

ПРИМЕНЕНИЕ ГРАФОВ В МУЛЬТИАГЕНТНОМ ОБУЧЕНИИ С ПОДКРЕПЛЕНИЕМ: ПРОГРЕСС ИССЛЕДОВАНИЙ И МЕТОДОЛОГИЧЕСКАЯ ОСНОВА

APPLICATIONS OF GRAPHS IN MULTI-AGENT REINFORCEMENT LEARNING: RESEARCH PROGRESS AND METHODOLOGICAL FRAMEWORK

**Zhang Che
V. Terekhov
G. Afanasyev**

Summary. Graph-based modeling is one of the key paradigms in multi-agent reinforcement learning (MARL), offering innovative solutions to complex problems in collaborative decision making. This paper explores the application of graph theory in MARL by providing a holistic methodological framework in three key areas: modeling of graph-structured environments, coordinated dynamic graph optimization, and graph-driven communication mechanisms. Based on this, a detailed comparison and analysis of state-of-the-art graph-based modeling methods in multi-agent reinforcement learning is provided. Case studies in traffic signal control and autonomous vehicle coordination demonstrate the versatility of the graph model in real-world applications, including intelligent transportation systems. However, significant theoretical challenges remain, such as ensuring convergence in graph-based dynamic communication and efficiently processing heterogeneous graph data.

Keywords: multi-agent reinforcement learning, multi-agent system, graph neural networks, coordination graph, dynamic graph.

Чжан Чэ

Аспирант, Московский государственный
технический университет им. Н.Э. Баумана
(национальный исследовательский университет)
chezhangcn@gmail.com

Терехов Валерий Игоревич

К.т.н., доцент. Московский государственный
технический университет им. Н.Э. Баумана
(национальный исследовательский университет)
terekchow@bmstu.ru

Афанасьев Геннадий Иванович

К.т.н., доцент, Московский государственный
технический университет им. Н.Э. Баумана
(национальный исследовательский университет)
gaipcs@bmstu.ru

Аннотация. Моделирование на основе графов является одной из ключевых парадигм в мультиагентном обучении с подкреплением (MARL), предлагая инновационные решения сложных задач в области совместного принятия решений. В данной работе проводится исследование применения теории графов в MARL, создавая целостную методологическую основу по трём ключевым направлениям: моделирование среды с графовой структурой, координированная динамическая оптимизация графов и механизмы коммуникации, управляемые графами. На основе этого проводится детальное сравнение и анализ современных методов моделирования на основе теории графов в мультиагентном обучении с подкреплением. Примеры из практики в области управления транспортными сигналами, координации автономных транспортных средств демонстрируют универсальность графовой модели в реальных приложениях, включая интеллектуальные транспортные системы. Однако остаются значительные теоретические проблемы, такие как обеспечение сходимости при динамической коммуникации на основе графов и эффективная обработка неоднородных графовых данных.

Ключевые слова: мультиагентное обучение с подкреплением, мультиагентная система, графовые нейронные сети, координационный граф, динамический граф.

Введение

В последние годы мультиагентное обучение с подкреплением (MARL) показало значительный потенциал в решении сложных динамических взаимосвязанных задач. Основные трудности здесь связаны с эффективным моделированием среды, координацией поведения агентов и оптимизацией механизмов коммуникации. Традиционные методы часто сталкиваются с такими проблемами, как динамика среды, фрагментация информации вследствие локальных наблюдений,

«проклятие размерности» из-за взрыва в пространстве совместных действий [11] и чрезмерные затраты на коммуникацию. Графы как инструменты структурированного представления [2] предлагают MARL новую методологию благодаря их возможностям топологического моделирования, характеристикам реляционного рассуждения и вычислительной эффективности.

Суть данного аналитического обзора заключается в обосновании подхода классификации «моделирование–координация–коммуникация», основанного

на функциональных характеристиках графов. Рассматриваемый в этой статье подход классификации «моделирование-координация-коммуникация» анализирует различные парадигмы применения графов в мультиагентном обучении с подкреплением с функциональной точки зрения.

