

УДК 004.81/.89/159.922

## Когнитивное развитие в Reinforcement Learning через Curriculum Learning

Лятифов Роман Эйнулла оглы<sup>(\*)</sup>

relyatifov@mail.ru

SPIN-код: 6774-2090

Косыгин Егор Владимирович

egor\_tech23@mail.ru

SPIN-код: 8561-7383

Кондрашов Тимофей Сергеевич

tumakondr@yandex.ru

МГТУ им. Н.Э. Баумана, Москва, Россия

**Аннотация.** Рассмотрена проблема адаптивности и генерализации агентов в Reinforcement Learning (RL) с позиции теорий когнитивного развития. Выдвинута гипотеза, что когнитивные функции агентов развиваются как решение когнитивных проблем, отраженных в структуре среды, аналогично онтогенезу человека. Обосновано применение принципов когнитивного развития Ж. Пиаже к проектированию адаптивных обучающих сред для RL-агентов. Предложена иерархия базовых когнитивных функций и соответствующее им семейство процедурно генерируемых сред, организованных по принципу curriculum learning. Показано, что последовательное обучение на взаимосвязанных задачах, прогрессирующих по сложности, повышает адаптивность и генерализацию агентов и делает ключевым объектом проектирования не архитектуру агента, а структуру среды.

**Ключевые слова:** reinforcement learning, curriculum learning, когнитивное развитие, адаптивность, процедурно генерируемые среды, иерархия задач, Пиаже

## Cognitive development in reinforcement learning through curriculum learning

Lyatifov Roman Eynulla ogly<sup>(\*)</sup>

relyatifov@mail.ru

SPIN-код: 6774-2090

Kosygin Egor Vladimirovich

egor\_tech23@mail.ru

SPIN-код: 8561-7383

Kondrashov Timofey Sergeevich

tumakondr@yandex.ru

BMSTU, Moscow, Russia

**Abstract.** The problem of adaptability and generalization of agents in Reinforcement Learning (RL) is considered from the perspective of cognitive development theories. A hypothesis is proposed that agents' cognitive functions develop as solutions to cognitive problems reflected in the environment structure, analogous to human ontogenetic development. The applicability of Piaget's cognitive development principles to the design of adaptive learning environments for RL agents is substantiated. A hierarchy of basic cognitive functions and a corresponding family of procedurally generated environments organized as a curriculum are proposed. It is shown that sequential training on interconnected tasks of

increasing complexity improves agent adaptability and generalization and shifts the design focus from agent architecture to environment structure.

**Keywords:** reinforcement learning, curriculum learning, cognitive development, adaptability, procedurally generated environments, task hierarchy, Piaget

**Введение.** Современные RL-агенты демонстрируют низкую адаптивность к новым условиям и склонность к переобучению. Попытки решить эту проблему через Meta-RL и Transfer Learning показывают частичный успех, но не адресуют глубинную причину: отсутствие систематического развития базовых когнитивных способностей [1].

Эволюционная теория показывает, что когнитивные изменения организмов являются ответом на проблемы, отраженные в среде [2]. По аналогии в RL когнитивные функции должны рассматриваться не как врожденные компоненты архитектуры, а как вычислительные решения, эволюционирующие в ответ на структурированные задачи среды. Следовательно, ключ к адаптивности лежит в проектировании обучающих сред, последовательно развивающих необходимые способности.

Теория Пиаже предоставляет формальную основу для описания иерархии когнитивных стадий, каждая из которых характеризуется качественно новыми возможностями [3, 4]. Применение этой последовательности к RL подразумевает разработку семейства взаимосвязанных сред, где агент последовательно развивает функции, необходимые для решения все более сложных задач [5].

Целью работы является теоретическое обоснование применения принципов когнитивного развития к проектированию адаптивных сред для RL и демонстрация того, что структура среды, а не архитектура агента, определяет спектр развиваемых способностей.

**Методы и материалы; результаты.** Исследование основано на методе обратного проектирования теорий когнитивного развития. Проанализированы: классические и современные теории развития (Пиаже [1], Паскуаль-Леоне, Демитриу [5], Коммонс [6]); исследования Curriculum Learning в RL [1]; работы по процедурной генерации сред [7, 8]; исследования воплощенного познания [2].

Теоретическое обоснование опирается на три ключевых принципа:

- адаптивное отражение — когнитивные функции отражают структуру проблем в среде;
- иерархия — базовые функции филогенетические (восприятие — действие) [9], высшие развиваются онтогенетически [5];
- последовательность — развитие следует информационной иерархии задач [6, 10].

Пиаже выделил четыре стадии: сенсомоторная (0–2 года), предоперационная (2–7 лет), конкретно-операционная (7–11 лет), формально-операционная (11+ лет) [3, 4]. Каждая характеризуется новыми операциями над информацией и не может быть пропущена.

Неопиажеанские теории объяснили переходы между стадиями: теория Паскуаль — Леоне связала их с ростом емкости рабочей памяти [5, 9], Коммонс показал, что стадии определяются иерархической организацией задач [6, 10]. Ключевой вывод: когнитивные функции — это вычислительные решения когнитивных задач среды. Поэтому правильная архитектура среды автоматически селектирует развитие адаптивных способностей через RL.

Можно выделить 4 уровня иерархии когнитивных функций.

Уровень 1 (сенсомоторное развитие): базовые функции, встроенные в архитектуру: восприятие/представление состояний, действие/исполнение команд. На этом уровне развиваются: (1) координация восприятия-действия; (2) причинно-следственные связи; (3) постоянство объектов.

