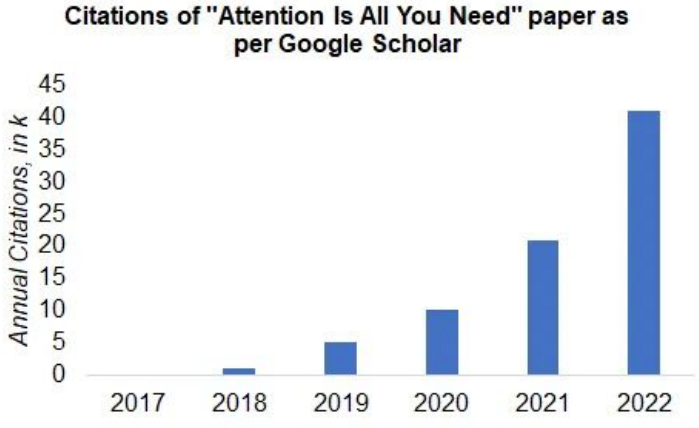
Обучение локомоции шарнирной системы методами обучения с подкреплением с учетом её структуры

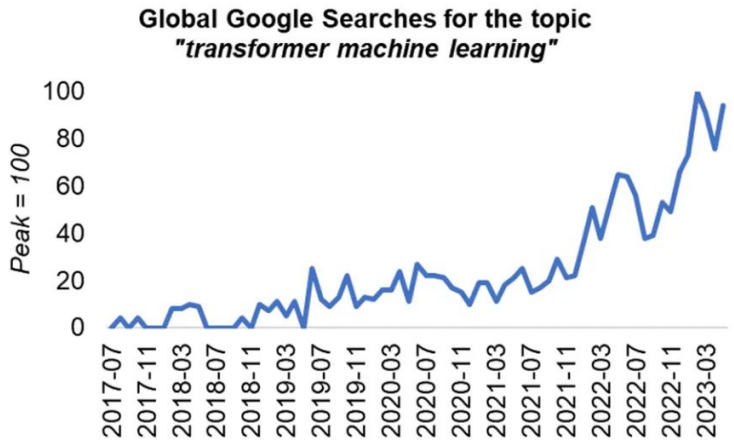
Афендульев Кирилл Дмитриевич

Содержание

Введение

Transformer захватил все сферы ML

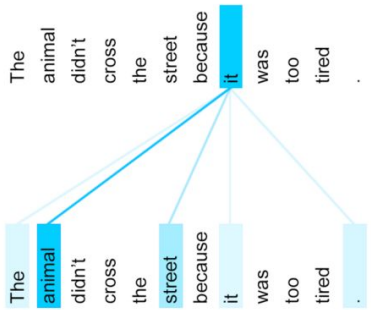


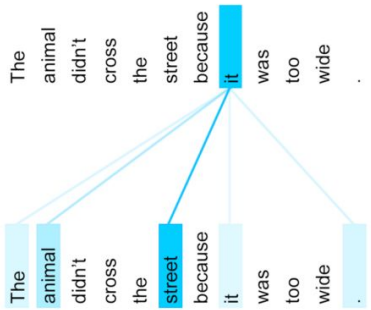


