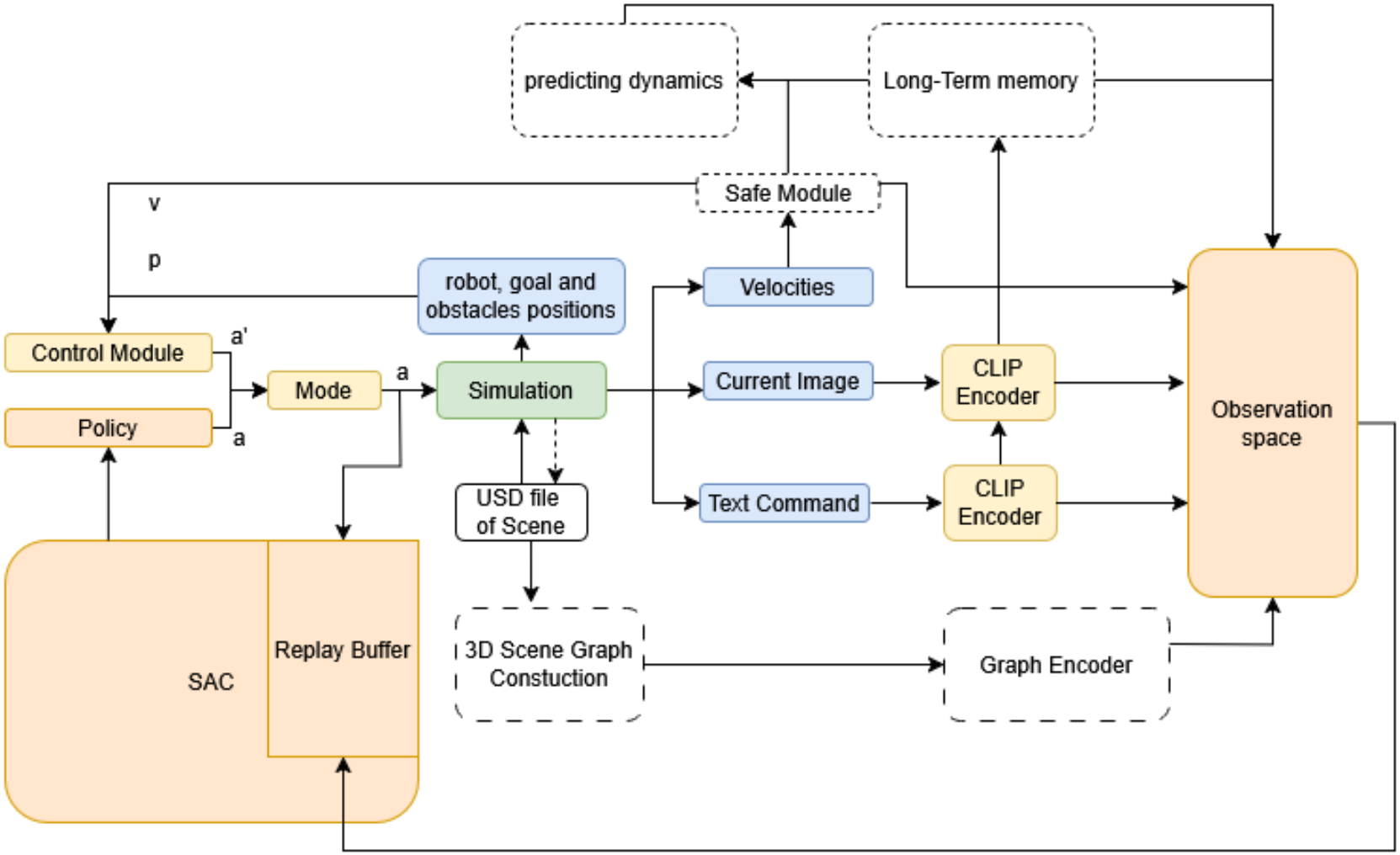


Рис.: Общая схема пайплайна

Описание используемого метода

1. Подзадача 1: ошибка по расстоянию < 1 м.
2. Подзадача 2: ошибка по углу $< \frac{\pi}{20}$

Функция награды.

