Чернов Иван ИУ7-11М

----------------------------------------------------------------------------------------------------------------

Пределы машинного обучения, мини-эссе

С каждым годом в обществе возрастает уровень беспокойства на счет того, что новые технологии лишат некоторую долю населения работы и их занятий. Кризис автоматизации заставляет многих нервничать и переживать за своё будущее.

Однако важно понимать, что у машинного обучения имеются свои пределы и ограничения. В выступлении рассматривалось два подхода к созданию искусственного интеллекта - дедуктивный подход и индуктивный подход. Оба подхода имеют значительные ограничения.

Дедуктивный подход предполагает серьезную подготовительную работу: необходимо привлечь экспертов, создать базы знаний, выделить правила. Индуктивный подход -- как можно догадаться из названия -- предполагает рассмотрение частных случаев и переход от частных случаев к общим выводам.

Первый подход обычно именуют экспертными системами, второй -- собственно, машинным обучением. У экспертных систем имеется ряд недостатков, характерных именно для них. Так, например, их необходимо постоянно модифицировать. Если человек способен обучаться и адаптироваться, то экспертные системы являются жесткой и негибкой системой. Экспертные системы не способны к творчеству и ограничены сенсорно. Они представляют собой отличных помощников, но не претендуют на место человека.

Настоящую тревогу у людей вызывает, как правило, именно машинное обучение. В новостях то и дело появляются анонсы новых достижений данной области -- нейронные сети умеют создавать изображения, писать тексты, генерировать музыку, водить автомобили, переводить на лету, проходить игры, словом, делать то, что традиционно делал человек. Несмотря на пугающий успех нейронных сетей, у них тоже есть ряд ограничений и пределов.

Во время выступления приводился пример с обучением искусственного интеллекта на играх Атари. Изначально нейросеть не обладала какими-либо знаниями о правилах игры, однако через несколько сотен итераций она вырабатывала некоторую стратегию игры. Нейросеть успешно справилась с большей частью игр, во многим превзойдя человека, однако была одна игра, с которой сеть не справилась.

К провалу привело то, что нейросеть полагалась на моментальные вознаграждения в виде очков, однако в данной игре было необходимо совершать довольно длинную последовательность действий, прежде чем игра вознаграждала игрока баллами. Нейронные сети плохо справляются с действиями, вознаграждение за которые наступает в долгосрочной перспективе.

Полностью нейронные сети нас пока не заменят. Как минимум потому, что в реальной жизни нет баллов и последствия за действия может наступать через значительное время. Также нейронным сетям зачастую не хватает фоновых знаний. Там, где человек поймет отсылки или догадается до чего-либо, нейронная сеть пройдет мимо. Нейронные сети не обладают набором знаний как таковых, только набором правил, необходимых для решения задачи.

Еще одним серьезным ограничением нейронных сетей является отсутствие у них здравого смысла. Здравый смысл представляет из себя широкий спектр неписаных предположений и эмпирических правил, неявных знаний о мире. Здравый смысл позволяет нам читать между строк, догадываться о значениях, избегать не очевидных опасностей.

Несмотря на выдающиеся успехи в области искусственного интеллекта, говорить о замене человека искусственным интеллектом пока не приходится. Пока решения нейронных сетей больше напоминают голубиные предрассудки, чем настоящее решение задачи.

---------------------------------------------------------------------------------------------------------------------------

Обобщающий ИИ

В продолжение темы ограничений искусственного интеллекта.

Несмотря на то, что и человек, и искусственный интеллект способны переходить от частных случаев к общим, бесспорно можно сказать, что в этой гонке человек уверенно лидирует. Способность человека к генерализации превосходит способность машины. Человек способен обучаться на нескольких примерах, а иногда строить суждения без предварительного обучения. Например, ребенок, впервые увидевший дельфина, способен распознать в нем живое существо.

В отличии от ребенка, искусственному интеллекту требуется огромный набор данных, зачастую предварительно размеченный. Недостаточная выборка может привести к катастрофическим последствиям. К примеру, распознавание лиц в камерах айфона плохо распознавало лица меньшинств, а лица представителей азиатского этноса считало одинаковым. Это привело к скандалам и обвинениям в расизме. К тому же, искусственный интеллект легко обмануть -- там, где человек с легкостью узнает объект на изображении, искусственный интеллект испытывает сложности.

Что позволяет человеческому мозгу быть настолько лучше? Исследователи полагают, что дело в петлях обратной связи. Именно они позволяют нам так хорошо определять предметы. Петли обратной связи оказываются отличными помощниками в устранении помех.

В примере испытуемым предложили классифицировать объект. Объект показывался двумя способами -- быстро (в этом случае у испытуемых не успевала возникнуть петля обратной связи) и достаточно долго. В первом случае испытуемые часто ошибались, во втором -- смогли правильно классифицировать объекты.

В выступлении приводится новая архитектура нейронных сетей с петлями обратных связей, позволяющая обрабатывать помехи, и исследования работы.