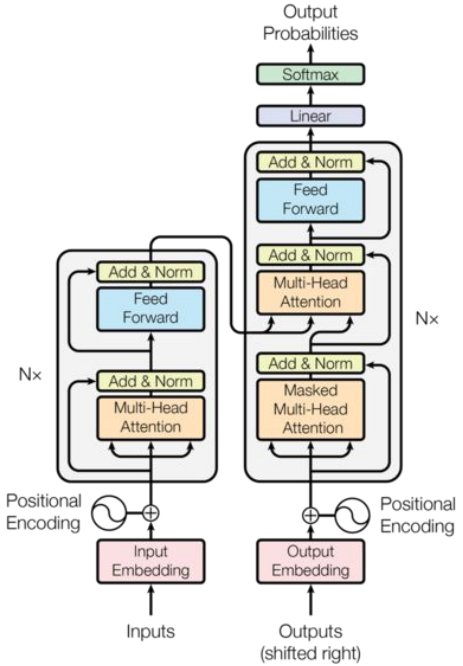
Примечание

Введение

Причина успеха в механизме Attention

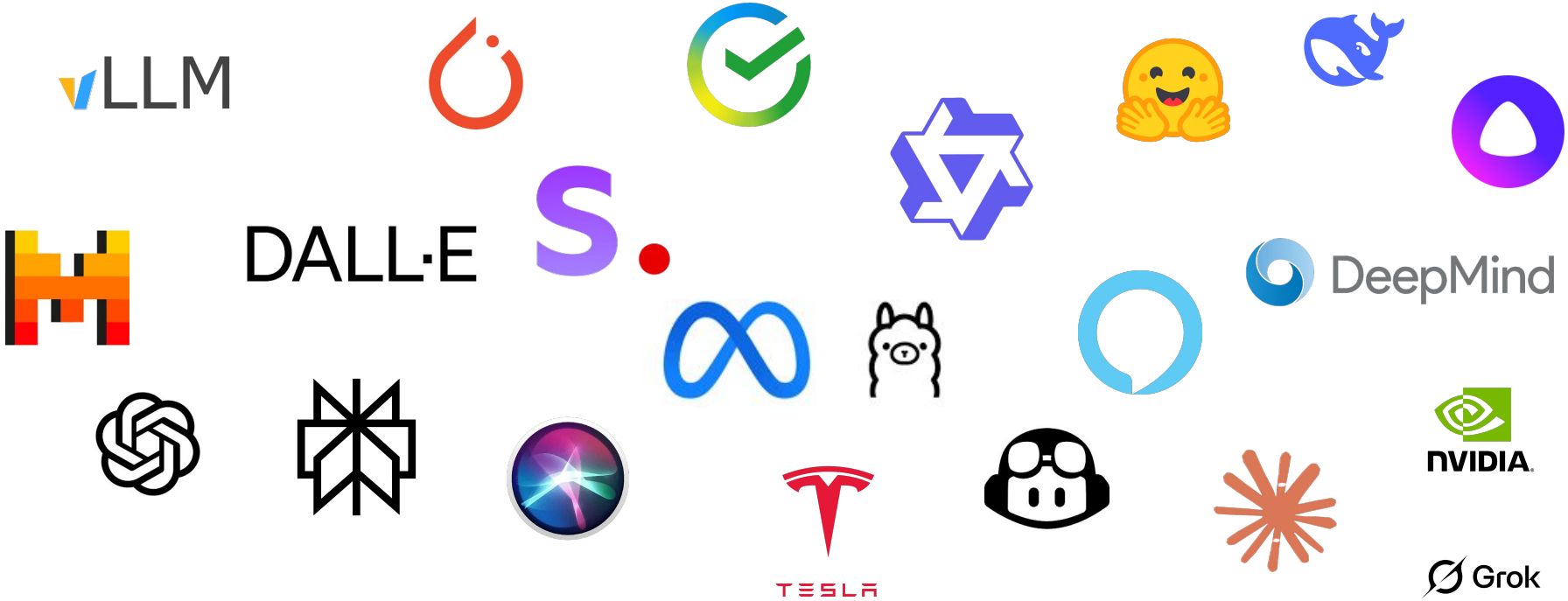






Введение

Transformer-based модели: стандарт де-факто в NLP и CV



Введение

Transformer-based модели в RL набирают обороты

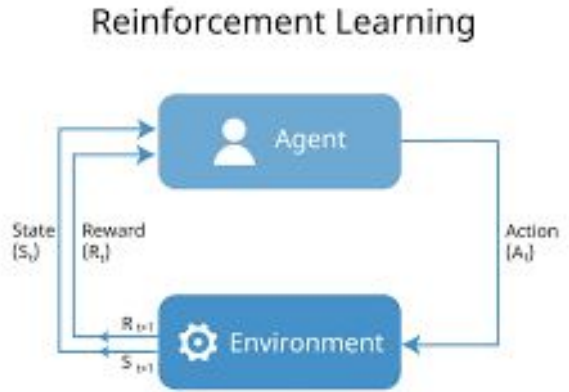
Динамика публикаций

2019: появление GTrXL (Parisotto et al.)

