



Российские ученые из Института искусственного интеллекта AIRI и студенты МФТИ создали первую в России биологически-правдоподобную модель памяти для систем искусственного интеллекта с внутренней мотивацией

Когнитивный агент, если выразаться простым языком, это программа, которая учится самостоятельно взаимодействовать с миром и обучаться на своих ошибках, выполняя конкретную задачу. Основой агента служит архитектура из алгоритмов, в том числе и нейросетевых, которая помогает ему выполнять инструкции разработчика.

В повседневной жизни мы регулярно сталкиваемся с результатами работы методов машинного обучения и искусственного интеллекта. За последнее десятилетие успех в этом направлении был связан с обучением глубоких нейронных сетей, (*Artificial Neural Networks, ANN*), построенных на основе модели искусственного нейрона. Исследователи также выделяют спайковые нейронные сети (*Spiking Neural Networks, SNN*), построенные на основе модели спайкового нейрона, которая является более близкой к биологическому нейрону. Искусственные нейронные сети обмениваются вещественными числами, а вторые – спайками, единичными событиями, происходящими в определенное время, максимально повторяя работу нервной системы.

Искусственные нейронные сети более распространены за счет простоты используемой модели нейрона, а архитектура графических ускорителей очень удобна для связанных с ними вычислений. Они используют для передачи информации все содержащиеся в них нейроны, в то время как спайковые нейросети подражают работе мозга животного или человека – задействуют только активные в конкретный момент времени нейроны, что обеспечивает значительную экономию ресурсов при их обучении и использовании. Кроме того, именно спайковые нейросети, биологически правдоподобные и гибридные модели и методы обучения ИИ считаются более перспективными с точки зрения прогресса в понимании принципов работы человеческого мозга за счет перспектив их использования в когнитивных науках. В основе таких разработок лежит использование модели пирамидального нейрона, который составляет основную долю нейронов коры головного мозга человека и обучается быстрее, чем искусственный нейрон.

Исследователи группы «Нейросимвольная интеграция» Института искусственного интеллекта AIRI и студенты МФТИ создали первую в России биологически-правдоподобную вычислительную модель памяти агента, который способен эффективно оперировать в незнакомой среде под воздействием внешнего сигнала подкрепления. Например, навигироваться и искать ресурсы в лабиринтах и помещениях.

Разработанная модель агента может оперировать абстракциями состояний и действий. Это значит, что он может совершать сложные действия на базе уже известных ему простых операций. Например, научившись искать дверь в помещении, агент сможет использовать этот навык для решения более сложных задач, в то время как большинство существующих сейчас программ требует создания новой инструкции для каждой конкретной задачи. Помимо внешней мотивации (награды за успешно совершенное действие) у разработанного научной

группой агента присутствует и внутренняя мотивация. Это делает его поведение более сложным и автономным. Внутренняя мотивация обеспечивает осмысленное поведение в отсутствие внешнего подкрепляющего сигнала. Это означает, что такой агент сможет не только искать решение заданной задачи, как большинство стандартных программ, но и изучать мир вокруг себя.

Статья выполнена в рамках долгосрочного фундаментального исследования на стыке компьютерных и когнитивных наук, которое приближает ученых к созданию более самостоятельных систем искусственного интеллекта и лучшему пониманию принципов работы мозга человека и животного. Построение таких больших гибридных биологически-правдоподобных моделей и последующее их тестирование в сложных средах-симуляторах — это принципиально новое и мало исследованное направление. Подобные проекты могут помочь исследователям-когнитивистам для проверки теорий и гипотез, а созданная модель может стать одним из примеров того, как целый комплекс моделей из нейронаук может быть соединен вместе, чтобы сделать работу систем ИИ более похожей на работу человеческого мозга. Кроме того, биологическое правдоподобие структуры такого агента потребует меньше вычислительных мощностей и сделает ИИ более экономным.

«Разработанная нашей группой архитектура действительно является уникальным комплексом интересных идей из области нейронаук и когнитивных исследований. Это оригинальные реализации иерархического представления информации, внутренней мотивации и предсказания последствий событий. Основное наше достижение – научить такую комплексную систему стабильно и согласованно действовать в среде. Наша модель дает исследователям со всего мира возможность создавать еще более сложные системы, которые еще лучше имитируют работу мозга и психики человека»

Александр Панов, к.ф.-м.н., ведущий научный сотрудник, руководитель группы «нейросимвольная интеграция» Института искусственного интеллекта AIRI, заведующий лабораторией когнитивных динамических систем МФТИ

Код когнитивного агента выложен в открытый доступ, любой желающий исследователь может им воспользоваться. Полный текст статьи был опубликован в авторитетном научном журнале с рейтингом Q1.

Прочитать статью:

<https://braininformatics.springeropen.com/articles/10.1186/s40708-022-00156-6#Sec1>

Научно-исследовательский Институт искусственного интеллекта AIRI — автономная некоммерческая организация, занимающаяся фундаментальными и прикладными исследованиями в области искусственного интеллекта. На сегодняшний день более 90 научных сотрудников AIRI задействовано в исследовательских проектах Института для работы совместно с глобальным сообществом разработчиков, академическими и индустриальными партнерами.