DOI 10.37882/2223-2966.2024.7.17

ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ ОБУЧЕНИЯ АГЕНТА НА ОСНОВЕ МОДЕЛИ ИЕРАРХИЧЕСКОЙ ТЕМПОРАЛЬНОЙ ПАМЯТИ

INCREASING THE EFFICIENCY OF AGENT TRAINING BASED ON THE HIERARCHICAL TEMPORAL MEMORY MODEL

G. Kanonir
A. Filchenkov

Summary. Modern methods of reinforcement learning have a number of limitations imposed by the used paradigm of artificial neural networks with a point model of a neuron. The use of the «hierarchical temporal memory» (HTM) model has the potential both for the development of already established training methods and for the creation of new ones. The aim of this paper is to propose a new design of a spatial-temporal memory unit that allows an agent based on the HTM model to take into account a temporal context of limited length and, due to this, to increase the efficiency of its learning when solving problems in which the actual reward received depends on a temporal context of a size smaller than the maximum length of the sequences of observations and actions considered within the framework of the problem being solved.

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

Канонир Георгий
Аспирант, Университет ИТМО
kanonirs@gmail.com
Фильченков Андрей Александрович
к.ф.-м.н., Университет ИТМО
afilchenkov@itmo.ru

Аннотация. Современные методы обучения с подкреплением имеют ряд ограничений, наложенных использующейся парадигмой искусственных нейронных сетей с точечной моделью нейрона. Использование модели «иерархической темпоральной памяти» (HTM) имеет потенциал как для развития уже устоявшихся методов обучения, так и для создания новых. Целью данной работы является предложение нового дизайна блока пространственно-темпоральной памяти, позволяющего агенту на основе модели HTM учитывать темпоральный контекст ограниченной длинны и, за счёт этого, повысить эффективность его обучения при решении задач, в которых фактически получаемое вознаграждение зависит от темпорального контекста размера меньшего, чем максимальная длинна рассматриваемых в рамках решаемой задачи последовательностей наблюдений и действия.

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

Введение

Современные методы обучения с подкреплением имеют ряд ограничений, наложенных использующейся 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 предложена простая и легко интерпретируемая одноуровневая архитектуры агента на основе модели HTM [6]. Архитектура структурно представляет собой три взаимодействующих модуля: (1) память агента, построенная на основе блоков пространственно-темпоральной памяти HTM, использующейся для хранения сенсорно-моторного опыта агента; (2) модуль, выполняющий оценивание выполняемых агентом действий, а также формирующий обратные синаптические связи, отражающие выработанную агентом стратегию поведения и играющие ключевую роль при выборе следующего действия; (3) модуль, ответственный за выбор следующего действия агента (учитывая предпочтения, основанные на предыдущем опыте агента). Несмотря на то, что предложенная архитектура была успешно апробирована как на задаче о классическом, так контекстуальном многоруком бандите с мгновенным или отложенным вознаграждением, для некоторых постановок задачи анализ полученных результатов показал низкую эффективность процесса обучения. В первую очередь данная проблема затрагивает случаи, когда фактически получаемое вознаграждение зависит от темпорального контекста размера меньшего, чем максимальная длинна рассматриваемых в рамках решаемой задачи последовательностей наблюдений и действия.

Целью данной работы является предложение нового дизайна блока пространственно-темпоральной памяти, позволяющего агенту учитывать темпоральный контекст различной длинны, но ограничением верхнего значения.

Модель HTM

Блок пространственно-темпоральной памяти HTM структурно представляет собой набор мини-колонок или групп нейронов. Любые два нейрона из одной мини-колонки имеют идентичное рецептивное поле, т.е. они реагируют на идентичные паттерны во входном образе, но для нейронов из разных колонок такой гарантии нет и, вероятнее всего, такие нейроны будут иметь в меньшей или большей степени различные рецептивные поля. При этом наличие множества нейронов в каждой мини-колонке используется для предоставления возможности формирования представления некоторого образа в различных темпоральных контекстах.

Число нейронов в каждой мини-колонке одинаковое и оно определяет количество контекстов, в которых может быть представлен входной образ с гарантированной возможностью отличить представления одного и того же образа в различных контекстах. При этом в оригинальном алгоритме модели HTM отсутствует возможность ограничения длинны используемого темпорального контекста, что и является тем фактором, что приводит к ранее описанной проблеме.