2021: Decision Transformer & Trajectory Transformer

2022–2023: Gato, Multi-Game DT, Bootstrapped Transformer

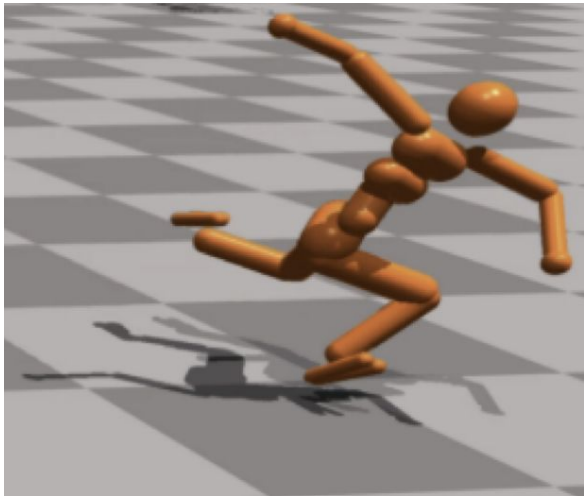
2024–2025: AGaLiTe, Transformer-based world models, Diffuser-RL



Проблема

Задача локомоции шарниой конструкции

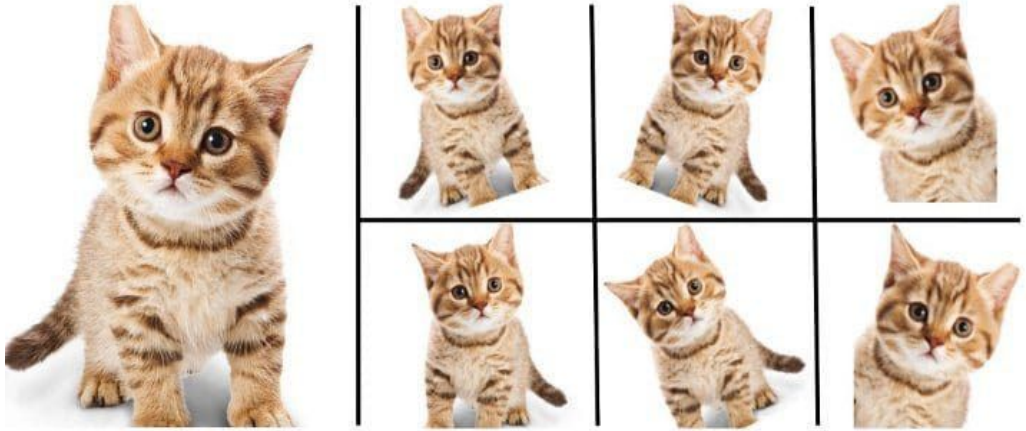
1. Награда за успешное прохождение маршрута
2. Штраф за нестаблиное поведение
3. Штраф за неэффективное энергопотребление



Постановка задачи

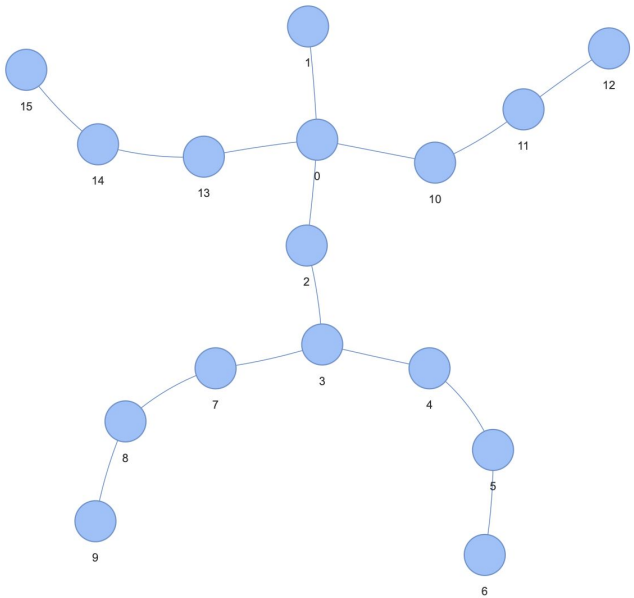
Внесение априорных знаний при обучении Transformer

1) Random Crop & Resize
2) Horizontal or Vertical Flip
3) Color Jitter
4) Synonym Replacement
5) Back-Translation



