УДК 004.4/.8

Мультиагентное обучение с подкреплением для адаптивного проектирования в иммерсивных VR средах

Морозов Кирилл Андреевич morozovka@student.bmstu.ru
SPIN-код: 5456-1403

МГТУ им. Н.Э. Баумана, Москва, Россия

Аннотация. *Работа посвящена применению мультиагентного обучения с подкреплением для адаптивного проектирования иммерсивных виртуальных сред. Предложена концепция, в которой задача формирования* VR *среды декомпозируется между несколькими кооперативными агентами, каждый из которых управляет отдельным слоем, в частности, структурой помещения, объектами, материалами и освещением. Взаимодействие агентов формализовано как децентрализованный частично наблюдаемый марковский процесс принятия решений с общей наградой. В качестве алгоритма обучения рассмотрен* MAPPO. *Для валидации концепции реализована дискретная среда с визуализацией всех слоев. Рассмотрены перспективы масштабирования подхода к трехмерным средам и интеграции с системами рендеринга.*

***Ключевые слова*:** *мультиагентное обучение с подкреплением, виртуальная реальность, иммерсивные среды, машинное обучение*

Multi-agent reinforcement learning for adaptive design in immersive VR environments

Morozov Kirill Andreevich morozovka@student.bmstu.ru
SPIN-code: 5456-1403

BMSTU, Moscow, Russia

Abstract. This paper explores the application of multi-agent reinforcement learning to the adaptive design of immersive virtual environments. A concept is proposed in which the task of creating a VR environment is decomposed among several cooperative agents, each of which controls a separate layer, specifically the room structure, objects, materials and lighting. Agent interaction is formalized as a decentralized, partially observable Markov decision process with a shared reward. MAPPO is considered as the learning algorithm. To validate the concept, a discrete environment with visualization of all layers is implemented. Prospects for scaling the approach to 3D environments and integration with rendering systems are discussed.

Keywords: multi-agent reinforcement learning, virtual reality, immersive environments, machine learning

Введение. Иммерсивные виртуальные среды (англ. VR) требуют высокой частоты кадров и низкой задержки для обеспечения ощущения погруженности, присутствия. Традиционные подходы часто используют настройки, ко-

торые могут не эффективно адаптироваться к изменяющейся сложности в рамках виртуального окружения [1]. Современные подходы к адаптивному рендерингу, как правило, централизованы и не учитывают локальные взаимодействия между объектами сцены [2].

Мультиагентное обучение с подкреплением (англ. MARL) предлагает перспективный подход для децентрализованного управления, позволяющего распределить задачу оптимизации между множеством агентов, каждый из которых управляет конкретным элементом среды [3–5]. Ключевая идея данной работы заключается в формировании концепции позволяющей, реализовав, использовать иммерсивные кооперативные среды, где агенты не конкурируют, а сотрудничают для достижения более глобальной цели, а именно баланса между производительностью и качеством.

Материалы и методы. Формирование иммерсивных VR сред возможно формализовать как кооперативный децентрализованный частично наблюдаемый марковский процесс принятия решений (англ. Dec-POMDP) [6]. В таком случае, среду допустимо представлять как дискретную сетку из заранее определенного количества ячеек, где внешний периметр формирует стены помещения, то есть самой виртуальной среды. Внутренняя область ячеек является рабочей частью для агентов, непосредственно их средой. В рассматриваемой ситуации каждая ячейка содержит несколько слоев информации, в частности, структура, тип объекта, материал и освещение. В разрабатываемой системе действуют четыре специализированных агента, каждый из которых отвечает за отдельный уровень проектирования дизайна VR среды.

Агенты, действуя последовательно в фиксированном порядке. Сначала управление структурой помещения, затем объектами, после материалами и текстурами поверхности, в завершении уже источникам света (их тип и расположение) в рамках одного шага (каждый из агентов таким образом отвечает за один из таких слоев). Каждый следующий агент наблюдает результаты действий предшествующих агентов, что формирует неявный канал координации. Один эпизод состоит из нескольких шагов, что позволяет агентам итеративно улучшать VR среду. Каждый агент видит свой собственный слой, которым он управляет, в полном объеме (точные значения в каждой ячейке), но при этом о других слоях ему известно только расположение элементов, но не какие именно. Эта информационная асимметрия создает нетривиальную задачу координации и стимулирует агентов к адаптивному поведению.

