

Проблема понимания в системах искусственного интеллекта

Максимов В.Ю., ИПМ им. М.В. Келдыша РАН
vadimmax2000@mail.ru

Клышинский Э.С., ИПМ им. М.В. Келдыша РАН
klyshinsky@mail.ru

Антонов Н.В., НИУ Высшая школа экономики, МИЭМ

Аннотация

Стремительное развитие искусственного интеллекта в последние годы заставило по-новому взглянуть на проблемы обработки естественного языка, и особенно – на проблему понимания. В статье дается обзор широкого круга проблем, связанных с пониманием в контексте использования систем искусственного интеллекта. Показано, что искусственному интеллекту стали открываться такие области, которые ранее традиционно исключали всякое вмешательство машины. Раскрытие проблемы понимания на сегодняшний день представляется ключевым для дальнейшего расширения области использования искусственного интеллекта, а также для разграничения двух типов искусственного интеллекта: в слабом и сильном смысле.

1 Введение

Успехи последних лет в области искусственного интеллекта (ИИ) существенным образом изменили взгляд на возможности использования ИИ в самых неожиданных областях. Едва ли кто-то мог уверенно предположить, что развитие ИИ будет продвигаться такими темпами. В 1996 году компьютер впервые одержал победу в шахматной игре у чемпиона мира, а через 10 лет уже не осталось людей, способных обыграть машину в шахматы. Прошло еще 10 лет, и совсем недавно появилось сообщение, что в конце января 2016 года система ИИ на основе нейронной сети впервые обыграла профессионального игрока в го [1]. Обучение сети производилось на 30 миллионах ходов реальных партий живых людей. В отличие от шахмат, число возможных комбинаций на доске здесь в 10^{100} раз больше, а общее число позиций превышает число атомов во Вселенной. Точное число комбинаций рассчитано 20 января 2016 Джоном Тромпом и опубликовано в сети [2].

За разработку системы искусственного интеллекта для игры в го взялась команда из DeepMind (принадлежит Google). Система

получила название AlphaGo. Для ее реализации была предложена схема на нейронных сетях с достаточно сложной архитектурой и алгоритмом, сочетающим в себе разные формы обучения (обучение с учителем и обучение с подкреплением), анализ позиций мастеров игры в го и позиций, полученных в результате автономного моделирования ходов [1].

В основе алгоритма лежит взаимодействие сети оценок (Value network) с сетью реализации политики (Policy network). Задача многослойной сети оценок – сужение поля возможных ходов, сеть реализации политики помогает выбирать оптимальные ходы в наиболее разветвленных деревьях.

Система длительное время обучалась в процессе игры с живыми профессиональными игроками в го, затем обучение продолжалось при обработке данных онлайн-матчей и игре системы с собой. В результате был достигнут уровень международного мастера (третий дан). В январе 2016 года AlphaGo обыграла со счетом 5:0 чемпиона Европы по игре в го, имеющего второй дан.

Это событие можно рассматривать как знаковый этап в развитии искусственного интеллекта. Дело не только и не столько в сложности задачи, сколько в ответе на вопрос о классе задач, которые можно поручать машине, а также о принципиальном отличии человеческого разума от ИИ. Долгое время бытовало мнение, что задачи подобной сложности решаются исключительно с помощью интуиции. И вот, на наших глазах одно за другим рушатся представления о монополии человеческого разума.

Естественно, на сегодняшний день речь не идет о состязании человека и машины в распознавании образов или сочинении стихов, вопрос стоит о принципиальной реализуемости того или иного круга задач системами искусственного интеллекта. В связи с этим на новом уровне приобретает актуальность давний вопрос: в чем заключается отличие чело-

века от автомата, пусть даже совершенного и биологического.

2 Прорыв в области ИИ

Активный прорыв в распространении идей ИИ на новые области, в том числе нехарактерные для него в прошлом, обусловлен рядом факторов. Среди важнейших, очевидно, следует назвать увеличение мощности процессоров и появление графических процессоров, позволяющих распараллеливать вычисления, что принципиально для построения нейронных сетей. С появлением облачных технологий стало возможно обрабатывать большие объемы данных. В свою очередь, перечисленные особенности подстегнули разработку новых алгоритмов и архитектур нейронных сетей. Многие современные идеи вполне могли быть реализованы еще 20 лет назад, но многодневные вычисления не давали возможности развернуться с экспериментированием. Многие решения начинают демонстрировать свою эффективность лишь на больших объемах данных и при длительном обучении.

