Новые архитектуры сильного ИИ, основанные на принципах работы мозга

С. А. Шумский[1]

Показано, как в ходе биологической эволюции мозг постепенно сформировал иерархическую архитектуру глубокого обучения с подкреплением. Основываясь на этой архитектуре предложена действующая модель сильного ИИ – ADAM, способная обучаться все более сложным поведенческим навыкам по мере увеличения глубины иерархии управляющих уровней.

Ключевые слова: общий искусственный интеллект, глубокое обучение с подкреплением, иерархическая система управления.

1. Введение

Представляется логичным при создании будущего сильного ИИ «учиться у мозга», поскольку никакой другой системы, обладающей сильным интеллектом, мы не знаем. При этом, разработчиков ИИ должны интересовать не тонкости работы и взаимодействия отдельных нейронов, а способы обучения и взаимодействия друг с другом больших подсистем мозга — не физика мозга, а его схемотехника. Чтобы создать искусственную психику роботов надо подняться от современного уровня отдельных подсистем (глубоких нейросетей) на системный уровень.

2. Постановка задачи

Мы будем рассматривать интеллект, как непременный атрибут автономных агентов. Действительно, когда говорят о человеческом интеллекте, под этим обычно подразумевается некая творческая личность. Творческая — означает способность решать нестандартные задачи, с которыми она раньше не сталкивалась. А личность — означает автономность. Личность не просто исполняет чужие команды, а самостоятельно определяет свое поведение, ставит себе задачи, повинуясь своим внутренним мотивам, максимизируя свою внутреннюю функцию ценности.

[1] *Шумский Сергей Александрович* — к.ф.-м.н., директор по системам сильного искусственного интеллекта Центра компетенций НТИ Искусственный интеллект при МФТИ, руководитель лаборатории когнитивных архитектур ЦНТИ МФТИ, e-mail: serge.shumsky@gmail.com.

Shumsky Sergey — PhD, director for AGI systems, NTI Centre of competence Artificial Intelligence, MIPT, head of laboratory for cognitive architectures, e-mail: serge.shumsky@gmail.com.

Для создания сильного ИИ необходимо создать искусственную творческую личность, которой мы могли бы управлять не напрямую, а через формирование у нее соответствующих ценностей — чтобы интеллектуальный агент мог сам ставить и достигать различные цели, действуя при этом в наших интересах. Так же как мы воспитываем человеческие личности с принятыми в нашем обществе ценностями. В итоге мы должны получить агентский интеллект, обладающий свободой воли. При этом сила такого агентского интеллекта определяется его способностью достигать достаточно сложные и далекие цели, состоящие из многих этапов, т.е. его способностью к долгосрочному планированию своего поведения.

Математическим выражением всех этих требований является т.н. *обучение с подкреплением*, где целью обучения агентов при их взаимодействии с миром является стремление максимизировать поток ожидаемых будущих наград. Для этого он в процессе обучения строит свою модель функции ценности сенсомоторной информации $Q(s, a)$, стремясь, чтобы она статистически приближала суммарные будущие награды на некотором временном горизонте, регулируемым параметром дисконтирования будущего γ. Чем ближе γ к 1, тем больше горизонт планирования поведения агента, тем он дальновиднее и «умнее».

Таким образом, задачу построения сильного ИИ можно сформулировать, как умение обучаться достигать отдаленные цели с минимальным числом ошибок (поскольку каждая ошибка имеет свою цену).

3. Эволюция мозга

Рассмотрим теперь, как эта задача решалась мозгом позвоночных по мере его эволюции от рыб до человека. У всех позвоночных нервная система имеет один и тот же генеральный план иерархического управления поведением. Нижние этажи с наиболее простыми примитивами движений находятся в спинном мозге. Более сложные поведенческие инстинкты находятся выше по уровням иерархии. Нас будет интересовать верхний этаж управления в *переднем мозге*, где принимаются окончательные решения о том, что делать в той или иной ситуации, где, так сказать, находится свободная воля.

Передний мозг состоит из *коры* (или *паллиума*) и церебральных ядер — *базальных ганглий (БГ)*. Последние представляют собой главную тормозную систему мозга, и все сигналы от коры к нижним уровням иерархии проходят через внешнее и внутреннее ядра БГ в одно или два касания. В зависимости от этого какие-то сигналы тормозятся, а какие-то наоборот — растормаживаются. Таким образом, БГ являются главной инстанцией, принимающей решения, и именно во внешнем ядре БГ (*стриатуме*) и формируется функция ценности.