Все агенты получают единую разделяемую награду после каждого шага, что обеспечивает полностью кооперативный режим. Награда вычисляется как взвешенная сумма трех компонент. Так, первая компонента накладывает штраф за коллизии объектов и избыточную плотность заполнения, поощряя согласованность расположения объектов. Вторая компонента отвечает за разнообразие используемых материалов и равномерность освещения по площади помещения. Третья оценивает функциональную правдоподобность заданного расположения объектов, например, размещение полок вдоль стен или нали-

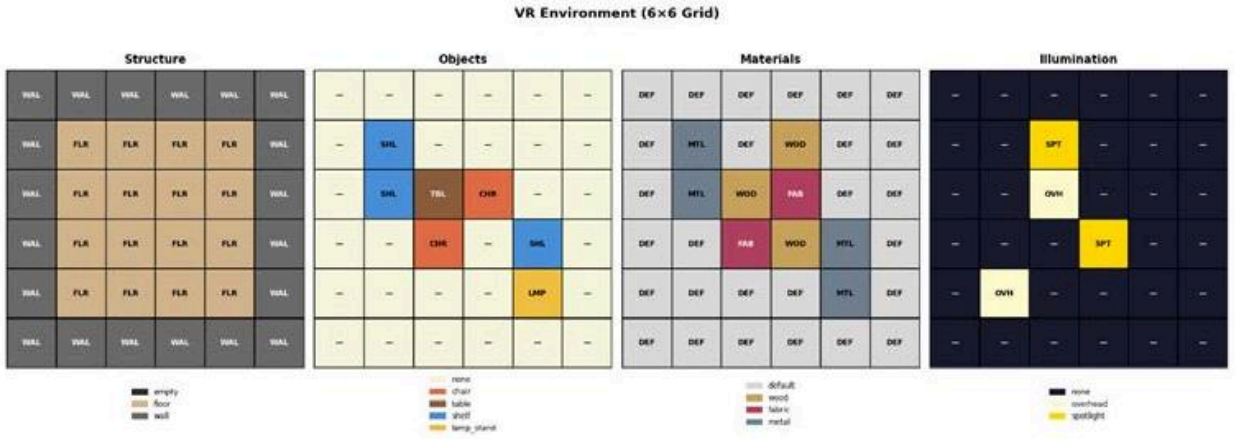
чие минимальных объектов характерных для данного типа помещения. Определение кумулятивной награды за эпизод можно представить следующим образом:

$$R_{ep} = \sum_{t=1}^{T_{max}} \gamma^{t-1} R_t \,, \tag{1}$$

где t — текущий шаг, T_{max} — общее число шагов в эпизоде, R_t — награда по сумме трех компонент (с учетом их значимости), γ — коэффициент дисконтирования.

В качестве подходящего алгоритма обучения допустимо использовать MAPPO (англ. Multi-Agent Proximal Policy Optimization) — мультиагентное расширение PPO в формате централизованного обучения с децентрализованным исполнением. Обоснованием такого выбора для данной задачи обусловлено доступной поддержкой дискретных пространства действий [7, 8].

Результаты. Для выполнения валидации и анализа предложенной концепции была реализована среда в виде дискретной сетки 6×6 ячеек с четырьмя слоями. На рисунке представлен пример визуализации такой среды, где каждый слой отображается отдельно. То есть структура помещения (стены и пол), размещенные объекты, назначенные материалы и источники освещения. Данная визуализация демонстрирует исходное представление пространства, с которым взаимодействуют агенты.



Визуализация слоев среды

Одним из критериев успешности сформированных сред является отсутствие вырождения, то есть в подавляющем большинстве генераций должно присутствовать не менее двух различных типов объектов и материалов, что свидетельствует об отсутствии коллапса политик в тривиальных решениях. Перевод подобного подхода в систему генерации уже более усложненных VR сред, включая полноценные 3D-среды, приведет к необходимости значительного расширения типов слоев, их содержимого, а также обновление кумулятивной награды, однако даже в рамках анализа данного примера можно судить о перспективности данного подхода. При этом следует отметить тот

факт, что обучение в мире-сетке не требует сильных вычислительных возможностей, что позволяет судить о пригодности для динамичного обновления, однако при движении в сторону ощущения погружения, потребуются значительно бо̀льшие мощности, в первую очередь для обучения агентов [9].

Обсуждение. Предложенная концепция мультиагентного обучения для проектирования VR сред обладает рядом свойств, определяющих направления дальнейшего развития и практического применения [10]. Так, модульная архитектура системы, в которой каждый агент отвечает за отдельный слой, открывает возможность независимой модификации отдельных слоев дизайна. То есть для смены визуального стиля помещения достаточно переобучить только агента материалов, сохранив политики остальных агентов без изменений. Аналогично допустимо расширение системы новыми слоями с добавлением агентов, при фиксации существующих. Подобная гибкость отсутствует в статичных моделях, где любое изменение затрагивает всю систему генерации сред целиком.