Теоретическая основа

1. Частично наблюдаемый марковский процесс принятия решений

Основа многоагентного обучения с подкреплением заключается в его математической формулировке, которая расширяет рамки одноагентных систем для решения задач распределённого принятия решений в частично наблюдаемых средах. Канонической моделью для MARL служит частично наблюдаемый марковский процесс принятия решений (POMDP) [1], определяемая кортежем $N, S, \{A^i\}, \{O^i\}, T, R, \gamma$:

N : множество агентов.

S : глобальное пространство состояний, часто недоступное отдельным агентам.

A^i : пространство действий агента i .

O^i : пространство наблюдений агента i , генерируемое функцией $\mathcal{E}^i: S \rightarrow O^i$.

$T(s'|s, a)$: вероятность перехода состояния для совместного действия $a = (a_1, \dots, a_N)$.

$R^i(s, a)$: индивидуальная функция вознаграждения (в кооперативных задачах часто предполагается общее вознаграждение).

γ : коэффициент дисконтирования вознаграждения.

Каждый агент i получает локальное наблюдение $o_t^i = \mathcal{E}^i(s_t)$ в момент времени t , формируя историю $h_t^i = (o_0^i, a_0^i, \dots, o_t^i)$. Децентрализованная стратегия $\pi^i(a_t^i | h_t^i)$ отображает истории в распределения действий.

Совместная стратегия $\pi = (\pi^1, \dots, \pi^N)$ порождает траекторию $\tau = (s_0, o_0, a_0, s_1, \dots)$, с целью максимизации ожидаемого возврата $\mathbb{E}_\tau \left[\sum_{t=0}^{\infty} \gamma^t \mathcal{R}(s_t, a_t) \right]$.

В частично наблюдаемых марковских играх (POMG) процесс обучения усложняется несколькими ключевыми проблемами. Во-первых, возникает нестационар-

ность, поскольку развивающиеся стратегии агентов постоянно изменяют динамику среды, дестабилизируя обучение и препятствуя сходимости к оптимальным стратегиям. Во-вторых, сложно распределить кредиты из-за взаимозависимостей между агентами. В-третьих, алиасинг наблюдений вводит неоднозначность, так как различные состояния могут порождать идентичные наблюдения. Эти проблемы создают препятствия для разработки эффективных алгоритмов обучения для многоагентных систем в условиях частичной наблюдаемости.

2. Фреймворк централизованного обучения и децентрализованного исполнения

Фреймворк централизованного обучения с децентрализованным выполнением (CTDE) решает проблему частичной наблюдаемости, используя глобальную информацию (например, полное состояние s_t) во время обучения и ограничивая политики локальными историями во время выполнения. На рисунке 1 показана рамочная структура CTDE. Методы, такие как VDN [24] и QMIX [20], раскладывают совместную функцию полезности $Q_{joint}(s, a)$ на индивидуальные полезности $Q^i(h^i, a^i)$, обеспечивая соответствие принципу «Individual-Global-Max»:

$$\arg\max_a Q_{joint}(s, a) = \left(\arg\max_{a^1} Q^1(h^1, a^1), \dots, \arg\max_{a^N} Q^N(h^N, a^N) \right). \quad (1)$$

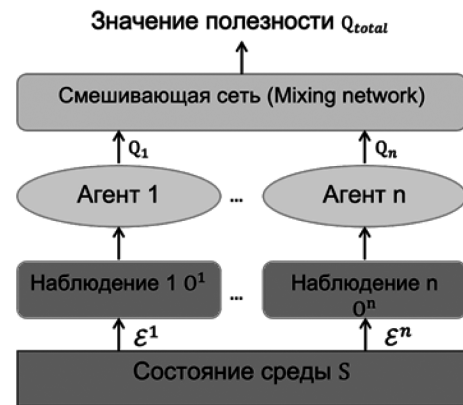


Рис. 1. Рамочная структура с централизованным обучением и децентрализованным исполнением

3. Роль графовых структур

Графовое мультиагентное обучение с подкреплением вводит реляционные индуктивные смещения [3] для смягчения проблемы частичной наблюдаемости. Моделируя взаимодействия агентов как рёбра в графе $G = (V, E)$, где узлы $v_i \in V$ представляют агентов или сущности, а рёбра $e_{ij} \in E$ кодируют отношения (например, коммуникационные связи или веса зависимостей), агенты могут распространять информацию с помощью

графовых нейронных сетей. Это позволяет агентам выводить скрытые состояния из локальных наблюдений и сообщений соседей, эффективно расширяя их «перцептивное поле» при сохранении децентрализации.