На сегодня системы ИИ распознают «капчи», объекты в видео потоке в режиме реального времени (катер у причала, машина, лицо человека, здание, фонарный столб...), раскрашивают черно-белые фильмы. Нейронные сети обучают распознавать юмор на картинках, доверяют им выбирать живописные маршруты для путешествий и даже определять победительниц в конкурсе красоты. Экспертные системы доказали свою эффективность при использовании в коммерческих проектах. Следует упомянуть о том, что с помощью систем ИИ доказываются математические теоремы, определяется степень новизны в патентах, решаются научные задачи. В частности, известная теорема о четырех красках, сформулированная в 1852 году, была доказана с помощью компьютера в 1976 году.

Сегодня система искусственного интеллекта уже выступает автором научной теории. В работе [3] описывается, как с помощью системы ИИ решена задача, поставленная в биологии 120 лет назад. Загадка регенерации червей вида *Planarian* беспокоила биологов давно. Наконец, появилась мысль привлечь к решению давней загадки машину. Система ИИ моделировала различные варианты сетей, сформированных генами и белками червя, сопоставляя их с результатами эксперимен-

тов. При достижении частичного соответствия СИИ вносила в генетическую сеть случайные изменения и возобновляла процесс оптимизации. В результате множества итераций была получена сеть генетических связей, полностью соответствующая результатам практических исследований.

Вне моделирования не осталась даже такая область, как человеческая этика. С помощью системы ИИ "Scheherazade system" генерировались тексты из краудсорсинговой платформы Amazon Mechanical Turk по признаку семантической схожести. Затем система определяла хронологию событий. Из множества событий создавалось дерево возможных бытовых историй, с какими мы сталкиваемся в жизни. Задачей виртуального агента была реализация заданного действия (получить лекарство в аптеке). Обучение велось так, что наибольший вес давался по наиболее характерным (т.е. наиболее часто наблюдаемым) поступкам людей в той или иной ситуации. В итоге виртуальный агент при выполнении задания выбирал не самый короткий путь, подразумевающий использование действий, противоречащих этике (например, украсть лекарство), а тот, который используется людьми, основывающимися свои поступки на этических принципах[4].

Все перечисленные примеры показывают, что мы действительно вступаем в новую эру, где искусственный интеллект будет отвоевывать всё новые и новые области. Наибольший акцент при этом приходится на естественный язык. Язык выполняет не только коммуникативную функцию, это в первую очередь средство мышления. Поэтому не удивительно, что основной акцент в направлении развития систем ИИ приходится именно на язык. Машина и человек должны научиться понимать друг друга. Термин «понимание», который на первый взгляд не требует особых комментариев для человека, в контексте соседства с ИИ сполна обнаруживает всю свою сложность и нетривиальность.

3 Проблемы понимания в контексте систем ИИ

«Понимание» плохо поддается формализации для реализации в машинных приложениях. Ко всему прочему, он скрывает в себе основную загадку когнитивистики о природе человеческого мышления, на которую до сих пор нет ответа. Тем не менее, попытки при-

близиться к решению проблемы установления понимания между естественным и искусственным интеллектом, очевидно, в ближайшем будущем будут все более интенсивными. Попробуем посмотреть, что делается в этом направлении, ни в коем случае не претендуя на полноту освещения этой темы.

3.1 Конкретизация "понимания"

Задачи конкретизации термина «понимание» в полном объеме возникли перед педагогами, т.к. именно в системе образования реализуется задача обучения пониманию. Одной из основ современной педагогики является таксономия Блума, предложенная в 1956 году [5]. В соответствии с ней, задачи образования распределяются по трем областям: когнитивной, аффективной и психомоторной. Методики постоянно совершенствуются, и постоянно появляются новые, где обучение пониманию раскрывается полнее. В частности, в качестве альтернативы таксономии Блума Wiggins and McTighe предложили "6 граней понимания" (6 Facets of Understanding) [6]. По их мнению, термин «понимать» раскрывается в следующих навыках:

- объяснять
- интерпретировать
- применять
- различать перспективу
- сопереживать
- иметь внутренне понятие

Как видно, научить машину более или менее полному «пониманию» непросто. Поскольку обучение и общение ведется посредством естественного языка, в среде искусственного интеллекта возник вопрос о моделировании языкового общения в искусственной среде.

3.2 Моделирование эволюции языка

Происхождение языка до сих пор вызывает дискуссии, но совсем недавно нейробиологи подтвердили теорию «врожденной грамматики» Хомского. [7]. С точки зрения ИИ важно понимать эволюционную составляющую, т.к. в механизме становления языка раскрываются фундаментальные процессы мышления. Первые попытки моделирования процесса зарождения языка предприняты в эволюционной лингвистике. Здесь особенного внимания заслуживают эксперименты с созданием искусственной грамматики, ориентированной на конкретное применение (Usage-

Based Grammar). В отличие от традиционной порождающей грамматики Хомского, в ней невозможны предложения, невыполнимые с точки зрения смысла, такие как "мысль ест насекомое". Более подробно эта тема разработана в работах Л. Стилса, посвященных Fluid Construction Grammar (FCG) [8]. FCG была разработана для экспериментов по взаимодействию «человек – робот» и «робот-робот». В требованиях грамматики учитываются не только лексические и синтаксические проблемы, но также семантика и «функциональность» (т.е. возможность практического применения) [9]. В системе «робот-робот» грамматика строится автоматами на структурах со спаренными свойствами. В числе параметров – тип, имя и вес. Грамматика возникает эволюционным путем в системе взаимодействующих агентов в результате взаимодействия этих агентов, причем правила используются в обе стороны: как для синтаксического разбора, так и для построения сообщения.

