Object-centric EfficientZeroV2

Вахитов Р.Р.
rodionvahitoff@yandex.ru

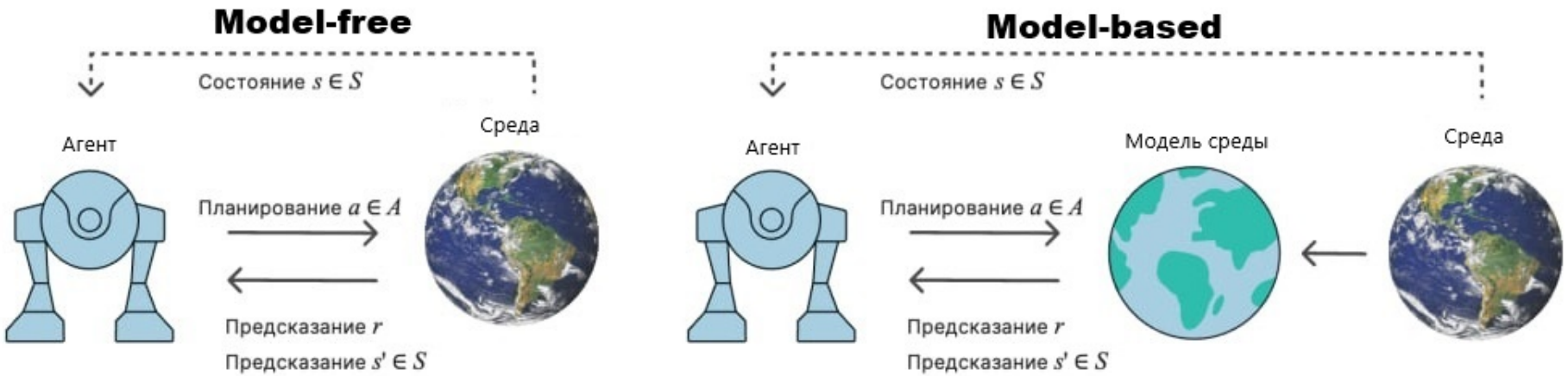
Центр когнитивного моделирования МФТИ

2025 – Научно-исследовательская работа





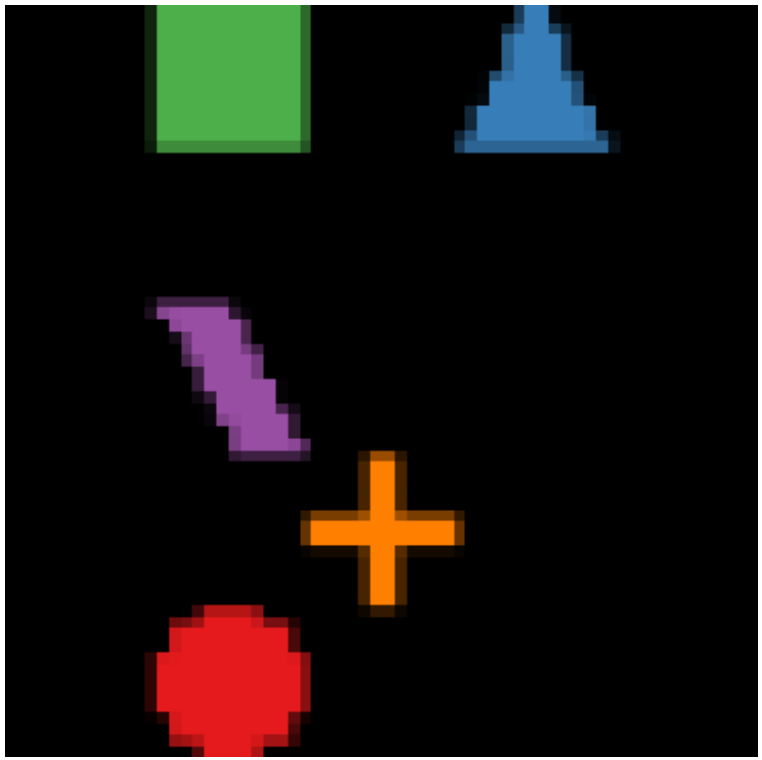
Мотивация

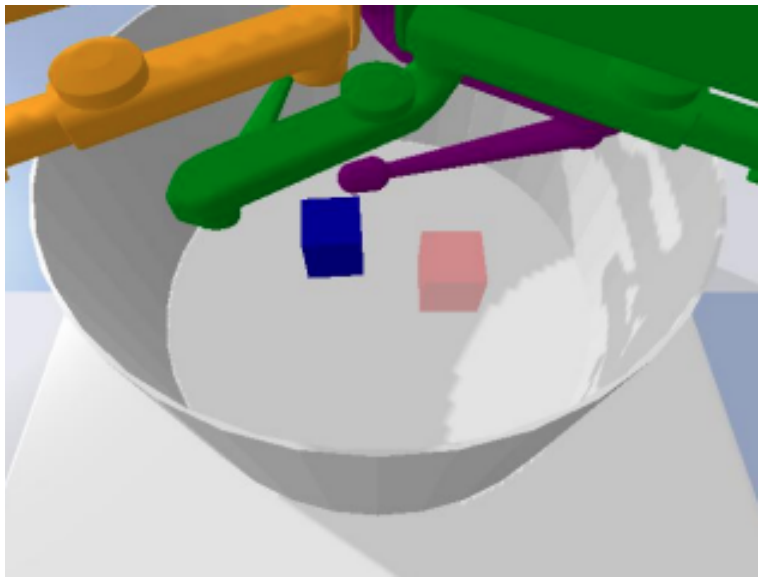


MBRL сталкивается с рядом недостатков и нерешённых проблем:

1. Слабая интерпретируемость нейросетевых моделей динамики
2. Нестабильность латентного пространства
3. Сложность переноса обученной модели на новые задачи
4. Низкая способность к обобщению

Цель исследования





Цель исследования — разработка алгоритма обучения с подкреплением на основе модели среды, который учитывает взаимодействия объектов в сложных средах для повышения эффективности обучения.

Постановка задачи

$(S, A, T, R, \gamma, \mathcal{O}, \Omega)$ - марковский процесс принятия решений, где:

- S: $S_1 \times \cdots \times S_k$: пространство состояний, где S_i - индивидуальное пространство состояний объекта i
- A: множество действий,
- $\mathcal{O}$: пространство наблюдений,
- $T : S \times A \to S$: функция переходов,
- $R : S \times A \to \mathbb{R}$: функция вознаграждения,
- $\Omega : \mathcal{S} \to \mathcal{O}$: функция наблюдений,
- γ: коэффициент дисконтирования.