Такая формализация подчёркивает, как графовое мультиагентное обучение с подкреплением устраняет разрыв между децентрализованным выполнением и структурированным кооперативным рассуждением в условиях частичной наблюдаемости.

Методологии на основе графов

Интеграция графовых структур в мультиагентное обучение с подкреплением стала ключевой парадигмой для решения сложных задач координации. Подходы к моделированию среды на основе графов (раздел 3.1) используют обучение представлению графов для кодирования взаимодействия между агентами и средой. Методы на основе координационных графов (раздел 3.2) раскладывают глобальные задачи на коллаборативные единицы, повышая масштабируемость обучения политике. Техники оптимизации коммуникации с учётом графов (раздел 3.3) разрабатывают протоколы для балансировки эффективности обмена информацией и совместной производительности.

На рисунке 2 показаны различные роли теории графов в трех типах методологий. В совокупности эти методологии создают унифицированный фреймворк, который использует теорию графов для решения основных задач MARL, способствуя созданию интерпретируемых, масштабируемых и эффективных с точки зрения комму-

никации решений для многоагентных систем в реальных условиях.

1. Методы моделирования среды на основе графов

Первая категория методов предполагает моделирование среды с использованием графовых структур. В частности, в этих методах агенты, ресурсы или состояния в среде представлены в виде узлов графа, а их взаимосвязи описываются с помощью рёбер для построения эффективной модели среды. На этой основе подход обычно использует графовые нейронные сети (GNN) или их варианты для извлечения высокоуровневых реляционных признаков, тем самым расширяя возможности обучения агентов.

MAGNet [17] — это подход к многоагентному обучению с подкреплением, использующий графовые структуры для моделирования среды. На этапе генерации графа система моделирует агентов и объекты среды в виде узлов графа, строит динамическую матрицу корреляций и для принятия решений метод интегрирует архитектуру глубокой Q-сети (DQN), что позволяет использовать сгенерированный граф топологии среды для обучения политике и эффективного распределения совместных задач в многоагентных системах.

ING-MA [26] предлагает использовать разнородный граф транспортной сети (TN-HetG) с четырьмя типами узлов: контроллеров транспортных сигналов, перекрёстков, полос движения и транспортных средств. Нейронные сети кодируют разнородные признаки узлов и агрегируют информацию из сгруппированных окрест-