Постановка задачи

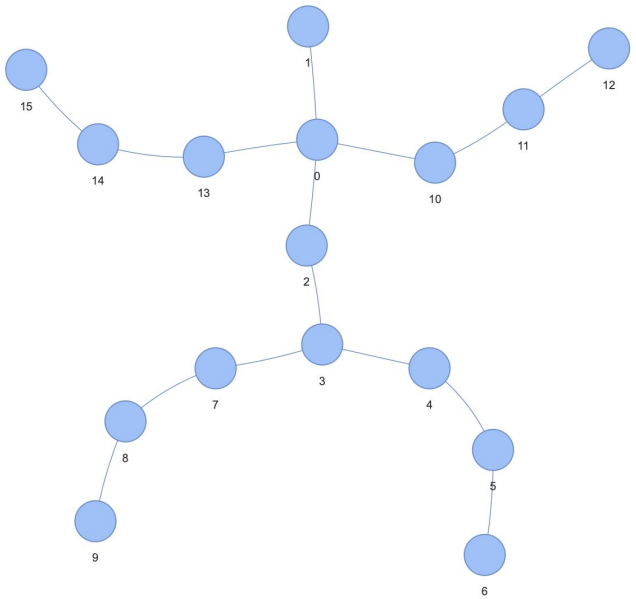
Внесение априорных знаний в RL агента при задаче локомоции



$$\begin{pmatrix}
0 & 1 & 1 & 2 & 3 & 4 & 5 & 3 & 4 & 5 & 1 & 2 & 3 & 1 & 2 & 3 \\
1 & 0 & 2 & 3 & 4 & 5 & 6 & 4 & 5 & 6 & 2 & 3 & 4 & 2 & 3 & 4 \\
1 & 2 & 0 & 1 & 2 & 3 & 4 & 2 & 3 & 4 & 2 & 3 & 4 & 2 & 3 & 4 \\
2 & 3 & 1 & 0 & 1 & 2 & 3 & 1 & 2 & 3 & 3 & 4 & 5 & 3 & 4 & 5 \\
3 & 4 & 2 & 1 & 0 & 1 & 2 & 2 & 3 & 4 & 4 & 5 & 6 & 4 & 5 & 6 \\
4 & 5 & 3 & 2 & 1 & 0 & 1 & 3 & 4 & 5 & 5 & 6 & 7 & 5 & 6 & 7 \\
5 & 6 & 4 & 3 & 2 & 1 & 0 & 4 & 5 & 6 & 6 & 7 & 8 & 6 & 7 & 8 \\
3 & 4 & 2 & 1 & 2 & 3 & 4 & 0 & 1 & 2 & 4 & 5 & 6 & 4 & 5 & 6 \\
4 & 5 & 3 & 2 & 3 & 4 & 5 & 1 & 0 & 1 & 5 & 6 & 7 & 5 & 6 & 7 \\
5 & 6 & 4 & 3 & 4 & 5 & 6 & 2 & 1 & 0 & 6 & 7 & 8 & 6 & 7 & 8 \\
1 & 2 & 2 & 3 & 4 & 5 & 6 & 4 & 5 & 6 & 0 & 1 & 2 & 2 & 3 & 4 \\
2 & 3 & 3 & 4 & 5 & 6 & 7 & 5 & 6 & 7 & 1 & 0 & 1 & 3 & 4 & 5 \\
3 & 4 & 4 & 5 & 6 & 7 & 8 & 6 & 7 & 8 & 2 & 1 & 0 & 4 & 5 & 6 \\
1 & 2 & 2 & 3 & 4 & 5 & 6 & 4 & 5 & 6 & 2 & 3 & 4 & 0 & 1 & 2 \\
2 & 3 & 3 & 4 & 5 & 6 & 7 & 5 & 6 & 7 & 3 & 4 & 5 & 1 & 0 & 1 \\
3 & 4 & 4 & 5 & 6 & 7 & 8 & 6 & 7 & 8 & 4 & 5 & 6 & 2 & 1 & 0
\end{pmatrix}$$

