

# СОЗДАНИЕ АГЕНТА НА ОСНОВЕ МОДЕЛИ ИЕРАРХИЧЕСКОЙ ТЕМПОРАЛЬНОЙ ПАМЯТИ

**Канонир Георгий**

Аспирант,

Университет ИТМО (Санкт-Петербург)

kanonirs@gmail.com

## CREATING AN AGENT BASED ON HIERARCHICAL TEMPORAL MEMORY MODEL

**G. Kanonirs**

*Summary:* Modern reinforcement learning methods have a number of limitations imposed by the used artificial neural networks paradigm with a point neuron model. The use of the latest achievements of neuroscience within a new theory of intelligence called The Thousand Brains Theory of Intelligence, as well as the application of the machine learning model called Hierarchical Temporal Memory (HTM), which implements some aspects of this theory, have the potential both to develop already established reinforcement learning methods, and to create new approaches for solving this problem. The purpose of this work is to propose and analyze a conceptual idea for creating an agent based on this model, allowing this task to be done in the most natural way, i.e. based on its underlying structural and functional principles, and without specific requirements for a reward signal.

*Keywords:* biologically-plausible machine learning methods, reinforcement learning, hierarchical temporal memory.

### Введение

Современные методы обучения с подкреплением имеют ряд ограничений, наложенных используемой парадигмой искусственных нейронных сетей с точечной моделью нейрона, включая слабую устойчивость к шуму во входных данных [1], низкую эффективность хранения информации в модели, приводящей к появлению проблемы катастрофического забывания и невозможности непрерывного обучения [2], а также низкую эффективность процесса обучения [3]. Использование последних достижений нейронаук в рамках новой теории интеллекта — «теории тысячи мозгов» (The Thousand Brains Theory of Intelligence) [4], а также применение модели «иерархической темпоральной памяти» (Hierarchical Temporal Memory, HTM) [5], частично реализующей данную теорию в виде модели машинного обучения, имеют потенциал как для развития уже устоявшихся методов обучения с подкреплением, так и для создания новых подходов решения этой задачи. Ранее автором данной работы была рассмотрена потенциальная возможность применения модели HTM в обучении с подкреплением, а также были проведены поиск и анализ научных исследований, посвящённых применению модели HTM в данной области [6].

В большинстве рассмотренных работ осуществлялась попытка использования только отдельных эле-

*Аннотация.* Современные методы обучения с подкреплением имеют ряд ограничений, наложенных используемой парадигмой искусственных нейронных сетей с точечной моделью нейрона. Использование последних достижений нейронаук в рамках новой теории интеллекта — «теории тысячи мозгов» (The Thousand Brains Theory of Intelligence), а также применение модели «иерархической темпоральной памяти» (Hierarchical Temporal Memory, HTM), частично реализующей данную теорию в виде модели машинного обучения, имеют потенциал как для развития уже устоявшихся методов обучения с подкреплением, так и для создания новых подходов решения этой задачи. Целью данной работы является предложение и анализ концептуальной идеи для создания агента на основе данной модели, позволяющая осуществить эту задачу наиболее естественно, т.е. на основе её основополагающих структурных и функциональных принципов, и без особых требований к сигналу вознаграждений.

*Ключевые слова:* биологически-правдоподобные методы машинного обучения, обучение с подкреплением, иерархическая темпоральная память.

ментов данной модели в комбинации с более традиционными подходами для обретения полезных свойств, но также опубликованы работы, авторы которых стремились создать самостоятельную модель для обучения с подкреплением на основе модели HTM. Последние представляют наибольший интерес, но даже они оставляют вопрос построения архитектуры агента открытым. Причиной тому является незавершенность модели HTM, которая в настоящее время предлагает структурные элементы и алгоритмы для формирования блока пространственно-темпоральной памяти для извлечения и хранения пространственно-темпоральных закономерностей на основе потока входных данных, но не предлагает готового решения для выполнения сенсорно-моторной интеграции и формирования памяти для хранения сенсорно-моторного опыта агента. Кроме того, выбор наиболее предпочтительного действия на основе хранящегося в модели сенсорно-моторного опыта мало изучен.

Целью данной работы является предложение и анализ концептуальной идеи для создания агента на основе модели HTM, позволяющая осуществить эту задачу наиболее естественно, то есть на основе её основополагающих структурных и функциональных принципов. Более того, предлагаемое решение не выдвигает особых требований к сигналу вознаграждения, в отличие от рассмотренных ранее архитектур, требующих наличия исключительно бинарного сигнала.

