

ОБУЧЕНИЕ С ПОДКРЕПЛЕНИЕМ НА ОСНОВЕ ИЕРАРХИЧЕСКОЙ ТЕМПОРАЛЬНОЙ МОДЕЛИ

Канонир Г. (Университет ИТМО)
Фильченков А.А. – к. ф.-м. н. (Университет ИТМО)

Современные методы обучения с подкреплением имеют ряд ограничений, наложенных используемой парадигмой искусственных нейронных сетей с точечной моделью нейрона. Использование последних достижений нейронаук в рамках теории интеллекта "The Thousand Brains Theory of Intelligence", а также применение модели машинного обучения иерархической темпоральной памяти (НТМ), которая реализует некоторые аспекты данной теории, имеют потенциал как для развития уже устоявшихся методов обучения с подкреплением, так и для создания новых подходов решения этой задачи. Целью данной работы является выявление перспектив применения модели машинного обучения НТМ в обучении с подкреплением.

Введение.

Обучение с подкреплением является направлением машинного обучения, в рамках которого моделируется взаимодействие агента с некоторой средой, в которой данный агент находится. Целью обучения является определение оптимальной стратегии принятия решений, основываясь на (возможно, неполных) наблюдениях среды и некотором сигнале вознаграждения, который косвенно даёт отложенную оценку принятых решений и таким образом ставит перед агентом задачу.

Для принятия эффективных решений агенту необходимо оценивать свое текущее состояние, основываясь на предыдущем опыте, т. к. именно оно определяет, какие решения агент может принять в данный момент и принятие какого решения может быть наиболее выгодно. В большинстве реальных задач пространство состояний, а в ряде случаев и пространство действий настолько велики, что единственным выходом является применение аппроксимирующих методов.

Современные методы обучения с подкреплением используют парадигму искусственных нейронных сетей с точечной моделью нейрона для выполнения таких аппроксимаций. Это налагает ряд ограничений на подобные методы, включая слабую устойчивость к шуму во входных данных, низкую эффективность хранения информации в модели, приводящей к появлению проблемы катастрофического забывания и невозможности непрерывного обучения, а также низкую эффективность самого процесса обучения в целом. Кроме того, в процессе обучения с подкреплением обычно используется буфер для хранения некоторого множества данных, обновление которого тоже представляет непростую задачу, полученных от взаимодействия агента с окружающей средой, которые используются в процессе обучения. В реальных задачах получение такого опыта может быть очень затратным и трудоёмким, а кроме того, и сам процесс обучения может неэффективным при отсутствии достаточного количества и разнообразия опыта.

Основная часть.

В настоящее время достигнуты огромные успехи при решении некоторых задач с помощью методов машинного обучения в целом и методов обучения с подкреплением в частности, но тем не менее наиболее продвинутыми агентами остаются живые существа,

обладающие мозгом, а точнее говоря, те из них, у которых есть неокортекс, представляющий собой отдел мозга, отвечающий за интеллект.

Одной из наиболее перспективных теорий интеллекта, учитывающей данные исследований принципов строения и функционирования мозга, является теория интеллекта «The Thousand Brains Theory of Intelligence», разрабатываемая Дж. Хоукинсом и его коллегами из компании Numenta. Помимо этого исследователи и инженеры компании занимаются разработкой модели машинного обучения иерархической темпоральной памяти (Hierarchical Temporal Memory, HTM), постепенно внедряя свои разработки в стремлении реализовать теорию интеллекта Дж. Хоукинса в виде вычислительной модели.

Модель HTM является нейронной сетью, но использует более сложную и приближенную к естественному модель нейрона, а также имеет более сложную организацию нейронов, формирующих нейронную сеть. Основной компонентой модели HTM является темпоральная память, обладающая способностью прогнозирования, формирование которой происходит за счет преобразования входных данных в распределенное разряженное представление и помещение этого представления в темпоральный контекст. Основопологающим в модели HTM является использование свойства разряженности и «активных дендритов», что и является главным фактором, позволяющим избежать проблем, возникающим при использовании традиционных искусственных нейронных сетей.

Как было отмечено, наличие и разнообразие опыта играет очень важную роль в процессе обучения, но при этом помимо использования опыта реального взаимодействия со средой возможно использование модели для имитации окружающей среды и порождения имитационного опыта. Благодаря планированию, можно использовать такой опыт для порождения/улучшения стратегии или более целенаправленного исследования среды. Основываясь на исследованиях, посвященных принципам, механизмам, свойствам и применениям модели HTM как при решении других, так и смежных задач, есть основания полагать, что ее использование может дать возможность не только эффективно строить модель окружающей среды, взаимодействия с ней агента и воспроизводить накопленный опыт, но и использовать прогнозирующую способность модели HTM для генерации ранее не наблюдавшегося, т. е. имитационного опыта.

В настоящем исследовании мы рассмотрим подходы для использования HTM в задачах обучения с подкреплением и очертим перспективы их применения.

Выводы.

Применение модели машинного обучения HTM имеет огромный потенциал для развития методов обучения с подкреплением как за счёт решения известных проблем, так и с помощью добавления новых возможностей (прим., непрерывное обучение). Кроме того, развитие направления обучения с подкреплением с применением биологически-правдоподобных принципов и методов может потенциально привести к лучшему пониманию работы мозга и интеллекта как следствия его функционирования.

Канонир Г. (автор)

Подпись

Фильченков А.А. (научный руководитель)

Подпись