Очень важная обратная связь между БГ и корой через таламус образовалась при выходе животных на сушу у древних пресмыкающихся — динозавров, что позволило им и их потомкам мыслить не отдельными действиями, а вырабатывать согласованные цепочки действий. Благодаря этому динозавры могли преследовать свою добычу даже после того, как она исчезала из виду, поскольку их поведение определялись ожиданием будущих наград на некотором временном горизонте.

Оказывается, что схемотехника БГ с подкрепляющими сигналами из допаминовой системы среднего мозга реализует итерационное обучение функции ценности, используя методы динамического программирования [1], что не требует построения дополнительных предиктивных моделей (т.н. model-free learning):

$$Q(s,a) = \left\langle \sum_{t \geqslant 0} \gamma^t r(t) \right\rangle_\pi = \langle r + \gamma Q(s',a') \rangle_\pi$$

$$\Delta Q(s,a) = \alpha \left(r + \gamma Q(s',a') - Q(s,a) \right), \quad (a \ll 1)$$

Однако такой дизайн мозга обладал очень важным недостатком — очень ограниченным горизонтом планирования. Оказывается, что при model-free обучении число итераций растет как куб горизонта планирования [2]. Т.е. при горизонте планирования 100 когнитивных актов (около минуты в реальном мире), требуется порядка миллиона прохождений через каждое из состояний поведенческого репертуара! Таким образом, интеллект динозавров, их способность заглядывать в будущее, был принципиально ограничен несколькими десятками когнитивных актов.

Выход из положения со временем нашли сменившие динозавров млекопитающие и птицы. Он состоял в том, чтобы отдельно от функции ценности в стриатуме обучать еще и модель мира в коре (у птиц — в паллиуме), которую можно использовать в качестве тренажера — эмулятора реальности для обучения функции ценности в режиме офлайн [3]. Точность модели мира не зависит от горизонта планирования, так что построение модели мира оказывается существенно более простой задачей, чем обучение поведению.

Кора у млекопитающих первоначально представляла собой эпизодическую память *архикортекса* — гиппокампа и энторинальной коры. Известно, что во сне и наяву во время отдыха животные, включая людей, проигрывают в уме различные эпизоды пережитого за день, тренируя тем самым свою модель поведения.

Эпизодическая память архикортекса со временем была дополнена гораздо более емкой категориальной памятью *неокортекса*, способной к обобщению, выделению типовых категорий, на которых и основывается наша картина мира. В неокортексе также имеется специальная система

обучения модели поведения, когда мозг не занят решением конкретных задач. Это т.н. *дефолтная система*, которая мысленно проигрывает различные воображаемые ситуации, тренируя модель поведения. Возможно это и есть самый древний способ использования мозгом способности неокортекса генерировать новые ситуации взаимодействия с реальным миром.

Простейшая модель неокортекса — *самоорганизующиеся карты признаков* — представляет собой двумерную нейросеть со взаимной конкуренцией между ее элементами [4]. Действительно, как экспериментально показал Маунткасл [5], неокортекс состоит из кортикальных колонок, каждая из которых является детектором определенного признака. Типичная колонка содержит порядка 10^4 нейронов, и ее размер определяется разбросом дендритов пирамидальных нейронов коры — около 300 мкм. Эти колонки объединены в более крупные модули — гиперколонки, размер которых определяется длиной аксонов тех же пирамидальных нейронов — около 1 мм. Внутри гиперколонки все колонки конкурируют друг с другом, т.е. тормозят активность соседей, в результате чего победителем в гиперколонке всегда оказывается какая-то одна колонка, наиболее сильно откликающаяся на данный внешний сигнал. Колонки, таким образом, формируют прототипы, т.е. *категории* сигналов и различают к какой категории принадлежит входящий сигнал. Поэтому кора является хранителем всех наших знаний о внешнем мире, выраженных на языке категорий, объединяющих сходные по каким-то признакам объекты.