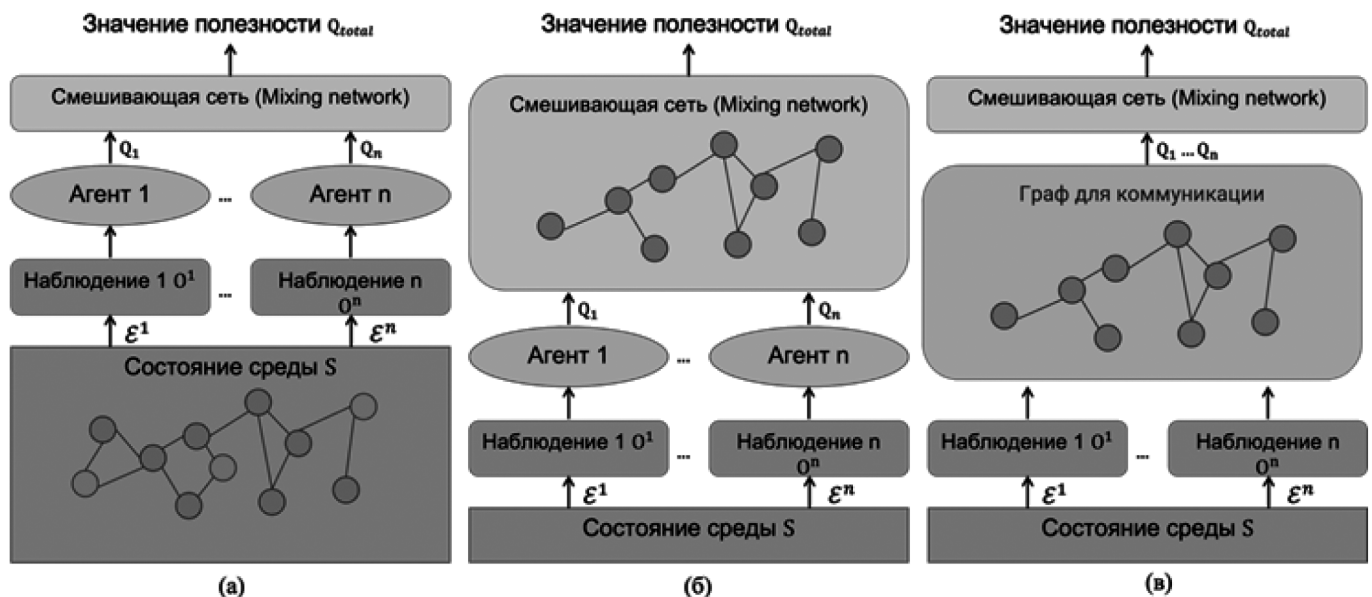


Рис. 2. 3 типа графовых методологий. (а) — использование графов для моделирования окружающей среды; (б) — метод координационного графа; (в) — использование графов для моделирования коммуникационных взаимодействий

ностей. Другая сеть, интегрированная с механизмами самовнимания, объединяет признаки узла и окрестностей для создания глобального представления состояния. На основе этого представления алгоритм использует фреймворк CTDE для обучения функции полезности или стратегии для каждого узла.

MAG-DQN [16] сочетает построение иерархической графовой структуры с графовым сверточным обучением с подкреплением для моделирования сложных сред и оптимизации политики. На этапе построения иерархической графовой структуры используется улучшенная нейронная сеть на основе теории адаптивного резонанса (ART) [6] для иерархической кластеризации карты.

2. Координационный граф

Координационные графы — основной инструмент для моделирования локального сотрудничества в мультиагентном обучении с подкреплением (MARL). Они определяют отношения взаимодействия между агентами с помощью графовых структур, решая проблему экспоненциального роста пространства совместных действий [13]. Данный метод объединил факторизацию из теории графов с динамическим программированием, снизив вычислительную сложность.

Глубокие координационные графы (DCG) [4] используют глубокие нейронные сети для автоматического изучения структуры и параметров координационного графа. DICG [14] объединяет механизмы самовнимания с графовыми свёрточными сетями (GCN), динамически выводятся неявные координационные графы с помощью дифференцируемых весов внимания.

Метод GCS [21] улучшает подход DCG, внедряя иерархический механизм топологической сортировки и динамически строя ориентированные ациклические графы (DAG). GACG [10] автоматически делит агентов на колаборативные подгруппы на основе их траекторий поведения. GraphMIX [18] использует сеть смешивания, которая отличается от традиционного подхода QMIX [20] благодаря применению архитектур GCN. HAMA [22] использует иерархическую графовую нейросеть с механизмом самовнимания (HGAT) для захвата межслойных отношений сотрудничества.

3. Оптимизация коммуникации

Методы оптимизации коммуникации направлены на разработку механизмов взаимодействия между агентами в многоагентных системах, управляемых графовыми структурами. Их цель — решить проблемы ограничений пропускной способности и избыточности информации в открытых средах. В отличие от традиционных моделей полностью связанного общения или

моделей с управляемыми коммуникациями, эти методы представляют агентов в виде узлов графа и обеспечивают эффективную фильтрацию и агрегацию информации через динамические рёберные соединения и структурированную передачу сообщений. Основная идея заключается в ограничении диапазона коммуникации топологией графа, оптимизации ценности информации с помощью графового рассуждения и балансировании индивидуального принятия решений и группового сотрудничества через графовое обучение. В этом разделе представлено несколько типичных методов.