Постановка проблемы

Рассмотрим проблему более наглядно, решая эпизодическую задачу о контекстуальном многоруком бандите с отложенным вознаграждением и предположим, что есть блок пространственно-темпоральной памяти HTM (в рамках рассматриваемой архитектуры агента называется памятью состояний агента), на вход которого подается представление, отражающее некоторое наблюдение в контексте предшествующего ему действия. Для примера, пусть набор данных будет состоять из двух последовательностей $x = [x_1, \ldots, x_m, z_1, \ldots, z_k]$ и $y = [y_1, \ldots, y_n, z_1, \ldots, z_k]$, имеющих различные префиксы-подпоследовательности: $[x_1, \ldots, x_m]$ и $[y_1, \ldots, y_n]$, за которым следует общая часть — подпоследовательность длинны k: $[z_1, \ldots, z_k]$. При этом получаемое в конце эпизода вознаграждение зависит от темпорального контекста длинны k т.е. зависит только от общей подпоследовательности, а принимаемое агентом решение зависит от представления z_k в темпоральном контексте.

Поскольку модель HTM не предоставляет возможности ограничить длину используемого темпорального контекста, представления z_k в темпоральном контексте для двух последовательностей будут различны. Следовательно, для выработки оптимальной стратегии поведения агенту необходимо проделать двойную работу, определяя её отдельно для каждой из двух наблюдаемых последовательностей.

Метод

Для снятия ранее рассмотренного ограничения предлагается концепция и дизайн памяти произвольного порядка (рис. 1), в которой максимальная длина используемого темпорального контекста является гиперпараметром модели. Создание такой памяти представляется наиболее естественными за счёт объединения блоков памяти первого порядка, состояние которого зависит только от его состояния в предшествующий момент времени, и памяти высокого порядка, состояния которого соответственно зависит от темпорального контекста высокого порядка. Далее новый дизайн блока пространственно-темпоральной памяти будет рассмотрен более подробно, но главная его суть заключается в формировании в каждый момент времени особого представления входных данных, являющегося суперпозицией представлений, построенных на основе темпоральных контекстов разных порядков.

Рис. 1. Блок пространственно-темпоральной памяти произвольного порядка

В оригинальном алгоритме модели HTM состояние блока пространственно-темпоральной памяти, представимое в виде бинарного вектора, является представлением его входного образа в данный момент времени, но учитывая темпоральный контекст. При этом каждый элемент состояния блока памяти отражает состояние соответствующего нейрона, т.е. является ли нейрон в данный момент времени активным или нет. Базовым изменением в рамках нового дизайна блока памяти является замена бинарного выхода HTM нейрона на выход в виде бинарного вектора $a = [a1, \ldots, ak]$, где на i-той позиции при $i = 1...k$ будет 1, если нейрон стал активен за счёт темпорального контекста i-ого порядка или 0, в противном случае.

Создание блока памяти первого порядка достигается за счёт использования оригинального блока памяти HTM, установив значение нейронов в каждой мини-колонке равным единице, и замены выхода каждого нейрона на бинарный вектор, в котором единица может

быть только на первой позиции, что отражает активацию нейрона за счёт темпорального контекста первого порядка.

Создание блока памяти высокого порядка достигается за счёт использования оригинального блока памяти HTM, но с изменением алгоритма активации и формирования выхода нейрона. Благодаря тому, что каждый нейрон теперь обладает информацией за счет темпорального контекста каких порядков были активны нейроны в предшествующий момент времени, на выходе нейрона формируется бинарный вектор, в котором на *i*-ой позиции будет единица, если данный нейрон стал активен за счет нейронов, которые были активны в предшествующий момент времени за счет темпорального контекста (*i*–1)-ого порядка.

Таким образом выход блока памяти первого порядка формирует своего рода реперное представление для начала последовательности темпорального контекста, а блок памяти высокого порядка использует полученное реперное представление и на его основе строит представления разных порядков. В свою очередь выход блока памяти произвольного порядка, представимый в виде бинарного тензора, является суперпозицией представлений, построенных на основе темпоральных контекстов разных порядков.

Апробация и результаты

Для апробации предлагаемого решения была выбрана задача о контекстуальном многоруком бандите с отложенным вознаграждением, в которой получаемое в конце каждого эпизода вознаграждение зависит от темпорального контекста ограниченной длинны.