В моделировании участвуют агенты, которые наделяются функциями «слушающего» и «говорящего». Изначально в системе словари различных агентов не совпадают между собой. «Говорящий» с помощью старается заставить «слушающего» понять сообщение, передаваемое посредством слов из его словаря. У «слушающего» в словаре содержится ряд слов, значения которых остаются не определены. Когда при взаимодействии значение слова с неопределенным значением задается самой ситуацией, слово переходит в разряд принятых в язык, и с тех пор используется обоими агентами [10].

При использовании одной тысячи агентов, объем словарей у каждого агента на начальной стадии эволюции стремительно растет. В проведенном эксперименте число слов в системе превышало 10 000, а число различных слов для обозначения одного и того же объекта достигало 500. После 50 тысяч взаимодействий между агентами для названия объекта осталось одно единственное слово.

Данный эксперимент можно рассматривать не только в качестве модели становления значений в языке, но и как эксперимент по достижению понимания в среде автоматов с помощью языка, создаваемого самими автоматами.

3.3 Моделирование самосознания

В робототехнике посредством экспериментов с человекообразными роботами изучается *привязка машинного понимания к телу машины*. Для обучения искусственного интеллекта движению и действию в окружающей среде, а также для «интеллектуального распознавания» движений человека, робот должен иметь модель тела и возможность обучаться. Чтобы проверить, может ли робот узнать что-либо о своем собственном теле, был создан робот Мюон (Muon). Таким образом, положено начало исследованию самосознания роботов. Авторы эксперимента уверены, что их методика позволит также изучать роль тела в возникновении естественного языка у человека. Искусственный интеллект робота настроен на обучение связям между зрительным образом и движением собственных двигателей [11].

Выбор языка для команд не принципиален, команды можно генерировать практически на любых языках. С помощью команд определяются действия, которые тело должно осуществить самостоятельно (например, «встать») или с привлечением посторонних объектов (например, «поднять красный кубик»). Языки, ориентированные на действие, предполагают систему навыков, объединяющих в себе систему знаний и моторные действия, в разных областях, в частности: способность выполнять действия и распознавать, какие действия выполнены другими (так называемая проблема зеркала); способность определять, какие объекты должны принимать участие в том или ином действии (например, при команде «поднять красный кубик», в действии, кроме самого робота, принимает участие красный кубик). При этом система должна понимать, какая роль отводится объектам в действии, например, должен ли объект совершить какое-то действие сам или действие нужно совершить над этим объектом, является ли объект прямой или косвенной целью (как, например, в ситуации, когда робот получает команду поставить красный кубик на зеленый).

В процитированной работе рассматриваются эксперименты с моделированием эволюционного возникновения и становления языка. В эволюционной игре с участием реальных человекоподобных роботов, исследуются конкурентные процессы в поведении и формировании языка, а также возможности подобного системного подхода самого по себе.

3.4 Антропомиметический робот

Еще один подход к изучению возможностей понимания в системе «человек-робот» осуществлен проф. Оуэном Холландом (Owen Holland), создавшим первый в мире антропомиметический робот (робот, разработанный, чтобы подражать человеческому движению). О.Холланд убежден, что исследовать возникновение естественного языка у людей невозможно без привязки языка к телу. С построением антропомиметических роботов стали возможны качественно новые эксперименты в эволюционной лингвистике и выявлении свойств искусственного интеллекта, которые могут быть в каком-то смысле приравнены к когнитивным способностям человека. В частности, в искусственный интеллект робота анализирует не только жесты человека, но и человеческую мимику.

Холланд отмечает, что для большинства человекоподобных роботов их человекоподобность весьма условна. Для роботов, которые призваны помогать в понимании образа мыслей человека и его действий, требуется несравненно большая степень взаимопроникновения (inspiration) и подобия. Его робот, также как и «Мюон» Стилса, ориентирован на распознавание и подражание движениям, но если роботы Стилса анализировали свое собственное отражение в зеркале или движения другого робота, имеющего аналогичную наружность, то интеллект антропомиметического робота настроен на распознавание движений человека и выявление аналогий между движениями человека и управлением собственным «телом». В частности, рука робота имеет кисть, которая позволяет воспроизвести ответное пожатие руки. Усилия «мышц» откалиброваны в соответствии с человеческими, и искусственный интеллект робота уже обучен осуществлять ответное рукопожатие.