Агент имеет доступ к обновляемой модели среды $M = \langle \hat{T}, \hat{R} \rangle$, которая задается графом, где вершины соответствуют объектным представлениям, а рёбра описывают возможные изменения свойств этих объектов при выполнении действий, и может строить предсказания траекторий с целью оптимизации стратегии выбора действий на основе получаемых данных:

$$\text{Traj} = \langle s_i, r_i, a_i, \hat{s}_{i+1}, \hat{r}_{i+1}, a_{i+1}, \ldots, a_{i+n}, \hat{s}_{i+n+1} \rangle$$

Цель агента — максимизировать ожидаемое кумулятивное вознаграждение, определяемое стратегией π:

$$\mathbb{E}_\pi \sum_{t=0}^{\tau} \gamma^t R(s_t, a_t)$$

Обзор работ

Center for Cognitive Modeling

- COBRA: Data-Efficient Model-Based RL through Unsupervised Object Discovery and Curiosity-Driven Exploration
- FOCUS: object-centric world models for robotic manipulation
- Objects matter: object-centric world models improve reinforcement learning in visually complex environments
- SOLD: Slot Object-Centric Latent Dynamics Models for Relational Manipulation Learning from Pixels

Задачи

- Получение baseline на средах Shapes2d, Maniskill, Robosuite и Causal World
  - Интеграция сред Shapes2d, Maniskill, Robosuite и Causal World в обучение агента
- Имплементация ОС версии EfficientZeroV2
- Проведение экспериментов на OCEZ-V2
- Технические доработки
- Доработка алгоритма, ablation study, дополнительные эксперименты
- Оформление статьи