Важно, что размер гиперколонок невелик — они объединяют порядка нескольких десятков колонок (число букв в типичных алфавитах порядка 30). Значит гиперколонки могут различать не так много признаков, т.е. выделяемые ими категории очень общие. Одна гиперколонка не способна, например, определить лицо человека. Однако несколько гиперколонок могут. Так, k гиперколонок, содержащих каждая по N колонок, могут распознать N^k образов, если они разбивают входные сигналы по разным признакам, т.е. работают с разными подпространствами входного сигнала.

Таким образом, небольшая область из нескольких гиперколонок, размером с горошину, может распознавать десятки тысяч образов любой природы. Например, в височной доле у нас есть небольшая зона примерно такого размера, отвечающая за распознавание лиц. При этом, процесс распознавания лиц в ней можно сопоставить составлению фоторобота путем комбинации небольшого числа типовых элементов лица. И таких модулей размером порядка 0.1 см^2 в человеческом мозге может быть порядка десятка тысяч.

Мы знаем, что наша кора умеет распознавать не только пространственные, но и временные образы, т.е. последовательности признаков, например, слова или мелодии, т.е. последовательности букв или нот. Эту задачу могут выполнять гипотетические рекурсивные модули, в которых гиперколонку, распознающую входные символы, окружают гиперколонки с локальными связями, «смотрящие» на нее и друг на друга. Такие модули способны преобразовывать временную последовательность символов в пространственный паттерн активности, причем модуль из k гиперколонок способен запоминать N^k последовательностей длины k.

Можно предположить, что именно преобладание таких рекурсивных модулей, т.е. специализация на работе с последовательностями, отличает мозг приматов от мозга остальных млекопитающих. Действительно, как выяснилось относительно недавно [6], у приматов количество нейронов в неокортексе растет пропорционально массе неокортекса (видимо, благодаря преобладанию локальных связей в рекурсивных модулях). Тогда как у остальных млекопитающих в коре преобладают длинные связи, поскольку число нейронов растет как масса неокортекса в степени 2/3. Поэтому с ростом массы коры приматы «умнеют» гораздо быстрее остальных млекопитающих. Так что уникален не сам по себе человеческий мозг, а мозг всех приматов. Люди же просто воспользовались этим преимуществом, максимально нарастив массу мозга (за счет уменьшения массы кишечника, видимо, в связи с освоением огня и переходом на вареную пищу).

Увеличение размера коры и умение работать с последовательностями позволило приматам использовать кору не только для обучения БГ в режиме офлайн, но и для планирования поведения в режиме онлайн. В результате кора высших приматов — и особенно человека — научилась «держать мысль», т.е. не просто грезить, а сознательно просчитывать варианты развития событий и сравнивать их между собой, запоминая по ходу дела результаты различных вариантов — используя механизм *рабочей памяти*. Сформировалась новая *центральная исполнительная система* планирования достижения целей в «виртуальной реальности» нашей модели мира.

Причем, что важно, планирование в нашем мозге происходит иерархически — сначала вырабатывается крупномасштабный замысел, который затем прорабатывается и воплощается в жизнь, адаптируясь к ситуации по мере развития событий. Соответственно, у нас в мозге иерархически организовано и взаимодействие неокортекса с БГ: *кортико-стриарная система* имеет четкую иерархическую структуру — состоит из типовых модулей с единообразной схемотехникой. Каждый такой модуль захватывает кусочек коры, обучаемый на ошибках своих предсказаний, и часть базальных ганглий, обучаемых допаминовыми подкрепле-

ниями из среднего мозга. С точки зрения автора эта особенность нашего мозга является ключевой для построения сильного ИИ, поскольку позволяет строить иерархические планы достижения очень далеких целей. И чем больше становилась площадь коры, тем более далекие и сложные планы были доступны нашим предкам, создавая давление отбора в пользу дальнейшего увеличения размеров мозга.

4. Конструирование искусственной психики роботов

Покажем теперь, как понимание описанной выше логики развития архитектуры мозга можно использовать при конструировании сильного ИИ, т.е. искусственной психики роботов.