В статьях Холланда [12, 13] излагаются требования к созданию антропомиметических роботов, не только повторяющих человеческую форму, но также имитирующих биологические структуры и функции, которые позволяют роботу подробно копировать движения человека на уровне управления конкретными мышцами. В статье дается описание устройства такого робота и его функционирование на начальных стадиях. Сочетание реалистично выполненного скелета, упругих связок, имитирующих мышцы, и ко-

нусной системы обзора дает уникальную возможность не только исследовать на новом уровне когнитивные аспекты, связанные с человеческим телом, но и выйти на новый уровень понимания проблем, которые человеческий мозг вынужден решать в сфере восприятия, моторики и движения человеческого тела.

4 Понимание при автоматической обработке текста

Предельной задачей искусственного интеллекта является понимание естественного языка. При этом в качестве ввода и вывода может использоваться как речь, так и текст на одном из естественных языков, представленный в письменной форме. Соответственно, обработка текстов на естественном языке (Natural Language Processing, NLP) распадается на две большие задачи: обработка речи и обработка текста.

Без решения задачи понимания текста обработка речи имеет прикладной характер, поэтому основной акцент все же ставится на понимании текста. Поэтому в последнее время появляется все больше работ, где вместо привычного термина «автоматическая обработка текста» используется термин «понимание». Основы машинного понимания текста (Natural Language Understanding, NLU) на современном уровне изложены, например, в монографии Е. Овчинниковой [14].

Задача машинного понимания текста делится в настоящее время разделяется на две области:

i) автоматическую обработку текста (NLP), в процессе которой сплошной поток символов, поступающий в машину, приобретает структуру текста, построенного в соответствии с законами естественного языка.

ii) представление знаний (Knowledge Representation, KR), т.е. отображение входной текстовой информации на естественном языке в форме, подходящей для дальнейшей машинной обработки.

После решения проблемы машинного понимания наступает черед генерирования выходной информации на естественном языке (Natural Language Generation, NLG). Последняя задача, опять же, носит вторичный характер, т.к. она обязательна отнюдь не во всех областях: к примеру, при формулировке задачи на естественном языке в робототехнике на выходе может быть конкретное действие.

К тому же состояние дел в области NLG на сегодняшний день обстоит намного лучше: задача акустического вывода текста несравненно легче задач понимания.

4.1 Стандартная схема автоматической обработки текстов

В задачах понимания содержания обработка текста разбивается на следующие стадии:

1. *Лексический (графематический) анализ.* На этом этапе машина преобразует набор символов в текст, разделенный на слова, предложения и параграфы. Данный этап обработки письменного текста в целом освоен для большинства широко распространенных естественных языков, хотя для конкретных естественных языков имеют место те или иные специфические проблемы.

2. *Морфологический и синтаксический анализ.* Модуль синтаксического анализа выявляет грамматические связи между словами. Для английского языка эта задача на сегодня решена на хорошем уровне. Часто в синтаксический анализ включается предсемантическая обработка, фильтрующая варианты, не несущие смысловой нагрузки.

3. *Семантический Анализ.* На этом этапе обработки происходит извлечение из текста нужных значений. Отдельные слова привязываются к определенным значениям в базе знаний или онтологии и получают дополнительные смысловые связи, определяющие взаимодействие между отдельными словами или частями предложения. Например, существительное «терминал» может означать устройство для ввода или вывода информации, часть аэропорта, часть порта, где происходит сортировка грузов, комплекс сооружений для накопления и сжижения газа. Специальные термины могут содержать свой ряд значений, включая специфический для данной области «жаргонный» ряд. К примеру, в контексте анализа предложений «терминал» может означать конечный узел в древовидном графе предложения. Семантический анализ должен также снимать семантическую омонимию там, где это возможно. Например, в предложении «доклад с юмором обсуждали» семантический анализ должен определить, к какому члену предложения относится дополнение «с юмором»: «доклад с юмором» или «с юмором обсуждали».

На сегодня семантический анализ неплохо работает с онтологиями в конкретной узкой сфере знаний. Однако область применения

остаётся сильно ограниченной там, где требуется интерпретировать диалоговый текст живых носителей языка: семантические модули пока не справляются с обилием вариантов, с помощью которых человек подчас выражает один и тот же смысл.

Решение проблем с синтаксическим и семантическим анализом для естественных языков позволило создать *системы машинного перевода* между различными естественными языками. За последние 5 лет в этой области достигнуты впечатляющие успехи, хотя и здесь без решения проблемы машинного понимания сложно надеяться, что программа будет способна переводить, например, художественные тексты на удовлетворительном уровне – даже если оставить в стороне художественные достоинства и ограничиться только адекватной смысловой передачей текста.