В рамках апробации было проведено два эксперимента — с использованием оригинального прототипа агента и с использованием модифицированного прототипа агента, способного учитывать темпоральный контекст разного порядка. В обоих случаях каждый эпизод состоял из n шагов, на каждом из которых агент выбирал одно из m возможных действий, и только в конце эпизода агент получал отличное от нуля вознаграждение. Отличие экспериментов заключалось в том, что в первом случае вознаграждение завесило от выбора агента на каждом шаге эпизода, т.е. оно зависело от полного темпорального контекста, а во втором — только от выбора агента на последних k шагах, где $k < n$, т.е. оно зависело от ограниченного темпорального контекста. Пример задачи (рис. 2) отражает распределение вознаграждений для каждой последовательности действий (состоящей из трех шагов). Ценности последовательностей действий $q^*(a)$, где a = `111`, `112`, …, `222`, выбирались из нормального распределения со средним равным нулю и единичной дисперсией. В конце каждого эпизода агент получал вознаграждение, которое выбиралось из нормального распределения со средним $q^*(a_t)$, где a_t — это фактически выбранная последовательность действий на шаге t, и единичной дисперсией.

На рис. 3 показаны результаты экспериментов при $n = 4$, $m = 2$, $k = 3$ — среднее вознаграждение (слева)

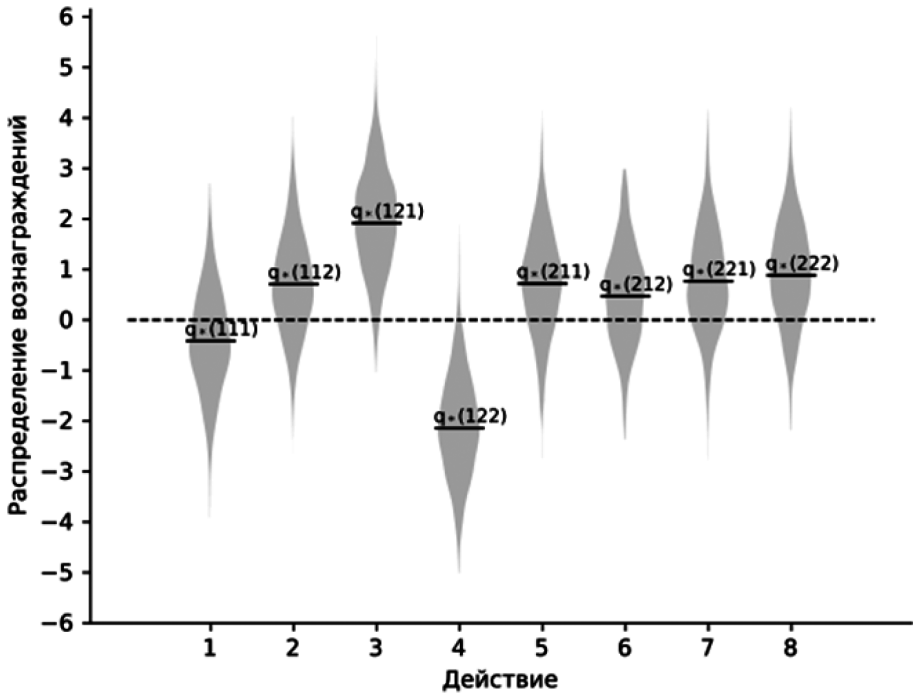

Рис. 2. Пример задачи о контекстуальном многоруком бандите с отложенным вознаграждением (распределения показаны серым цветом)