EfficientZeroV2

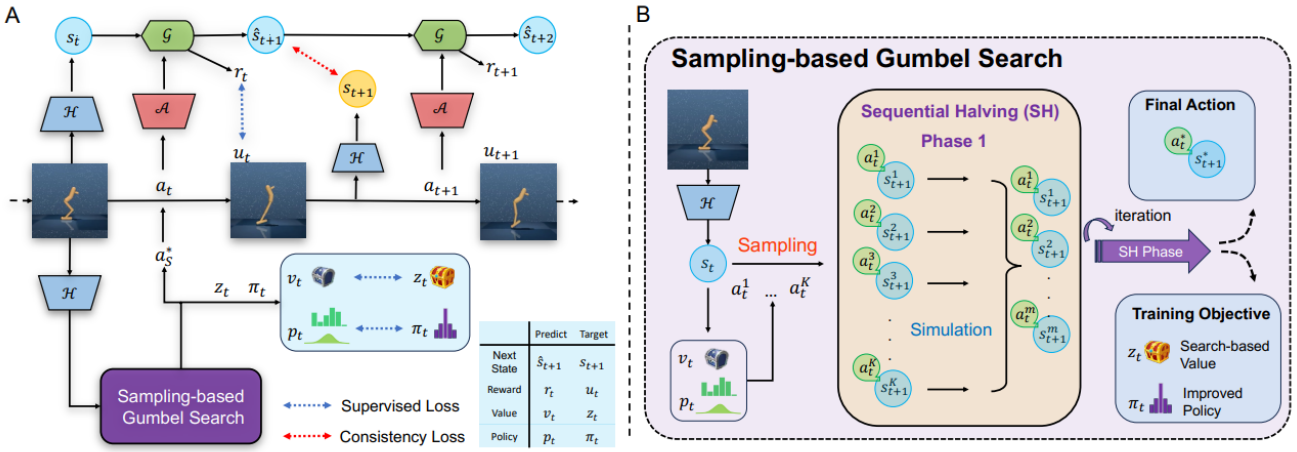


Рис.: Схема работы EZ-V2

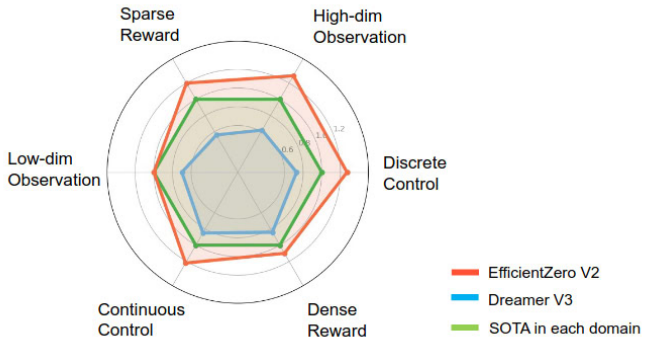


Рис.: Результаты EZ-V2

Object-centric EfficientZeroV2

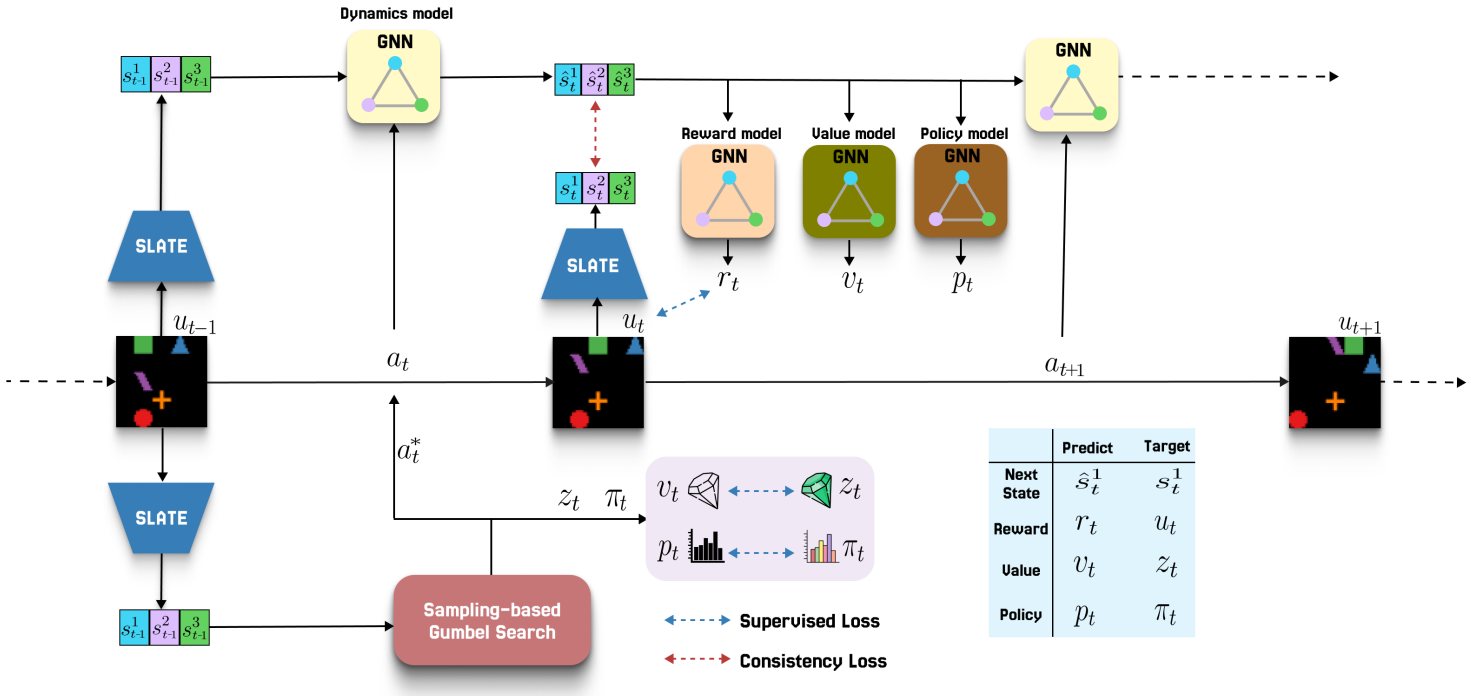


Рис.: Схема работы OCEZ-V2

Первые эксперименты

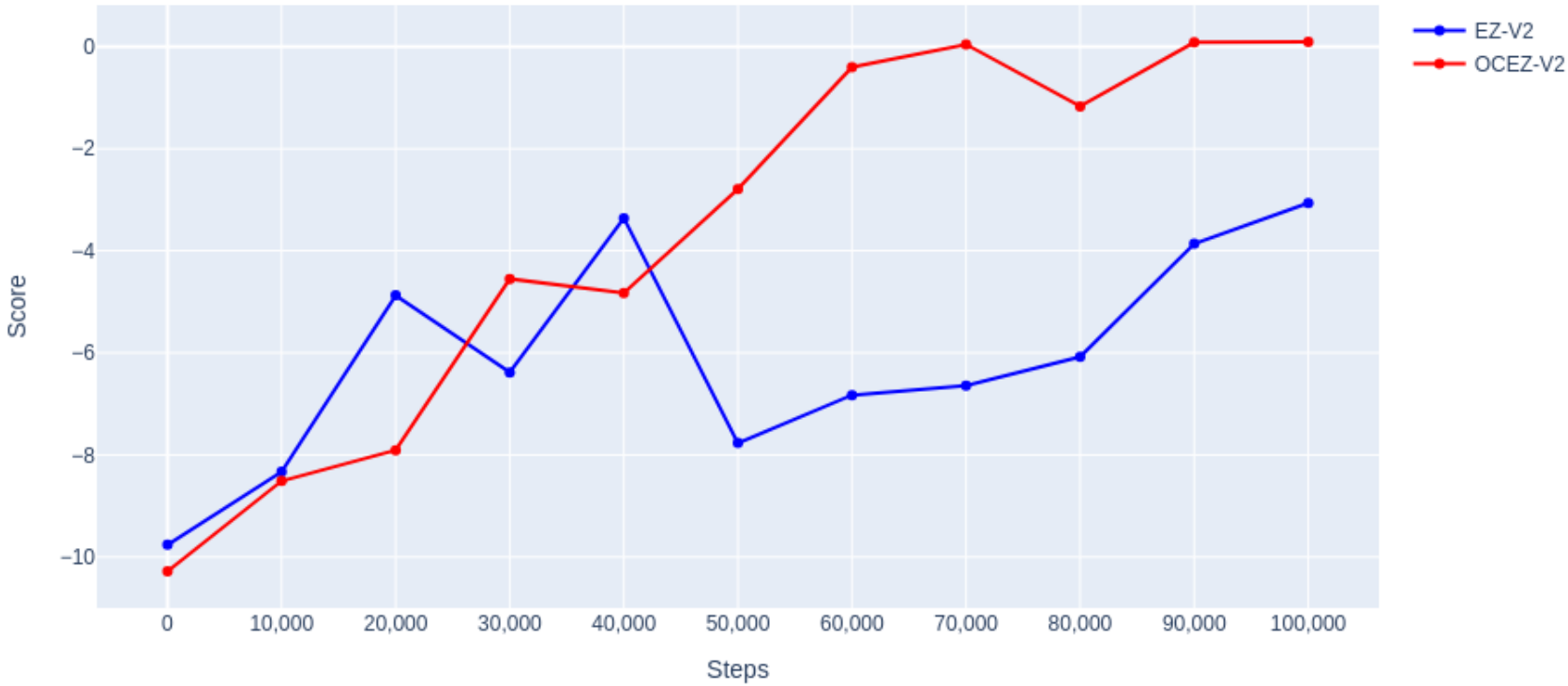


Рис.: Усредненная отдача по 100 эпизодам и трем seed для EZ-V2 и OCEZ-V2 на Shapes2D Navigation Task

Прогресс и дальнейшие планы исследований

Center for Cognitive Modeling

- Получение baseline на средах Shapes2d, Maniskill, Robosuite и Causal World
  - Интеграция сред Shapes2d, Maniskill, Robosuite и Causal World в обучение агента
- Имплементация ОС версии EfficientZeroV2
- Проведение экспериментов на OCEZ-V2 на представленных средах
- Технические доработки
- Доработка алгоритма, ablation study, дополнительные эксперименты
- Оформление статьи
- Модернизация MCTS в парадигме объектно-центричного подхода

Спасибо за внимание!

cogmodel.mipt.ru

rodionvahitoff@yandex.ru