4. Анализ дискурса (*Discourse integration*).

Семантические связи в некоторых случаях можно определить только с учетом контекста. Значение предложения часто становится понятным только в окружении других, как правило, соседних предложений из анализируемого текста. Так, в упомянутом чуть выше примере контекст может быть следующим: «Запомнились два доклада: один с юмором, и другой, на котором все заснули. Доклад с юмором обсуждали. О втором никто ничего не мог сказать.» Семантическая связь «доклад» - «с юмором» становится очевидной только исходя из контекста.

Модуль анализа дискурса должен выдавать непротиворечивую структуру семантических связей внутри одного текста (или части текста). Границы этой задачи, как и предыдущей, определены нечетко. На данный момент сложно говорить об успешном решении в данной области, несмотря на обилие подходов и моделей.

5. *Прагматический анализ*. На этой завершающей стадии машинной обработки текста высказывание преобразуется в то значение, которое было заложено при его формировании. Конкретное высказывание кодирует некий мыслеобраз, который в итоге должен быть заключен в адекватную форму, «понятную» машине. Другими словами, машина должна извлечь смысл, который подразумевался при формулировке конкретного высказывания, и представить его в той форме, в какой будет возможна его дальнейшая обра-

ботка: генерирование ответного высказывания или запуск определенной процедуры.

Прагматика использует те аспекты языка, которые подразумевают реальные знания о мире, в котором живет и действует человек. На этом этапе обработки текста начинают проявляться индивидуальные свойства конкретных естественных языков. Если тексты научно-технической направленности могут передаваться на другом естественном языке практически без искажений, то художественные тексты переводятся на другой язык с известными ограничениями даже человеком. Это связано с тем, что между понятиями носителей различных естественных языков невозможно установить однозначные соответствия. Таким образом, мысль, выраженная на одном естественном языке, не всегда может быть передана адекватно средствами другого языка.

Может показаться, что в тезисе о невозможности полноценного художественного перевода заключается парадокс, т.к. он означает, что носители разных языков лишены возможности полноценно понимать друг друга. В некотором смысле этот парадокс имеет место, но на сегодня подобные тонкости едва ли актуальны для проблем искусственного интеллекта.

Человек на практике способен находить эквиваленты, пусть не тождественные, но вполне адекватно передающие начальный смысл средствами другого языка. С машинной точки зрения, такой перевод далек от эквивалента, зато с точки зрения полноты передачи целостного образа он признается наиболее точным. Талантливые переводчики находят поразительно точные соответствия. Приведем иллюстрацию из [15]:

The mountain tops were hidden in a grey waste of sky... (A. Cronin)

При переводе на русский переводчик нашел неожиданный эквивалент:

Вершины гор *тонули* в сером небе. (Пер. М. Абкиной).

Шанс обнаружить соответствие «to hide» – «тонуть» в словарях близок к нулю, однако в данном случае это соответствие представляется более адекватным, чем словарное значение «to hide» – «прятать, скрывать».

Таким образом, переводчику удалось понять изначальный образ, заложенный в тексте, на более глубоком уровне, нежели тот, который был доступен для передачи образа

средствами искусственного интеллекта (словари и онтологии).

Описанная проблема пока находится вне сферы разработок в области искусственного интеллекта. К настоящему моменту нет даже ответа на вопрос о возможности принципиальной реализации подобных задач. Определенные надежды возлагаются на обучение нейронных сетей с участием человека, но едва ли кто-либо даст прогноз, до каких пределов удастся продвинуться в этом направлении.

Автоматическая обработка текста давно стала классической дисциплиной компьютерной лингвистики, и здесь нет необходимости подробно рассматривать основы, тщательно изложенные во многих монографиях, например [16]. Весьма интересный взгляд на историю машинного перевода, на наш взгляд, представлен Джоном Хатчинсоном в работе [17].

4.2 Представление знаний

В искусственном интеллекте и в человеческом сознании *форма представления знаний* принципиально различается. Поиск решений в области машинного представления знаний идет по многим направлениям, постоянно предлагаются новые подходы к решению этой проблемы. Поскольку понятие «знание» включает в себя невербализируемые составляющие, мы обратимся к этой теме ниже, после рассмотрения проблемы метафоры.

4.3 Омонимия

Если инструменты для грамматического разбора (parsing) предложений разработаны сегодня на хорошем уровне для большинства распространенных естественных языков, то проблема снятия омонимии до сих пор является одной из самых острых при анализе текста. Виды омонимии обсуждаются, например, авторами [18]. Даже такая простая фраза как «Иванов держит деньги в банке» на стадии морфологии анализа даст как минимум два значения, поскольку предложение допускает прочтение с существительным «банк» (мужского рода) и «банка» (женского рода). На стадии семантического анализа возникнет дополнительная неоднозначность, поскольку «банка» может означать не только емкость, но и песчаную отмель, а также поперечину шлюпки. В специальных словарях существительное «банка» обрстет еще множеством значений. Глагол держать в данном случае

может иметь синоним «хранить», тем более если он входит в идиоматическое выражение «держат деньги», но вполне можно допустить интерпретацию «Иванов держит деньги в банке, зажатой в руке».