### Иерархическая Темпоральная Память

Прежде всего рассмотрим саму модель НТМ [5]. Её ключевым элементом является блок темпоральной памяти, выполняющий извлечение и хранение пространственно-темпоральных закономерностей на основе потока входных данных. На практике обычно используется один блок темпоральной памяти, которому на вход подаются данные из потока, а его состояние в каждый момент времени используется как вход для внешнего алгоритма с целью прогнозирования данных или выявления аномалий в потоке данных. Построение классической иерархии теоретически возможно, но в настоящее время данный вопрос мало изучен.

Темпоральная память представляет собой набор мини-колонок, каждая из которых представляет собой множество НТМ нейронов имеющих идентичное репетитивное поле, определяемое синаптическими связями, формируемыми на единственном проксимальном дендритном сегменте каждого нейрона. При этом нейрон может иметь один или более дистальных дендритных сегментов, на каждом из которых формируются модулирующие синаптические связи. Последние определяют темпоральный контекст, при достаточной активации которого будет активирован и сам дистальный сегмент. Кроме того, нейрон может иметь один или более апикальных дендритных сегментов, которые играют похожую роль, но получают данные из источника, отличного от того, от которого получают данные дистальные дендритные сегменты. Как будет показано далее, наличие двух различных источников модуляции имеет ключевое значение при создании агента на основе модели НТМ.

Функционирование темпоральной памяти на каждой итерации осуществляется в три шага. На первом шаге происходит преобразование входного образа в разряженное распределённое представление. На втором шаге полученное представление помещается в темпоральный контекст. На заключительном шаге выполняется прогнозирование множества представлений потенциальных входных образов в следующий момент времени.

Агенты создаются для выполнения поставленной задачи, а для успешного выполнения задачи агенту необходимо знание своего текущего состояния и наличие соответствующей стратегии принятия решений. Состояние агента определяется исключительно его сенсорно-моторным взаимодействием со средой. А именно, в каждый момент времени оно определяется сенсорно-моторной интеграцией текущих наблюдений и действий в контексте ранее накопленного опыта. Для выполнения упомянутой задачи требуется определение как архитектуры, так и принципов её функционирования, но в данный момент предположим, что на каждой итерации

у нас уже имеется разряженное распределённое представление, отражающее состояние агента в текущий момент времени. Отталкиваясь от этого предположения, далее будет предложено решение для превращения модели НТМ в модель машинного обучения с подкреплением, осуществляя это наиболее естественным образом, то есть на основе её основополагающих структурных и функциональных принципов.

Ключевым элементом предлагаемого решения является блок темпоральной памяти НТМ, использующийся для формирования памяти действий. На его основной (проксимальный) вход поступает представление действия, а на его модулирующий (дистальный) вход поступает представление состояния агента в предшествующий выбору этого действия момент времени. В таком случае на заключительном шаге каждой итерации в темпоральной памяти выполняется прогнозирование множества представлений потенциальных действий на следующий момент времени. Соответственно, агенту остается только решить дилемму эксплуатации/исследования, то есть выбрать либо одно из прогнозируемых, либо ранее не выбиравшееся действие.

Для определения оптимальной стратегии и появления возможности более целенаправленного выбора следующего действия добавляем модуль оценивания. Данный модуль получает представление действия в контексте предшествующего выбору этого действия состояния агента, то есть аналог пары состояния/действия для Q-функции. Далее данный модуль выполняет оценивание и формирует обратные связи с блоком памяти действий на основе полученных оценок. Сформированные обратные связи отражают выработанную агентом стратегию поведения, а сам модуль с их помощью создаёт дополнительное предсказательное смещение, позволяющее из множества потенциальных действий выбрать то, что наиболее предпочтительно с точки зрения сигнала вознаграждения.

При этом модуль оценивания структурно может представлять собой просто множество НТМ нейронов, формирующих прямые (проксимальные для данного модуля) и обратные (апикальные для блока памяти действий) синаптические связи с памятью действий, а также латеральные (дистальные для данного модуля) синаптические связи между нейронами данного модуля. Прямые синаптические связи необходимы для получения входных данных и формирования на их основе рецептивных полей нейронов. Обратные синаптические связи, как уже было сказано ранее, формируются на основе полученных результатов оценивания и отражают выработанную агентом стратегию поведения. Формирование латеральных синаптических связей, осуществляемое исключительно между активными нейронами текущего представления, необходимо для формирования дис-

тальных дендритных сегментов, хранящих результаты оценивания текущего представления.

