О ПОНИМАНИИ В ИСКУССТВЕННОМ ИНТЕЛЛЕКТЕ

Искусственный интеллект сейчас представляется различными системами, но о понимании можно говорить только в системах компьютер-человек, что представлено диалоговыми системами Искусственного Интеллекта (ИИ). И сама тема понимания в ИИ сводится к нескольким аспектам диалогового взаимодействия искусственного агента с человеком:

1. Порождаемых диалоговой системой тексты отвечают «здравому смыслу»
2. Ответы системы соответствуют контексту диалога и ожиданиям человека
3. Понимание целей, намерений и смысла высказываний человека

Понимания смысла в полной мере нельзя отнести к теме понимания контекста диалога, так как смысл высказывания собеседника может быть по-разному интерпретирован, и какой интерпретации должно соответствовать состояние понимания, не ясно. Можно ли «ошибки» по мнению собеседника (человека) интерпретировать как иное понимание смысла выражения системой? В большей степени понимания смысла относится к пониманию намерений и целей высказывания, а это отдельная тема *theory of mind*.

«Здравый смысл» как критерий понимания можно интерпретировать точнее. В общем смысле это соответствие ответа картине мира, что поддается проверке. И на сегодня это является лучшим критерием понимания искусственными агентами, такими как диалоговые боты, контекста диалога. Но пока в этом боты не демонстрируют успехи.

Релевантный ответ является самым простым критерием понимания ботом собеседника (человека). Но этот критерий легко «подделать», что ни раз демонстрировалось участниками Премии Лёбнера. Он достигается закладыванием большого числа вариативных шаблонов ответов на распознаваемые нейронной сетью «интентов». Это трудно назвать пониманием. Но и успехи таких ботов скромны – смешанные интенты они распознают крайне плохо. Один вопрос вне тем шаблонов и система проваливается. Это легко проверить на таких ботах как Алиса от Яндекс и Siri от Apple. *Можно сказать, что знания мира у таких систем фрагментарны.*

Еще один путь – построение онтологий, покрывающие все знания о мире, чтобы иметь ответить на вопросы. Это реализуется через возможность идентифицировать вопрос на ветке онтологии и построить верный ответ на основе заложенных в ней знаний. Такой подход применялся таких проектах как СУС и АВВУ Comreno. Теперь можно констатировать, что такой способ решения проблемы «понимания» провалился. Десятилетия ручного построения онтоинженерами таких онтологий не позволили построить такую, чтобы она не проваливалась даже от простых детских вопросов. И проблемы в основном связаны с метафоричностью языка и омонимией, которые невозможно непротиворечиво заложить в древовидную «логичную» структуру онтологий. К этим же попыткам можно отнести проекты FrameNet (Филмор), заключающиеся в построении сценариев на основе семантического фрейма для отражения фактов мира. Но по тем же причинам проект не смог получить развитие. *Можно сказать, что такие системы не обладают учитывать экономность языка как способность ограниченным словарным запасом описывать большое число концептов через механизм метафоры.*

С 2010 года больше упор делается на нейронные сети, но они ограничены тем, что могут делать предположения только в рамках той картины мира, которая присутствовала в обучающей выборке. Такие сети как GPT-3 (февраль 2019) в проекте OpenAI, построенными на архитектуре Transformer (генеративная сеть, обученная на миллиардных по объему слов текстах), порождают грамматически правильные «человеческие» тексты, совершенно лишённые смысла, что говорит о полном непонимании ими здравого смысла как в примере ниже. Можно сказать, что такие системы могут сгенерировать высказывание по любому вопросу, встречающемуся в миллиардных текстах, но они не могут строить из них факты «здравого смысла», соответствующие картине мира.

Есть комбинированные подходы как COMET, основанным Yejin Choi (1), в котором для дообучения предобученной языковой модели GPT-2 применили базу знаний фактов мира Atomic. В результате порождение сеть стала порождать достоверно больше новых правдоподобных фактов, отсутствующий и в Atomic, и в обучающей выборке GPT. Но успехи такого проекта пока тоже скромны.

Интерес представляют и системы DeepMind, которые в дополнение к нейросети имеют внешнюю память фактов (или опыта), что позволяет им обучаться без учителя «правилам игры», просто проявляя активность в среде и записывая результат. И таким способом обучаться даже в игре друг с другом, что позволило обыгрывать игроков-людей даже в таких играх как Го. Сейчас это считается мейнстримом в построении агентов, «понимающих мир». Но архитектура такой самообучающейся системы не позволяет масштабировать ее на какую-то более сложную реальность, чем игра в черные-белые камешки или примитивную компьютерную игру Atari. *Можно сказать, что такие системы реализуют свои знания не при помощи использования знаний для построения новых знаний, чтобы экономить ресурсы системы.* Поэтому им требуется слишком много ресурсов для обучения даже в бедных средах.

Что же тогда можно называть «пониманием» у искусственных систем с прагматической точки зрения? Обыденный ответ – агент должен обладать знанием. Но, исходя из сделанного анализа выше, это неверный ответ.

Резюме

Понимание системой ИИ означает ее способность ДОСТРАИВАТЬ правдоподобные гипотезы о картине мира по отрывочным знаниям фактов из этого мира. И для экономии система должна уметь для описания бесконечного числа фактов использовать ограниченный язык, что достигается механизмами, подобными метафоре. На сегодняшний момент, правда, этот механизм не известен настолько, чтобы его можно было воплотить в программном коде. Имеющиеся концепции метафоры не конкретны алгоритмически, например, концептуальная метафора или бленды. К ним пока не применима математика, но работа автора ведется именно в этом направлении

По мнению автора такое достраивание и есть основной критерий способности искусственной системы к пониманию. Когда «картина мира» ограничена, например, в шахматах, мы способны явно заложить алгоритмы продукции знаний, то есть возможных ходов, чтобы шахматы ориентировались в любой, даже не встречавшейся ранее расстановке фигур. Но как это сделать в реальном мире, где правил на много порядков больше, пока не известно, что и составляет основное направление исследования автора.

Библиография

Common Sense Comes Closer to Computers, Quantamagazin, April 30 2020,
<https://www.quantamagazine.org/common-sense-comes-to-computers-20200430/>