G2ANet [15] использует механизм жёсткого внимания для выбора ключевых соседних узлов, строит разреженные графы взаимодействий и применяет мягкое внимание для назначения весов сохранённым рёбрам взаимодействия. MAGIC [19] использует дифференцируемый кодировщик графового внимания для генерации направленных динамических графов в реальном времени, определяя таким образом цели коммуникации в каждый момент. GAMA [5] вводит динамическую матрицу смежности, выбирая k ближайших соседей в качестве эффективных целей взаимодействия на основе состояния или статуса движения, и использует механизм графового внимания для взвешенного агрегирования полученной разнородной информации. Sparse-Att [23] применяет двухэтапный механизм внимания, сочетая жёсткое и мягкое внимание, чтобы устанавливать связи только с теми, кто значительно влияет на оптимизацию его политики.

Перспективы применения и теоретические вызовы

Методы мультиагентного обучения с подкреплением (MARL) на основе графов продемонстрировали большой потенциал применения в различных областях. В интеллектуальных транспортных системах этот подход моделирует топологические взаимосвязи перекрёстков и взаимодействия транспортных средств с помощью графовых нейронных сетей, что позволяет оптимизировать работу светофоров на нескольких перекрёстках совместно [26]. В автономном вождении графовые структуры эффективно фиксируют пространственно-временные зависимости между транспортными средствами, а модели принятия решений на основе графовой свёртки [12] повышают безопасность совместной работы в сложных интерактивных сценариях. Кроме того, графовые свёрточные сети (GCN) [7],[8],[9] поддерживают корректировку динамических формаций в реальном времени для адаптации к изменениям в среде выполнения задач.

Тем не менее в этом направлении остаются значительные теоретические проблемы. Во-первых, это вопрос гарантий сходимости [25] для графовых механизмов коммуникации. Когда многоагенты используют

графовые структуры для передачи информации, динамические изменения топологии могут привести к нестабильности при обновлении стратегий. Например, частые изменения топологии сети агентов могут вызывать колебания при обучении.

Выводы

Введение графовых структур в мультиагентное обучение с подкреплением (MARL) предлагает методологическую основу для решения сложных задач совместного принятия решений. Глубокая интеграция теоретико-графового моделирования с основными механизмами MARL

позволила исследователям разработать систему методологий, состоящую из трёх категорий: моделирование среды, оптимизация графа взаимодействия и оптимизация коммуникаций. Эти методы продемонстрировали значительные преимущества в различных сценариях, таких как распределение ресурсов, координация интеллектуальных транспортных средств при автономном вождении. Это подтверждает ценность графовых структур для представления отношений между агентами и агрегации глобальной информации. Основные теоретические проблемы в текущих исследованиях связаны с моделированием динамических графовых данных и фундаментальной теорией коммуникационных механизмов.