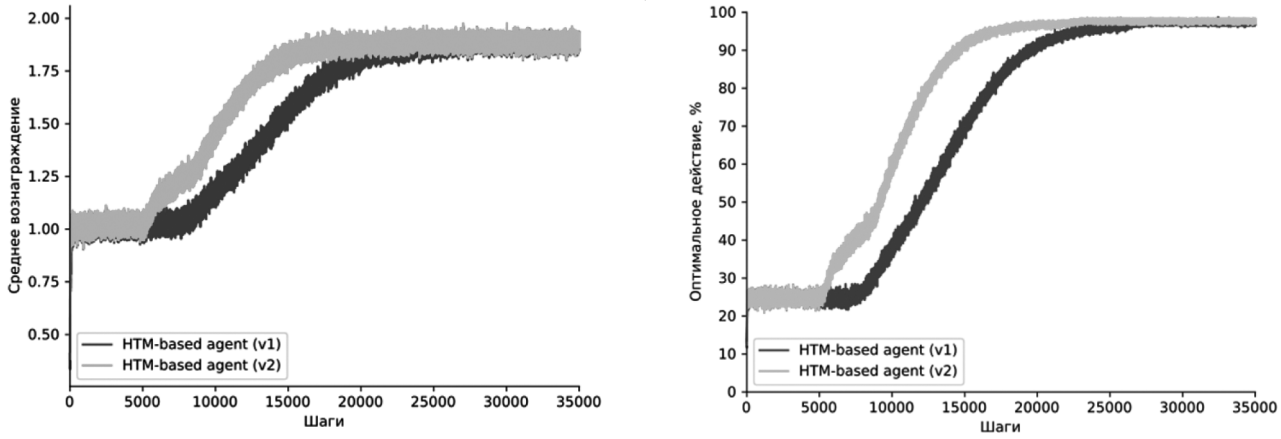

Рис. 3. Задача контекстуальном многоруком бандите с отложенным вознаграждением: среднее вознаграждение (слева) и доля выбора оптимального действия (справа)

и доля выбора оптимальной последовательности действий справа, полученные с использованием оригинального и модифицированного прототипов агента на основе модели HTM. Оценки были получены усреднением по выборке после проведения эксперимента на фиксированной задаче 2000 раз. Как оригинальный, так и модифицированный прототипы агента успешно справляются с поставленной задачей, но последний вырабатывает и переходит к использованию оптимальной стратегии поведение даже на такой простой задаче заметно быстрее.

Заключение

Модель HTM обладает множеством преимуществ по сравнению с традиционными искусственными нейронными сетями. В настоящее время ее основополагающие принципы, а также алгоритм, обеспечивающий ее функционирование, позволяют наиболее успешно применять данную модель для решения задач, требующих выявления темпоральных закономерностей, например, с целью прогнозирования или выявления аномалий на основе потока входных данных. Тем не менее, в рамках исследования о возможности применения модели HTM для решения задач обучения с подкреплением, было выявлено значительное ограничение модели, приводящее к низкой эффективности процесса обучения из-за невозможности ограничить длину используемого в каждый момент времени темпорального контекста.

Предлагаемый в рамках данной работы новый дизайн блока пространственно-темпоральной памяти позволяет естественным образом ограничить максимальный размер используемого темпорального контекста. Данная возможность позволяет существенно повысить эффективность обучения для случаев, когда получаемое агентом вознаграждение зависит от темпорального контекста меньшего порядка, чем длина возможных темпоральных последовательностей. При этом необходимо отметить, что предлагаемое решение не исключает полностью возможности «заглянуть» в более далекое прошлое, поскольку модель HTM предполагает в принципе наличие иерархии, формирование на каждом ее уровне темпорально стабильных представлений, охватывающих более длительные интервалы времени и образование обратных связей, обеспечивающих ниже расположенные уровни информацией об обобщенном темпоральном контексте более высокого уровня.

ЛИТЕРАТУРА

1. Liu M. et al. Analyzing the noise robustness of deep neural networks // 2018 IEEE Conference on Visual Analytics Science and Technology (VAST). — IEEE, 2018. — с. 60–71.
2. Goodfellow I.J. et al. An empirical investigation of catastrophic forgetting in gradient-based neural networks // arXiv preprint arXiv:1312.6211. — 2013.
3. Thompson N. C. et al. The computational limits of deep learning // arXiv preprint arXiv:2007.05558. — 2020.
4. Hawkins J. 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: 01.05.2024).
6. Канонир Г. (науч. рук. Фильченков А.А.) Одноуровневая архитектура агента на основе модели иерархической темпоральной памяти // Сборник тезисов докладов конгресса молодых ученых. Электронное издание. — СПб: Университет ИТМО, [2023]. URL: https://kmu.itmo.ru/digests/article/9901

© Канонир Георгий (kanonirs@gmail.com); Фильченков Андрей Александрович (afilchenkov@itmo.ru)
Журнал «Современная наука: актуальные проблемы теории и практики»