Уровень 2 (предоперациональное развитие): на основе базовых функций развиваются: (1) символическое представление; (2) внутренние модели; (3) рабочая память с неполной информацией. Агент начинает планировать и моделировать.

Уровень 3 (конкретно-операциональное развитие): развиваются логическое мышление, многошаговое планирование, способность к многозадачности.

Уровень 4 (формально-операциональное развитие): развиваются абстрактное мышление, метаучение, комбинаторное исследование пространства стратегий.

В машинном обучении существует понятие Curriculum Learning — метод, представляющий задачи в возрастающем порядке сложности [4, 7]. Адаптивный curriculum динамически подстраивает сложность в зависимости от возможностей агента, соответствуя принципу зоны ближайшего развития.

Процедурно генерируемые среды (PCG) позволяют создавать бесконечное разнообразие задач в рамках одного семейства [7, 8]. Adversarial PCG автоматически поддерживает оптимальный уровень сложности.

Поэтому предлагается иерархия взаимосвязанных сред, каждая развивающая определенный набор функций.

1. Уровень 1: простые среды координации восприятия-действия (объекты скрываются/ищутся, манипуляции с целями).

2. Уровень 2: среды со средней сложностью — предсказание будущих состояний, работа с обобщенными концепциями, работа при неполной информации.

3. Уровень 3: сложные среды — многошаговое планирование, логические связи, многозадачность.

4. Уровень 4: адаптивные среды — новые домены, метаучение, комбинаторное исследование.

Параметры масштабирования: горизонт планирования, размерность состояния, стохастичность, задержка вознаграждения, глубина причинных связей. Агент переходит на следующий уровень при достижении порога производительности.

**Обсуждение полученных результатов.** Предложенный подход резонирует с современными направлениями: Meta-RL подразумевает обучение ал-

горитму обучения [1]; Multi-task Learning демонстрирует, что совместное обучение на связанных задачах улучшает генерализацию [11]; Embodied RL подчеркивает роль взаимодействия морфологии и среды [2].

Преимущества подхода являются: улучшение генерализации — агент применяет развитые функции к новым задачам; снижение требований к данным — последовательное обучение требует меньше данных; стабильность обучения — постепенное усложнение снижает нестабильность; масштабируемость — новые уровни добавляются без переучивания.

При этом выделяется и ряд ограничений: определение оптимальной последовательности остается открытым вопросом; не все функции переносятся автоматически; вычислительная стоимость может быть выше при плохом дизайне среды; требуются надежные метрики для оценки развития функций.

Центральная гипотеза — что функции эволюционируют как отражение задач среды — получает поддержку от исследований воплощенного интеллекта [2, 12].

**Заключение.** Основной вывод: адаптивность RL-агентов повышается не путем совершенствования архитектур, а проектированием структурированных сред, последовательно развивающих когнитивные функции. Выдвинута гипотеза о том, что когнитивные функции — это вычислительные решения проблем, отраженных в среде.

Предложена практическая архитектура многоуровневой иерархии сред, соответствующей стадиям Пиаже, с параметрами масштабирования и адаптивным управлением переходами. Подход открывает перспективы как для понимания адаптации в RL, так и для разработки гибких систем ИИ.

Дальнейшее развитие требует: экспериментальной проверки на RL-платформах; разработки метрик оценки когнитивных функций; исследования оптимальной последовательности для различных доменов; интеграции с Meta-RL и In-Context Learning.

## Список источников

- [1] Narvekar S. et al. Curriculum learning for reinforcement learning domains: A framework and survey. *Journal of Machine Learning Research*, 2020, vol. 21, no. 181, pp. 1–50.
- [2] Gupta A. et al. Embodied intelligence via learning and evolution. *Nature communications*, 2021, vol. 12, no. 1, art. no. 5721.
- [3] Piaget J. *The construction of reality in the child*. Basic Books, New York, 1954.
- [4] McLeod S. Piaget's theory and stages of cognitive development. *Simply Psychology*, 2018, pp. 1–34. <https://doi.org/10.5281/zenodo.15241970>
- [5] Demetriou A., Spanoudis G., Mouyi A. *A three-level model of the developing mind: Functional and neuronal substantiation and educational implications*. Dordrecht, Springer Netherlands, 2010, pp. 9–48.
- [6] Commons M.L. Introduction to the model of hierarchical complexity. *Behavioral Development Bulletin*, 2007, vol. 13, no. 1, art. no. 1.
- [7] Khalifa A. et al. Pcgrl: Procedural content generation via reinforcement learning. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, 2020, vol. 16, no. 1, pp. 95–101.

- [8] Liu J. et al. Deep learning for procedural content generation. *Neural Computing and Applications*, 2021, vol. 33, no. 1, pp. 19–37.
- [9] Constantinidis C., Klingberg T. The neuroscience of working memory capacity and training. *Nature Reviews Neuroscience*, 2016, vol. 17, no. 7, pp. 438–449.
- [10] Commons M.L., Richards F.A. A general model of stage theory. *Beyond formal operations*, 1984, vol. 1, pp. 120–140.
- [11] Bengio Y., Courville A., Vincent P. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 2013, vol. 35, no. 8, pp. 1798–1828.
- [12] Lu R. et al. Discovery of the reward function for embodied reinforcement learning agents. *Nature Communications*, 2025, vol. 16, no. 1, art. no. 11064.