Если отвлечься от биологических деталей, архитектура управления мышлением и поведением в нашем мозге представляет собой иерархию управляющих уровней, устроенных единообразно и осуществляющих управление каждый на своем масштабе времени. При этом, более низкие уровни ищут пути реализации планов, спущенных из более высоких уровней [7]. На основе такой архитектуры в лаборатории когнитивных архитектур МФТИ создан программный код ADAM (Adaptive Deep Autonomous Machine, [8]) — действующий прототип искусственной психики, способный планировать свое поведение на многих масштабах времени. По мере накопления опыта взаимодействия с внешней средой, ADAM наращивает число своих уровней, обучаясь целенаправленному поведению на все более долгих временных масштабах. Разработанная и отлаженная в рамках проекта ADAM архитектура сильного ИИ сможет использоваться в самых разных сервисах и продуктах, требующих креативного машинного мышления.

5. Заключение

Резюмируя, в данной работе описан подход к созданию сильного ИИ, основанного на принципах работы мозга. Впервые в мире создана действующая модель искусственной психики с иерархической архитектурой глубокого обучения с подкреплением, способная самостоятельно планировать свое поведение на сколь угодно большом временном горизонте. Этот подход может быть использован при создании операционных систем автономных роботов, способных в процессе своей жизни накапливать опыт решения самых разных задач.

Список литературы

[1] R. Bellman, “The theory of dynamic programming”, *Bulletin of the American Mathematical Society*, **60**:6 (1954), 503-515.

[2] M.Wainwright, “Stochastic approximation with cone-contractive operators: Sharp l_∞-bounds for Q-learning”, *arXiv preprint arXiv:1905.06265*, (2019) May 15.

[3] R. Sutton, “Dyna, an integrated architecture for learning, planning, and reacting”, *ACM Sigart Bulletin*, **2**:4 (1991), 160-163.

[4] T. Kohonen, “Self-organized formation of topologically correct feature maps”, *Biological cybernetics*, **43**:1 (1982), 59-69.

[5] V. Mountcastle, “The columnar organization of the neocortex”, *Brain: a journal of neurology*, **120**:4 (1997), 701-722.

[6] S. Herculano-Houzel, *The human advantage: A new understanding of how our brain became remarkable*, MIT Press, 2016.

[7] Шумский С.А., “Глубокое структурное обучение: новый взгляд на обучение с подкреплением”, *Сборник научных трудов XX Всероссийской научной конференции Нейроинформатика-2018, Лекции по нейроинформатике*, 2018, 11-43.

[8] Шумский С.А., Басков О.В., “Программный Агент глубокого иерархического обучения с подкреплением ADAM Deep Control”, *Государственная регистрация программ для ЭВМ*, 2021, RU 2021660307.

Brain-inspired new AGI architectures

Shumsky S.A.

We consider how in the course of biological evolution the brain gradually formed a hierarchical architecture of deep reinforcement learning. Based on this architecture, a working AGI model, ADAM, is proposed, capable of learning more and more complex behavioral skills as the depth of the hierarchy of control levels increases.

*Keywords:*artificial general intelligence, deep reinforcement learning, hierarchical control systems.

References

[1] R. Bellman, “The theory of dynamic programming”, *Bulletin of the American Mathematical Society*, **60**:6 (1954), 503-515.

[2] M.Wainwright, “Stochastic approximation with cone-contractive operators: Sharp l_∞-bounds for Q-learning”, *arXiv preprint arXiv:1905.06265*, (2019) May 15.

[3] R. Sutton, “Dyna, an integrated architecture for learning, planning, and reacting”, *ACM Sigart Bulletin*, **2**:4 (1991), 160-163.

[4] T. Kohonen, “Self-organized formation of topologically correct feature maps”, *Biological cybernetics*, **43**:1 (1982), 59-69.

[5] V. Mountcastle, "The columnar organization of the neocortex", *Brain: a journal of neurology*, **120**:4 (1997), 701-722.

[6] S. Herculano-Houzel, *The human advantage: A new understanding of how our brain became remarkable*, MIT Press, 2016.

[7] Shumsky S.A., "Deep structural learning: a new look at reinforcement learning", *Collection of scientific papers of the XX All-Russian scientific conference Neuroinformatics-2018, Lectures on neuroinformatics*, 2018, 11-43 (In Russian).

[8] Shumsky S.A., Baskov O.V., "Software Agent with deep hierarchical reinforcement learning ADAM Deep Control", *State registration of computer programs*, 2021, 2021660307 (In Russian).