1) не выполнена ни одна из подзадач: $reward = -\frac{2}{10}$
2) выполнена только 1 подзадача: $reward = -\frac{1}{10}$
3) выполнены обе подзадачи: $reward = 2$
4) за превышение времени или коллизии: $reward = -5$

Выделенные подзадачи

1. реализовать обучение с расписанием ✓
2. реализовать имитационное обучение ✓
    1. реализовать глобальное планирование пути (оптимальное планирование) ✓
    2. реализовать алгоритм следования (эврестическое управление) ✓
3. реализовать граф знаний и локализацию внутри графа
4. реализовать кратковременную память
5. реализовать долговременную память
6. реализовать адаптивное управление

Модуль контроля (имитационное обучение)

Дейкстра

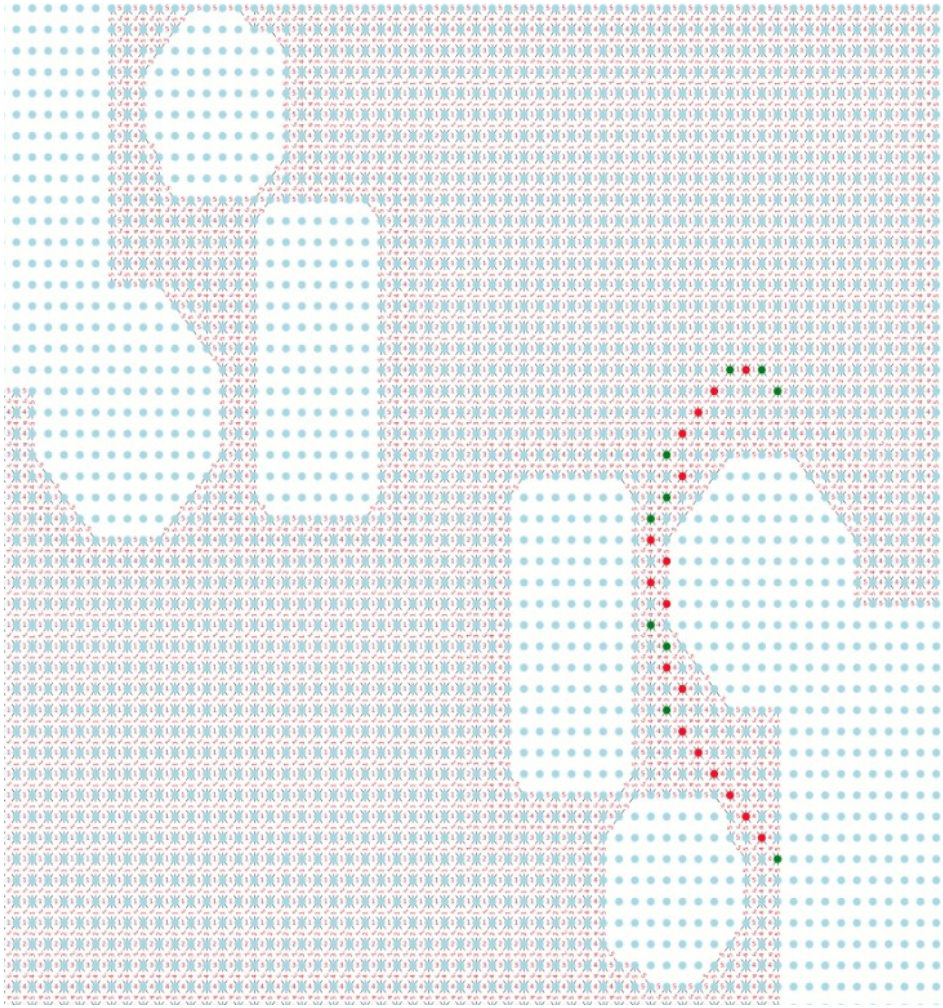


Рис.: Пример построенного алгоритмом Дейкстры и спрямленного пути

Модуль контроля (имитационное обучение)