Постановка задачи

Внесение априорных знаний в RL агента при задаче локомоции



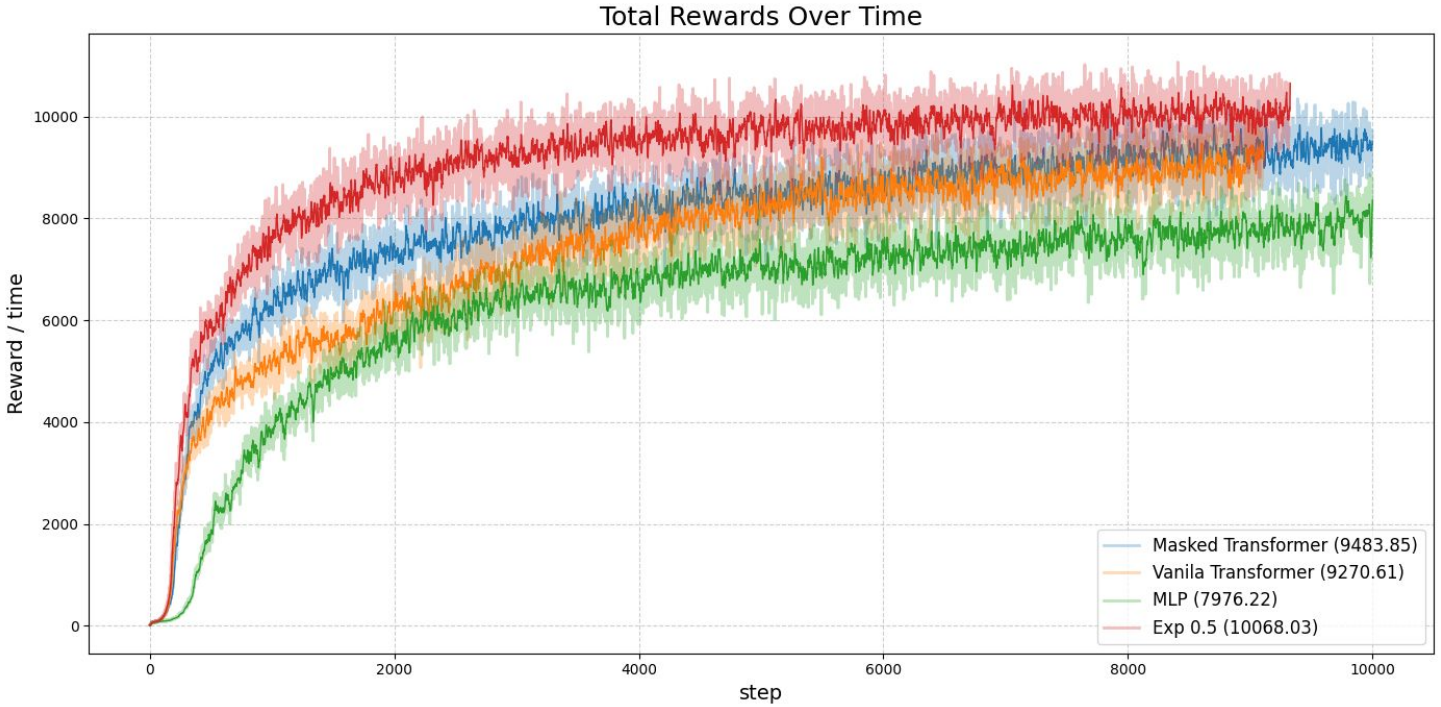
$$\begin{bmatrix}
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1
\end{bmatrix}$$

Положения, выносимые на защиту

- Показано, что априорная инъекция физической информации о кинематической структуре робота существенно повышает эффективность обучения политики локомоции методом PPO в Isaac Gym.
- Проведено систематическое сравнение трёх способов структурного маскирования механизма внимания (бинарная маска, экспоненциальная маска, TDA-маска), продемонстрировав их влияние на эпизодическую награду и скорость сходимости.
- Исследован топологический подход к построению масок внимания для шарнирных конструкций, выявлен выигрыш в плотности связей при сохранении качества управления.