В практических задачах учитывается статистический вес того или иного варианта [19]. С точки зрения проблемы понимания в искусственном интеллекте подобное решение не является вполне «честным», поскольку даже в задачах машинного перевода статистика не в состоянии обеспечить подбор адекватных эквивалентных синонимов. Более сложные варианты подразумевают использование машинного обучения (machine learning) [20], разметку или иные методы [21, 22]

Однако для более или менее адекватного «понимания» естественного языка и оперирования со множеством смысловых значений, содержащихся в тексте, машина должна не только иметь очень богатую онтологию, сопоставимую по объему с человеческими представлениями о мире, но и алгоритм, позволяющий работать с актуальным набором вариантов. В указанном примере «Иванов держит деньги в банке» машина должна иметь критерий, по которому существительное «банка» в смысле сосуда или песчаной отмели не подходит или наоборот, является единственно правильным.

5 Распознавание эмоций

В коммуникации с участием человека «понимание» включает в себя эмоциональную составляющую. В последние годы в области распознавания эмоций достигнут значительный прогресс (Emotion Detection). Для классификации часто используется модель Экмана, предложившего шесть основных эмоций:

1. Злость
2. Отвращение
3. Страх
4. Счастье
5. Грусть
6. Удивление

Впоследствии список был расширен следующими эмоциональными состояниями: развлечение, презрение, приятное спокойствие, стыд, возбуждение, вина, гордость от достижения, облегчение, удовлетворение, сенсорное удовольствие, позор. Распознавание эмоций в языке – относительно новая задача, но в этой области уже наметились чет-

кие направления. Одна ветвь занимается текстовым представлением, другая работает с голосом и речью.

Эмоциональный анализ текстов получил название «Анализ тональности текста» (Sentiment analysis, Opinion mining) Создан ряд тезаурусов, содержащих эмоциональную разметку. Основная часть модулей распознавания эмоций использует их. Среди наиболее часто используемых - WordNet-Affect. Это семантический тезаурус, в котором слова, несущие эмоциональную нагрузку, наделяются признаком семантической валентности. Разметка содержит дополнительные маркеры, распределяющие эмоционально окрашенные слова по группам: *позитивная, негативная, неоднозначная и нейтральная*. WordNet-Affect содержит разметку по шести основным эмоциям в соответствии с моделью Экмана.

SentiWordNet получен с помощью автоматической обработки синсетов WordNet при обучении системы с частичным привлечением эксперта (semi-supervised learning). На выходе множество синонимов (синсеты) получает двубальную (0,1) оценку по шкалам позитивности, негативности и объективности.

SenticNet – сравнительно новый проект, в отличие от предыдущих двух работающих с семантическим значением слов, объединенных в семантически связанную группу.

Методы работы модулей эмоционального анализа текста активно используют машинное обучение Свежий обзор методов, использующихся с привлечением машинного обучения, предлагается в статье [23].

В статье [24] предлагается любопытная модель регистрации эмоций, используя исключительно структуру семантических и синтаксических связей предложения, без участия человека и использования каких-либо баз данных или словарей, созданных с участием человека. Предложенная методика по мнению авторов дает возможность оторваться от модели шести основных эмоций (модель Экмана). Авторы убеждены, что предлагаемая ими модель более универсальна и не только обеспечивает практическое решение проблемы классификации эмоций, но также приводит к значительно более точным результатам, чем предыдущие методы классификации эмоций без участия человека.

Однако, основным каналом передачи эмоций, безусловно, является не текст, а речь. Для большинства людей невербальные реплики, тон голоса, его громкость, частотный

спектр, скорость произнесения слов передают их эмоциональное состояние, которое распознается человеком без усилий. Звук речи несет в себе огромное количество информации, которую искусственный интеллект должен научиться распознавать. Без учета эмоций понимание речи не может быть полным, к тому же эмоции часто кодируют вполне конкретную информацию, которая может передаваться порой вообще без семантически значимых конструкций.

В работе [25] рассматриваются шесть различных типов классификаторов, которые служат для передачи шести основных универсальных человеческих эмоций. В методах распознавания использовались 6 аудио файлов из аудио-визуальной базы эмоций. Чтобы выбрать самые значимые речевые особенности из ряда акустических особенностей, обычно извлекаемых при анализе эмоции, использовалось дерево принятия решений. Оценка классификаторов производилась на особенностях, о которых выносились конкретные предположения, и на тех, которые были отобраны деревом решений. С таким выделением признаков наблюдалось повышение точности определения для каждого классификатора. Наилучшие результаты были получены для метода опорных векторов (Support Vector Machine) и байесовской сети.