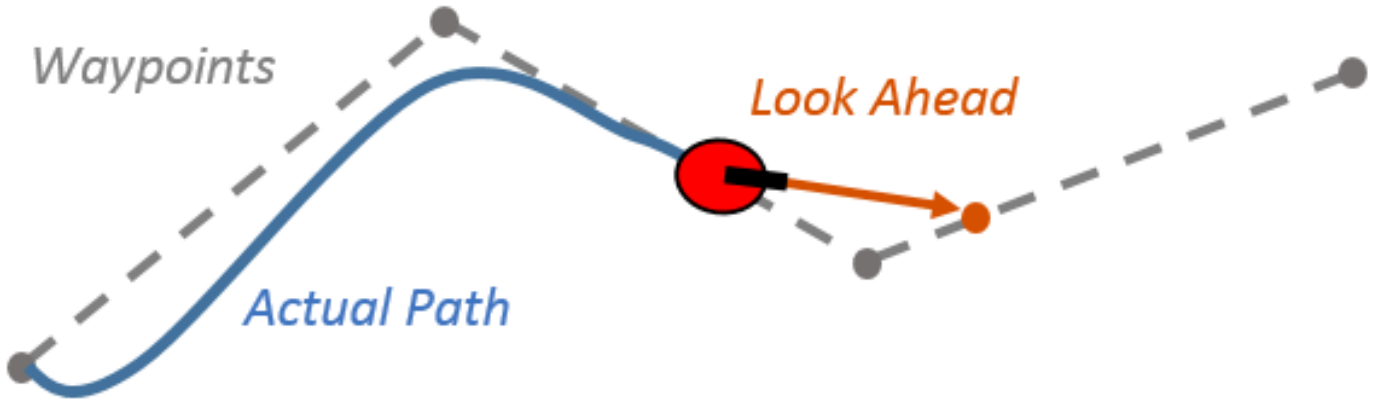


Рис.: Алгоритм следования Pure Persuit[1]

[1] Coulter RC. Implementation of the pure pursuit path tracking algorithm.

Обучение с расписанием

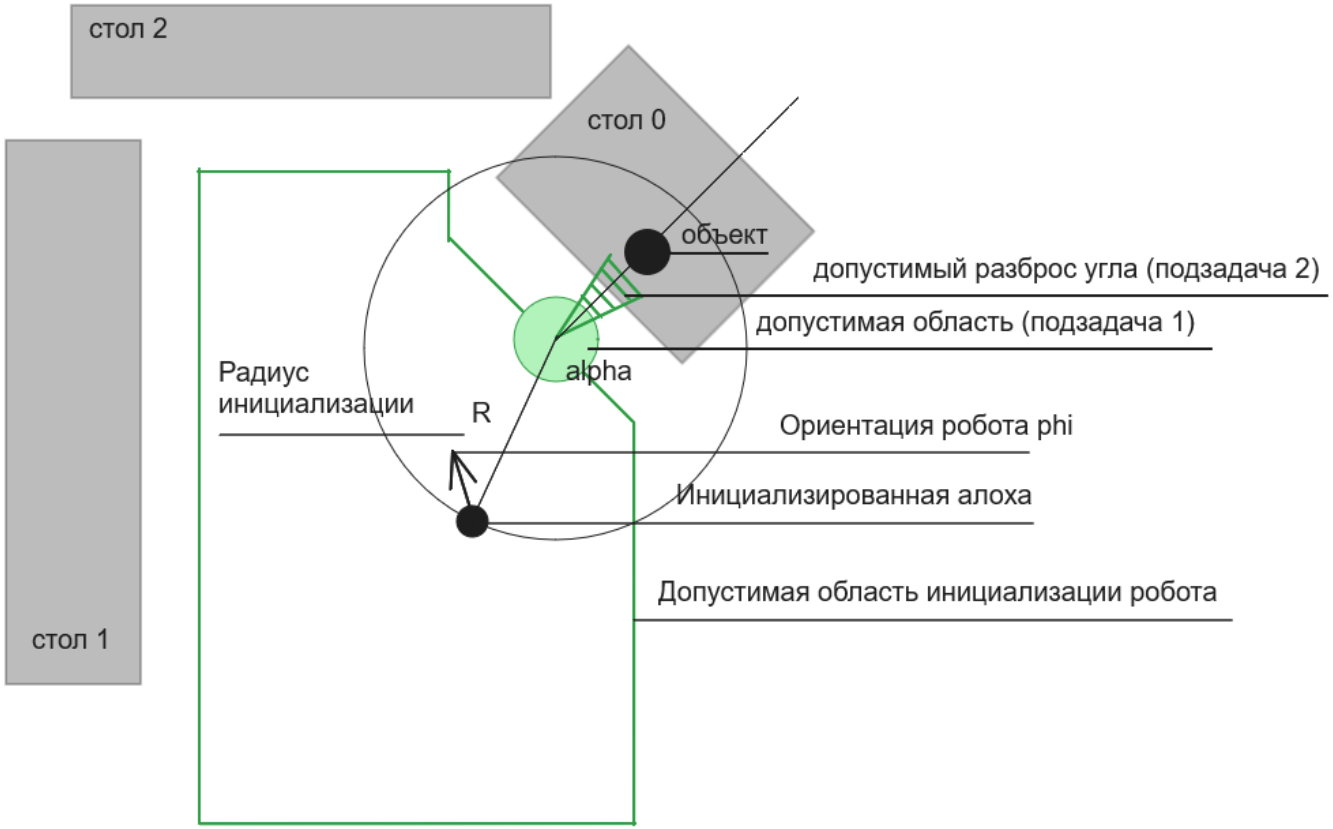


Рис.: Схема пайплайна эксперимента

Первые эксперименты

Технические характеристики: эксперементы производились в среде Isaac Sim на колесном роботе ALoha[2]