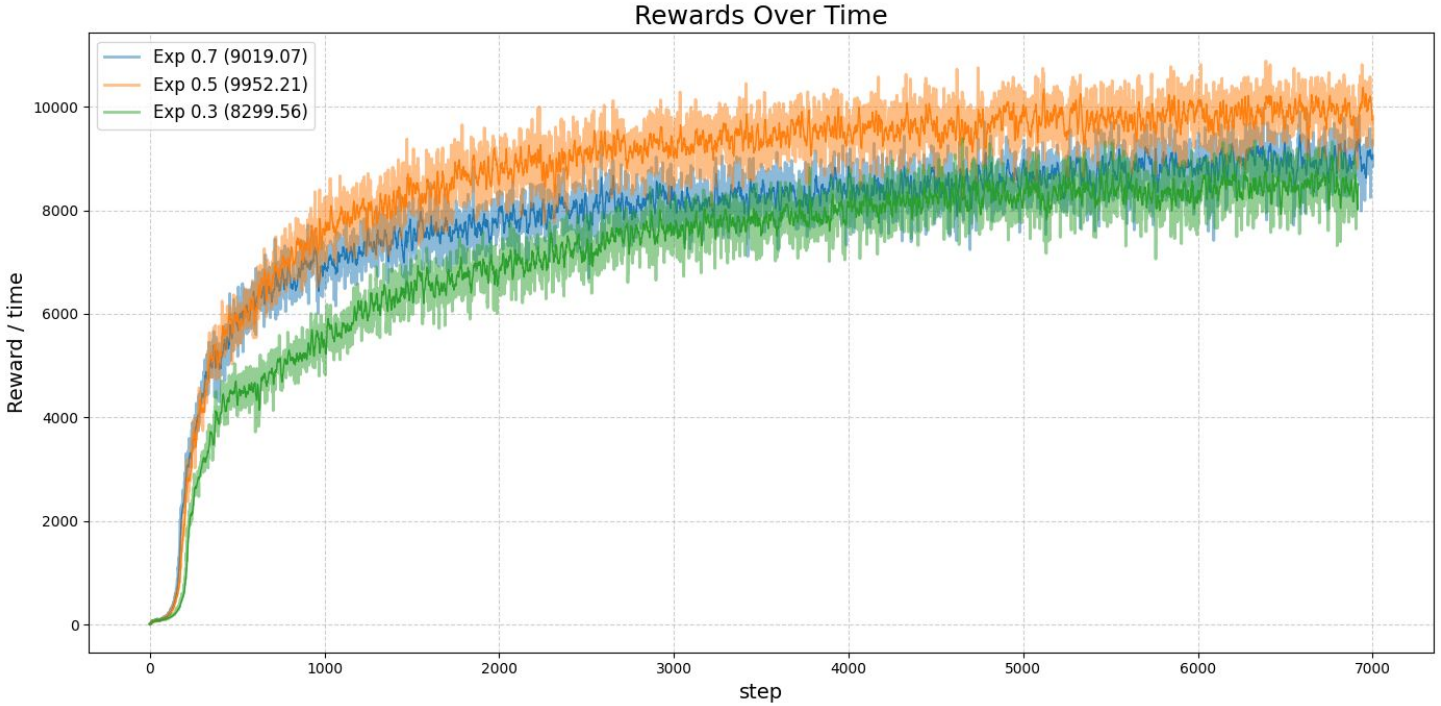
Результаты исследования

Сравнение различных типов маскирования



Результаты исследования

Сравнение различных типов экспоненциального маскирования



Результаты исследования

Pipeline анализа карт внимания

1

Оучение base-line transformer политики на задачу

2

Тестирование и сохранение карт внимания

3

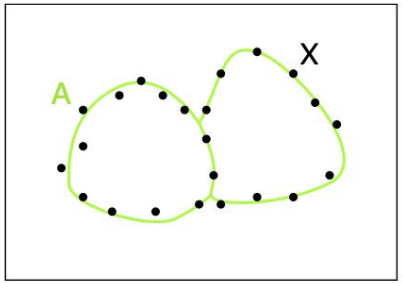
Анализ получившихся карт внимания. Определение вапжных ребер