Обзор основных методов определения тональности текста можно найти в работе [26].

С современным состоянием дел в данной области можно ознакомиться, например, по работе [27].

В настоящее время удастся достичь достаточно хорошей точности эмоционального анализа текста. С помощью полуавтоматических методов еще в 2005 году был составлен лексикон оценочных прилагательных и было показано, что методы анализа тональности текста могут достигать точности 90% [28]

6 Проблема метафоры и творчества

Машинная интерпретация естественного языка наталкивается на серьезную проблему: естественный язык метафоричен. Метафора лежит в самой основе языка, и в силу своей изначальности плохо поддается объяснению. «Тайна метафоры привлекала к себе крупнейших мыслителей - от Аристотеля до Руссо и Гегеля и далее до Э. Кассирера, Х. Ортеги-и-Гассета и многих других. О метафоре написано множество работ. О ней высказывались

не только ученые, но и сами ее творцы - писатели, поэты, художники, кинематографисты. Нет критика, который не имел бы собственного мнения о природе и эстетической ценности метафоры» [29].

Взаимодействие человека с миром происходит при посредстве языка, язык обуславливает ключевые аспекты бытия – познание и разумная деятельность. Для искусственного интеллекта метафора, пожалуй, является тем фокусом, где сходятся воедино две различные формы представления знаний. Связь между научным и художественным мышлением, где присутствие метафоры очевидно, гораздо глубже: при внимательном рассмотрении оказывается, что метафоричность мышления играет определяющую роль в научных открытиях. В частности, метафоричность мышления сыграла ключевую роль в становлении корпускулярно-волновой теории [30]. Метафора обеспечивает перенос знаний между областями, метафорическое мышление выделяет общую структуру в разных, далеких друг от друга областях знаний и дает возможность успешно применять наработки из одной области науки в другой.

Важной вехой в когнитивной теории метафоры является книга Дж. Лакоффа и М. Джонсона "Metaphors We Live by" (1980) [31]. Отводя метафоре ключевую роль в мышлении и поведении, авторы придают метафоре статус основного механизма когнитивного метода. Метафора является местом встречи объективного и субъективного, рационального и иррационального.

Все рассмотренные исследования способствовали становлению когнитивного подхода к метафоре, но именно в книге Дж. Лакоффа и М. Джонсона была разработана теория, которая привнесла системность в описание метафоры как когнитивного механизма и продемонстрировала большой эвристический потенциал применения теории в практическом исследовании. Авторы постулировали, что метафора не ограничивается лишь сферой языка, что сами процессы мышления человека в значительной степени метафоричны. Метафора как феномен сознания проявляется не только в языке, но и в мышлении, и в деятельности. "Наша обыденная понятийная система, в рамках которой мы думаем и действуем, по сути своей метафорична" [30]. Такой подход позволил окончательно вывести метафору за рамки языковой системы и рассматривать ее как феномен взаимодействия языка, мышле-

ния и культуры. В работе "The Contemporary Theory of Metaphor" Дж.Лакоффа строго ограничил метафорическое выражение и концептуальную метафору, подчеркивая, что "локус метафоры - в мысли, а не в языке" [31].

Вопрос о понимании метафоры является ключевым для искусственного интеллекта. Многие авторы обращают внимание на основное отличие от машинной обработки знания: понимание метафоры, как и ее создание, «есть результат творческого усилия оно столь же мало подчинено правилам» [32].

Когда мы говорим о понимании, мы подразумеваем, что понимание встроено в систему коммуникации. Осмысленная коммуникация состоит из нескольких стадий: мысленный образ облекается в текст, текст заключается в некую материальную форму (в случае речи это физические звуковые колебания), которая воспринимается непосредственно (на месте) или впоследствии с помощью воспроизведения записи на материальном носителе. Далее этот текст (последовательность знаков), представляющий материальный образ мысли, считывается субъектом или машиной и интерпретируется. В случае интерпретации человеком конечным результатом в случае понимания является по существу «тот же» исходный мысленный образ, который послужил началом в создании текста. Степень тождественности можно считать **мерой понимания**. В случае коммуникации человек-человек суждение о мере понимания возможно лишь на основе уточняющих взаимодействий (аналогичных коммуникаций), которые передают более полное контекстное окружение *исходного образа*.

Человеческое общение допускает широкий спектр оттенков или модусов текста. Блэк пишет: «Автор выражает не тот смысл, который он имеет в виду, а некоторую функцию от него, $f(m)$; в задачу читателя входит применить обратную функцию f^{-1} , чтобы достичь $f^{-1}(f(m))$, т. е. m или *буквального* значения высказывания. Различные тропы задаются различными функциями. Так, в случае иронии автор говорит *противоположное* тому, что он имеет в виду; в случае гиперболы — *преувеличивает* его значение и т. д.» [33].