Поскольку в будущем предполагается функционирование агента полностью онлайн для решения как эпизодических, так и непрерывных задач, то наиболее предпочтительными кандидатами на роль методов оценивания представляются методы на основе временных различий.

### Заключение

Современные методы обучения с подкреплением имеют ряд ограничений, наложенных используемой парадигмой искусственных нейронных сетей с точечной моделью нейрона. Модель НТМ имеет потенциал для решения всех вышеупомянутых проблем, в первую очередь, благодаря использованию принципа разряженности (синаптических связей и активации нейронов) и оперированию разряженными распределёнными представлениями. Данное обстоятельство приводит к тому, что представления являются более устойчивыми к шуму и ёмкость модели значительно возрастает [7–9],

а её функционирование может быть более эффективным и менее ресурсоёмким с точки зрения вычислений [10]. Кроме того, использование активных дендритов позволяет в значительной мере смягчить проблему катастрофического забывания и позволяет осуществлять непрерывное обучение [11].

В настоящее время модель НТМ далека от своего завершения и, с точки зрения теории, лежащей в её основе, и не является моделью машинного обучения с подкреплением. Предлагаемое решение позволяет достичь последнего и сделать это наиболее естественным образом, то есть на основе её основополагающих структурных и функциональных принципов. Кроме того, процесс оценивания не выдвигает никаких специфических требований к сигналу вознаграждения, что позволит в дальнейшем использовать любые ресурсы для сравнения с устоявшимися методами на известных бенчмарках. В будущем планируется более детальная проработка предлагаемого решения, а также его реализация и апробация на задаче о контекстуальном многооруком бандите с мгновенным или отложенным вознаграждением.

### ЛИТЕРАТУРА

1. Liu M. et al. Analyzing the noise robustness of deep neural networks // Proceedings of 2018 IEEE Conference on Visual Analytics Science and Technology (VAST). — p. 60–71.
2. Goodfellow I.J. et al. An empirical investigation of catastrophic forgetting in gradient-based neural networks. — 2013. URL: <https://arxiv.org/abs/1312.6211> (дата обращения: 07.04.2023).
3. Thompson N. C. et al. The computational limits of deep learning. — 2020. URL: <https://arxiv.org/abs/2007.05558>. (дата обращения: 07.04.2023).
4. Hawkins J.A. A thousand brains: A new theory of intelligence. Монография. — 2021. — 288 с.
5. Hawkins, J. et al. Biological and Machine Intelligence. — 2016 — 2020. URL: <https://numenta.com/resources/biological-and-machine-intelligence/> (дата обращения: 07.04.2023).
6. Канонир Г. Обучение с подкреплением на основе модели иерархической темпоральной памяти // Современная наука: актуальные проблемы теории и практики. Серия: Естественные и технические науки. — 2022. — № 10. — С. 80–83.
7. Ahmad S., Hawkins J. Properties of sparse distributed representations and their application to hierarchical temporal memory. — 2015. URL: <https://arxiv.org/abs/1503.07469>. (дата обращения: 07.04.2023).
8. Ahmad S., Scheinkman L. How can we be so dense? The benefits of using highly sparse representations. — 2019. URL: <https://arxiv.org/abs/1903.11257>. (дата обращения: 07.04.2023).
9. Ahmad S., Hawkins J. How do neurons operate on sparse distributed representations? A mathematical theory of sparsity, neurons and active dendrites. — 2016. URL: <https://arxiv.org/abs/1601.00720>. (дата обращения: 07.04.2023).
10. Hunter K., Spracklen L., Ahmad S. Two sparsities are better than one: unlocking the performance benefits of sparse–sparse networks // Neuromorphic Computing and Engineering. — 2022. — vol. 2. — no. 3. — p. 1–26.
11. Iyer A. et al. Avoiding catastrophe: Active dendrites enable multi-task learning in dynamic environments // Frontiers in neurobotics. — 2022. — vol. 16. — p. 1–23.

© Канонир Георгий (kanonirs@gmail.com)

Журнал «Современная наука: актуальные проблемы теории и практики»