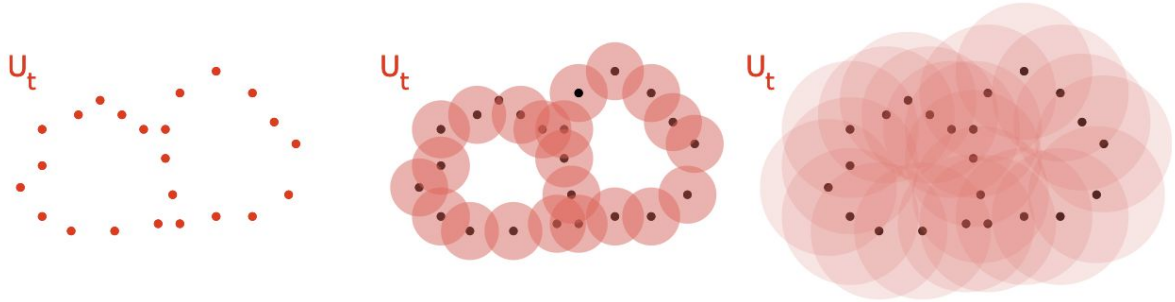
4

Включение новых ребер в карту внимание. Проведение эксперимента

Результаты исследования

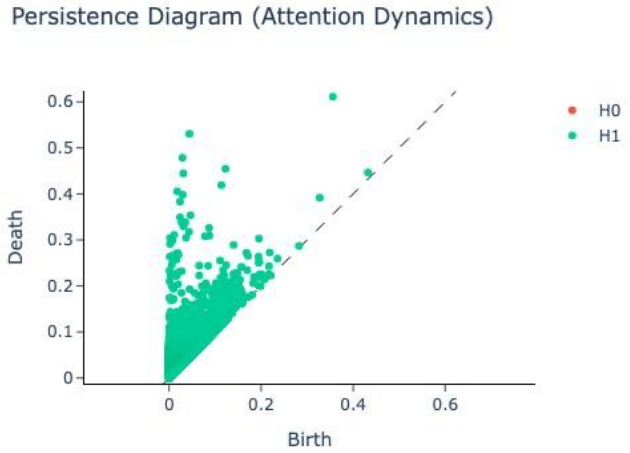
Анализ свойств системы

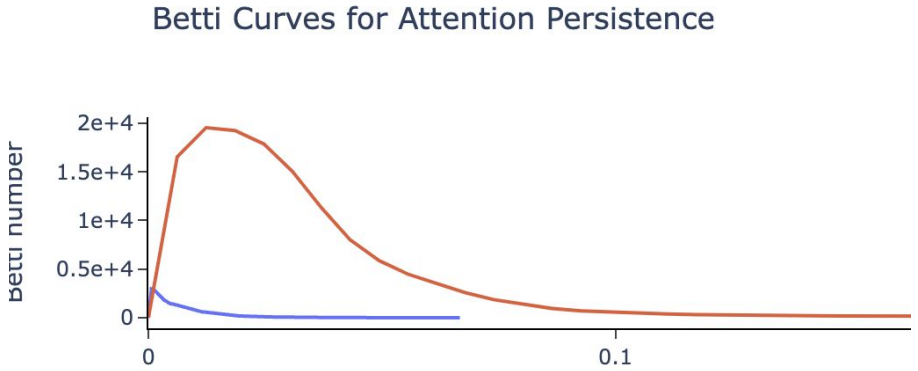




Результаты исследования

Анализ свойств системы

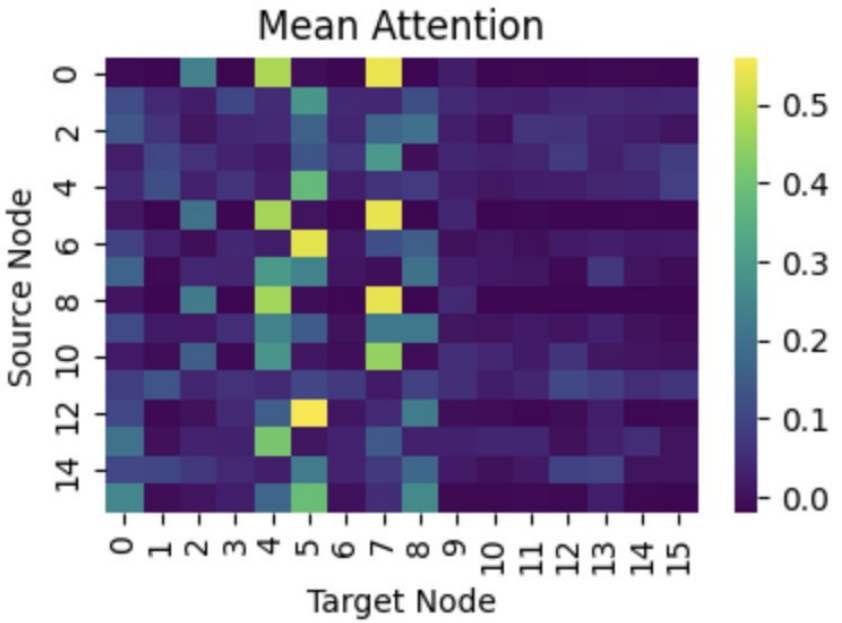




Результаты исследования

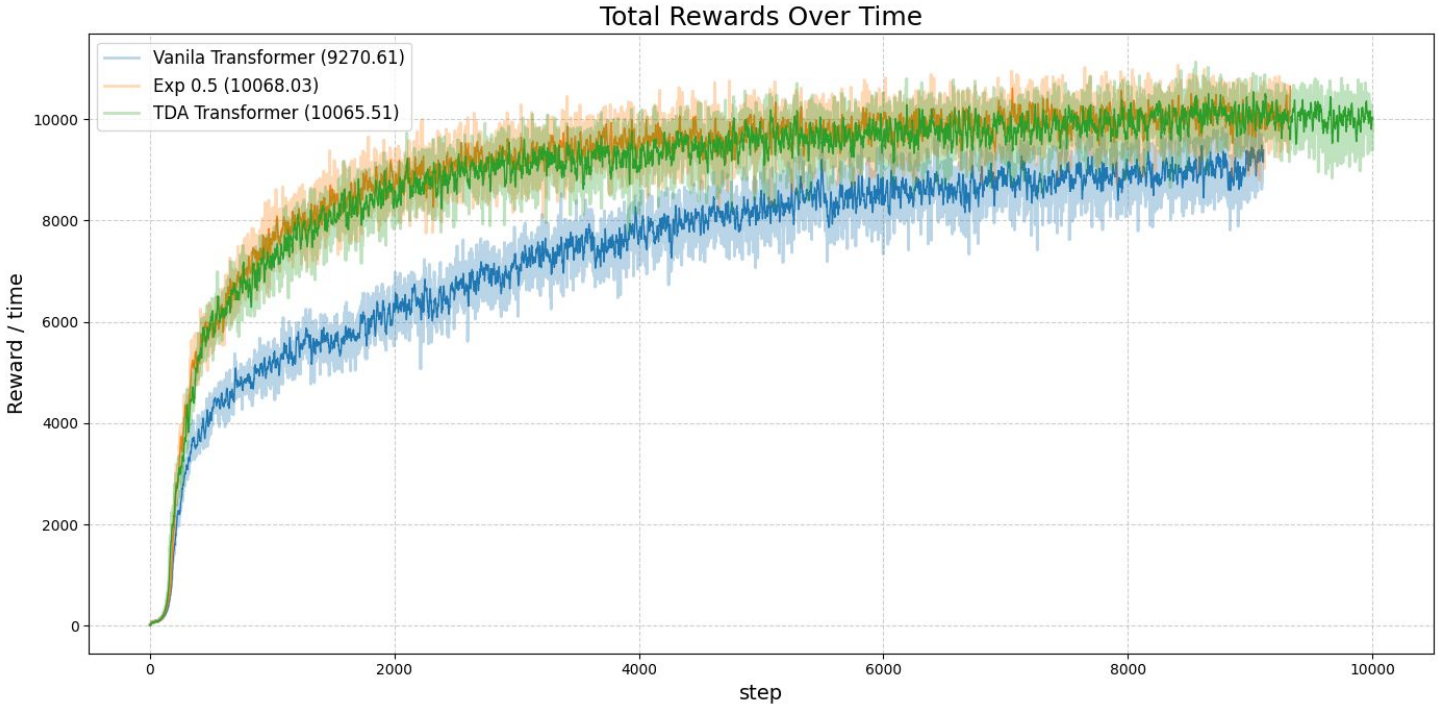
Анализ свойств системы

Топ вершин по в.с.	Суммарный вес
Вершина 7 (бедро)	1.786
Вершина 5 (колено)	1.730
Вершина 4 (бедро)	1.724
Вершина 8 (колено)	1.212
Вершина 0 (тело)	1.133



Результаты исследования

Анализ свойств системы



Выводы

- Учет априорных физических знаний через маскирование внимания — повышает качество и устойчивость обучения трансформер-политик в задаче локомоции.
- Бинарные маски внимания — обеспечивают заметный прирост эпизодической награды и ускоряют сходимость PPO по сравнению с немаскированным базовым трансформером.
- Экспоненциальное маскирование attention — даёт сопоставимый прирост качества, но замедляет обучение из-за отключения нативных оптимизаций attention-механизма.
- TDA-маски на основе persistent homology — сохраняют разреженность порядка 20 % (то есть лишь 20 % непустых связей) и обеспечивают качество управления на уровне экспоненциальных масок, что значительно сокращает вычислительные затраты.