ЛИТЕРАТУРА

1. Алфимцев, А.Н. Мультиагентное обучение с подкреплением: учебное пособие / А.Н. Алфимцев. — Москва: Издательство МГТУ им. Н.Э. Баумана, 2021. — С. 79–81.
2. Белянова М.А., Ревунков Г.И., Афанасьев Г.И., Гапанюк Ю.Е. Автоматическая генерация вопросов на основе текстов и графов знаний // Динамика сложных систем — XXI век. 2020. Т. 14. № 4. С. 55–64.
3. Battaglia P.W. et al. Relational inductive biases, deep learning, and graph networks //arXiv preprint arXiv:1806.01261. — 2018.
4. Böhmer W., Kurin V., Whiteson S. Deep coordination graphs //International Conference on Machine Learning. — PMLR, 2020. — С. 980–991.
5. Chen H. et al. Gama: Graph attention multi-agent reinforcement learning algorithm for cooperation //Applied Intelligence. — 2020. — Т. 50. — С. 4195–4205.
6. da Silva L.E.B., Elnabrawy I., Wunsch II D. C. A survey of adaptive resonance theory neural network models for engineering applications //Neural Networks. — 2019. — Т. 120. — С. 167–203.
7. Dai A. et al. Graph convolutional multi-agent reinforcement learning for UAV coverage control //2020 International Conference on Wireless Communications and Signal Processing (WCSP). — IEEE, 2020. — С. 1106–1111.
8. Ding Z., Huang T., Lu Z. Learning individually inferred communication for multi-agent cooperation //Advances in neural information processing systems. — 2020. — Т. 33. — С. 22069–22079.
9. Du Y. et al. Learning correlated communication topology in multi-agent reinforcement learning //Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems. — 2021. — С. 456–464.
10. Duan W., Lu J., Xuan J. Group-aware coordination graph for multi-agent reinforcement learning //arXiv preprint arXiv:2404.10976. — 2024.
11. Foerster J. et al. Learning to communicate with deep multi-agent reinforcement learning //Advances in neural information processing systems. — 2016. — Т. 29.
12. Gao X. et al. Multi-agent decision-making modes in uncertain interactive traffic scenarios via graph convolution-based deep reinforcement learning //Sensors. — 2022. — Т. 22. — №. 12. — С. 4586.
13. Guestrin C., Lagoudakis M., Parr R. Coordinated reinforcement learning //ICML. — 2002. — Т. 2. — С. 227–234.
14. Li S. et al. Deep implicit coordination graphs for multi-agent reinforcement learning //arXiv preprint arXiv:2006.11438. — 2020.
15. Liu Y. et al. Multi-agent game abstraction via graph attention neural network //Proceedings of the AAAI conference on artificial intelligence. — 2020. — Т. 34. — №. 05. — С. 7211–7218.
16. Luo T. et al. Multi-agent collaborative exploration through graph-based deep reinforcement learning //2019 IEEE International Conference on Agents (ICA). — IEEE, 2019. — С. 2–7.
17. Malysheva A., Kudenko D., Shpilman A. Magnet: Multi-agent graph network for deep multi-agent reinforcement learning //2019 XVI International Symposium» Problems of Redundancy in Information and Control Systems»(REDUNDANCY). — IEEE, 2019. — С. 171–176.
18. Naderializadeh N. et al. Graph convolutional value decomposition in multi-agent reinforcement learning //arXiv preprint arXiv:2010.04740. — 2020.
19. Niu Y., Paleja R.R., Gombolay M.C. Multi-Agent Graph-Attention Communication and Teaming //AAMAS. — 2021. — Т. 21. — С. 20th.
20. Rashid T. et al. Monotonic value function factorisation for deep multi-agent reinforcement learning //Journal of Machine Learning Research. — 2020. — Т. 21. — №. 178. — С. 1–51.
21. Ruan J. et al. GCS: Graph-based coordination strategy for multi-agent reinforcement learning //arXiv preprint arXiv:2201.06257. — 2022.
22. Ryu H., Shin H., Park J. Multi-agent actor-critic with hierarchical graph attention network //Proceedings of the AAAI Conference on Artificial Intelligence. — 2020. — Т. 34. — №. 05. — С. 7236–7243.
23. Sun C., Shen M., How J. P. Scaling up multiagent reinforcement learning for robotic systems: Learn an adaptive sparse communication graph //2020 IEEE/RSJ international conference on intelligent robots and systems (IROS). — IEEE, 2020. — С. 11755–11762.
24. Sunehag P. et al. Value-decomposition networks for cooperative multi-agent learning //arXiv preprint arXiv:1706.05296. — 2017.
25. Xu X. et al. The gradient convergence bound of federated multi-agent reinforcement learning with efficient communication //IEEE Transactions on Wireless Communications. — 2023. — Т. 23. — №. 1. — С. 507–528.
26. Yang S. et al. IHG-MA: Inductive heterogeneous graph multi-agent reinforcement learning for multi-intersection traffic signal control //Neural networks. — 2021. — Т. 139. — С. 265–277.