Рис.: Изображение с камеры робота в эксперимента с множеством выборов

[2]Fu Z, Zhao TZ, Finn C. Mobile aloha: Learning bimanual mobile manipulation with low-cost whole-body teleoperation.

Первые эксперименты (С множественным выбором)

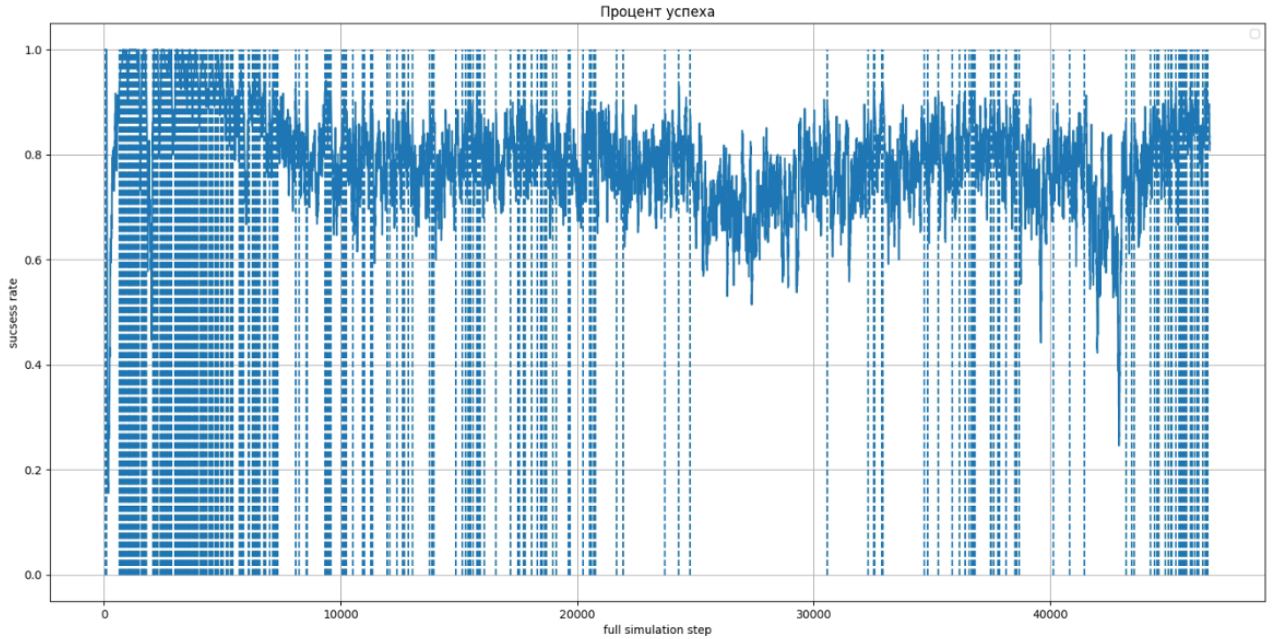


Рис.: График прцоента успеха от количесвта эпизодов в процессе обучения

Процент успеха 70% на стартовой дистанции в 2 метра