Как уже ранее было отмечено, переход от дискретной двумерной сетки к трехмерным VR средам предполагает множество технических расширений. С увеличения размерности сетки и объема словарей до полноценной интеграции с системами рендеринга, позволяющими визуализировать сгенерированные сетки в виде трехмерных сцен. При этом следует учитывать и необходимость поддержки непрерывного позиционирования объектов, их произвольная ориентации потребует перехода к непрерывным пространствам действий, что, в свою очередь, может потребовать адаптации алгоритма обучения.

Заключение. В данной работе представлена концепция применения мультиагентного обучения с подкреплением для проектирования иммерсивных VR сред. Задача декомпозирована между четырьмя агентами, отвечающими за структуру, объекты, материалы и освещение, которые действуют последовательно и координируются через общую разделяемую награду. Формализация задачи выполнена в рамках Dec-POMDP с частичной наблюдаемостью, а в качестве алгоритма обучения предложен MAPPO. Представлена среда в виде дискретной сетки 6×6, демонстрирующая работоспособность подхода для потенциального расширения на большую структуру. Модульность архитектуры обеспечивает возможность независимой модификации отдельных слоев дизайна и допускает расширение новыми без переобучения всей системы. Дальнейшие исследования и разработки предполагают переход к трехмерным средам, интеграцию с элементами 3D рендеринга и применение предварительного обучения на экспертных данных.

Список источников

[1] Lili Wang, Xuehuai Shi, Yi Liu. Foveated Rendering: a State-of-the-Art Survey. 2022, 32 p. URL: https://arxiv.org/abs/2211.07969 (accessed 12.01.2026).

[2] Sirohi P., Agarwal A., Maheshwari P. A survey on Augmented Virtual Reality: Applications and Future Directions. Seventh International Conference on Information Technology Trends

(ITT), Abu Dhabi, United Arab Emirates, 2020, pp. 99–106. https://doi.org/10.1109/ITT51279.2020.9320869

[3] Morozov K.A. Models as a Key Factor of Environments Design in Multi-Agent Reinforcement Learning. 6th International Youth Conference on Radio Electronics, Electrical and Power Engineering (REEPE), Moscow, Russian Federation, 2024, pp. 1–5. https://doi.org/10.1109/REEPE60449.2024.10479882

[4] Velichko N.A. Distributed Multi-Agent Reinforcement Learning Based on Feudal Networks. 6th International Youth Conference on Radio Electronics, Electrical and Power Engineering (REEPE), Moscow, Russian Federation, 2024, pp. 1–5. https://doi.org/10.1109/REEPE60449.2024.10479775

[5] Sakulin S., Alfimtsev A. Multicriteria decision making in tourismindustry based on visualization of aggregation operators. Applied System Innovation, 2023, vol. 6, no. 5, art. no. 74. https://doi.org/10.3390/asi6050074 (accessed 12.01.2026).

[6] Bernstein D.S., Shlomo Zilberstein, Neil Immerman. The Complexity of Decentralized Control of Markov Decision Processes. Uncertainty in Artificial Intelligence Proceedings, 2000, pp. 32–37. URL: https://arxiv.org/abs/1301.3836 (accessed 12.01.2026).

[7] Schulman J., Wolski F., Dhariwal P., Radford A., Klimov O. Proximal Policy Optimization Algorithms. 2017, 12 p. URL: https://arxiv.org/abs/1707.06347 (accessed 12.01.2026).

[8] Chao Yu, Akash Velu, Vinitsky E., Jiaxuan Gao, Yu Wang, Bayen A., Yi Wu. The Surprising Effectiveness of PPO in Cooperative, Multi-Agent Games. 36th Conference on Neural Information Processing Systems (NeurIPS 2022) Track on Datasets and Benchmarks, 2022, 30 p. URL: https://arxiv.org/abs/2103.01955 (accessed 12.01.2026).

[9] Hady, M.A., Hu, S., Pratama, M. et al. Multi-agent reinforcement learning for resources allocation optimization: a survey. Artif Intell Rev., 2025, vol. 58, art. no. 354, 49 p. https://doi.org/10.1007/s10462-025-11340-5

[10] Oyedokun O., Alkahtani M., Duffy V.G. Exploration of User Experience in Virtual Reality Environment. A Systematic Review. Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management. HCII 2024. Lecture Notes in Computer Science, Springer, Cham, 2024, vol. 14709, pp. 320–338. https://doi.org/10.1007/978-3-031-61060-8_23