Первые эксперименты

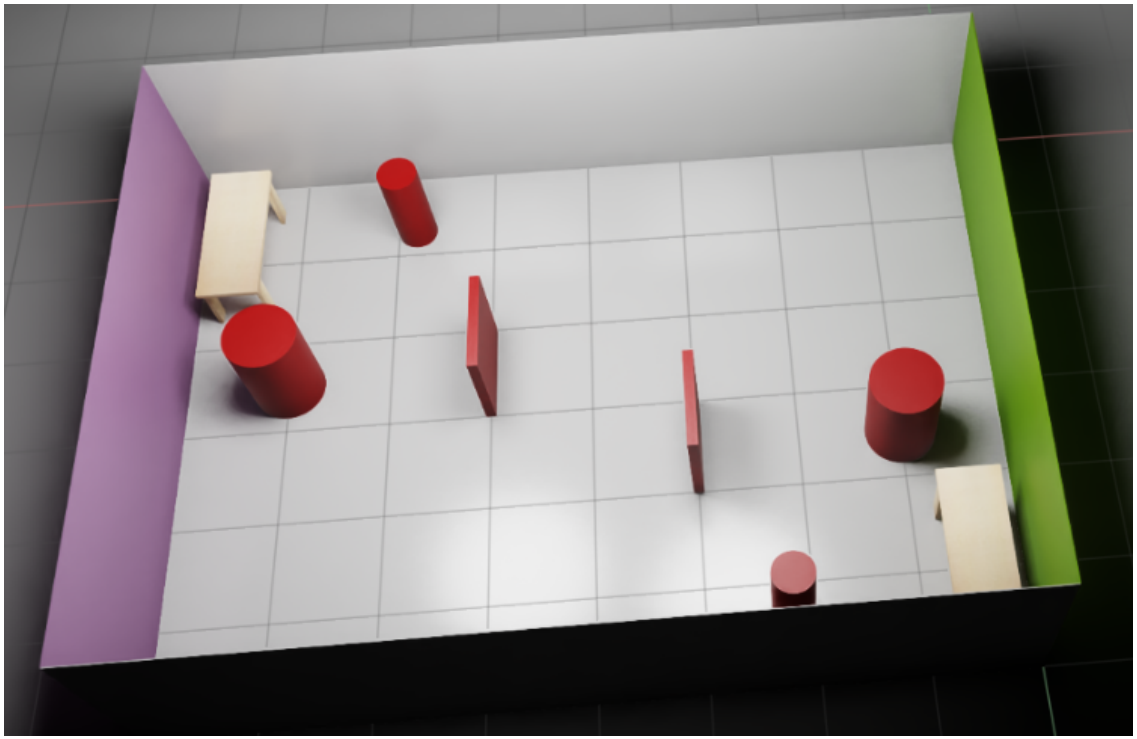
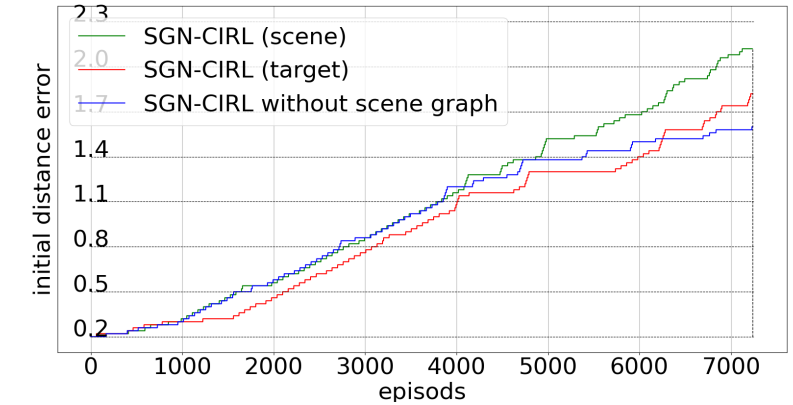


Рис.: Эксперимент со статическими препятствиями

Первые эксперименты (Со статическими препятствиями)



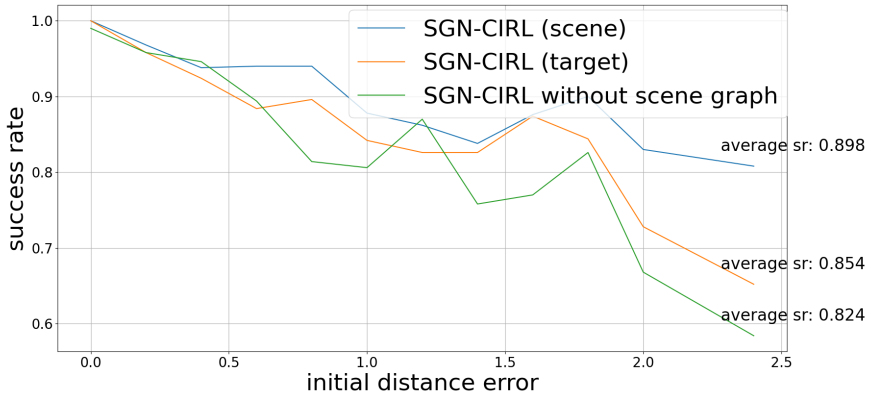


Рис.: Графики достигнутого уровня сложности от кол. эпизодов во премя обучения и процента успеха от стартовой дситанции между агентом и целевым объектом во время инференса