В том случае, когда результат обрабатывается искусственным интеллектом, для установления меры понимания мы должны иметь механизм, проверяющий меру тождественности той структуры, которая сформировалась в

искусственном интеллекте на выходе в результате коммуникации, и исходного образа. Но на сегодняшний день трудно представить себе возможность создания подобного механизма: онтологическая пропасть между человеческим мышлением и искусственным интеллектом кажется непреодолимой.

В самом начале компьютерной эры существовала надежда посредством способности распознавать метафоры провести грань между наличием у некоего объекта мышления и его отсутствием. Можно сказать, что на эту способность опирался тест, изобретенный Тьюрингом. Однако именно в силу того, что творчество заложено в самой природе человека, и человек не может воспринимать текст не творчески, практическая реализация этого теста лишь отодвигает вынесение окончательного вердикта: имитирование разумного ответа и разум, творчески решающий проблему – это различные вещи. Так, практическое прохождение теста системой «Eugene Goostman» на уровне 13-летнего подростка в 2014 году не было воспринято как событие, подтверждающее наличие интеллекта у машины.

Те же соображения относятся и к художественному творчеству. Творчество также можно имитировать, внося элемент случайности в создание форм, и картина будет восприниматься как художественная не потому, что она является продуктом творчества, а потому что ее оценивает творческая личность. Картины, полученные в экспериментах с обработкой изображений с помощью нейронных сетей и обнаруженные Google летом 1915 года [34], вполне могут претендовать на роль **произведений искусства**.

Аналогичные результаты есть и в музыке. Разработаны методики по использованию компьютерных алгоритмов в композиции на основе нейронных сетей, сегодня уже каждый имеет возможность оценить образцы сочинений искусственного интеллекта [35, 36], нашла свое воплощение в музыке и идея использования корпусов [37]. Процессы использования компьютерных преобразований человеком при написании музыки нельзя считать участием компьютера в создании произведения: в случайных ходах, предлагаемых компьютером, композитор может находить материал для творчества [38], но он один остается тем, кто творит.

Машина способна на более или менее правдоподобную имитацию, но *имитация не*

перестает быть таковой, если даже она вызывает у человека отклик, сопоставимый с реальными произведениями искусства.

Поиск различий устройства человеческого разума и машины продолжается. Можно разглядеть в проблемах привлечения искусственного интеллекта к творчеству все ту же проблему метафоры. В метафоре всегда присутствует сравнение. Уэйтли пишет: «Сравнение можно рассматривать как отличное от Метафоры только по форме: в случае Сравнения сходство *утверждается*, а в случае Метафоры — *подразумевается*» [39]. Машину можно научить находить общее, а значит, она может *предлагать для сравнения* какие-то варианты точно так же, как она предлагает распознанные формы деревьев или животных в экспериментах Google, или варианты изменения музыкальной темы. Отличие человека проявляется там, где начинается «подразумевание».

Д. Дэвидсон считает, что человек при реализации своего творческого потенциала *пользуется несемантическими ресурсами*. «Любая коммуникация — это взаимодействие мысли изреченной и мысли, извлеченной из речи. Вопрос лишь в степени разрыва. Метафора его увеличивает тем, что пользуется в дополнение к обычным языковым механизмам несемантическими ресурсами. Для создания метафор не существует инструкций, нет справочников для определения того, что она “означает” или “о чем сообщает” [40].

7 Представление знаний

Очевидно, что *формы представления знаний* в искусственном интеллекте и в человеческом сознании принципиально различаются. Цель представления знаний в ИИ - организация человеческого знания таким способом, чтобы машина могла оперировать с ним, представляя на выходе продукт в форме знания, являющегося ответом на запрос.

Не отвлекаясь на пересказ общеизвестных вещей, уделим внимание некоторым работам, имеющим непосредственное отношение к обсуждаемым здесь проблемам. В частности, проблеме метафоры в аспекте организации представления знаний посвящено исследование [41]. Здесь обсуждается широкий спектр проблем когнитивной лингвистики: немонотонная логика, логика первого порядка против семантических сетей, сравнивается процедурное и декларативное представление,

обсуждается проблема «ясного» подхода к искусственному интеллекту против «сложного» (“neats vs. scruffies” problem). Большой объем занимает подробное изложение теории «Иерархии динамического типа» (Dynamic Type Hierarchy). За основу этой теории взят тип иерархии со множественным наследованием. Метафора рассматривается как способ выделения и изменения структуры иерархии. Основная идея заключается в том, что интерпретация метафоры осуществляется с помощью поиска «снизу вверх» по иерархии с целью обнаружения общего предка содержания и оболочки метафоры (tenor and vehicle of the metaphor). Этот поиск приводит к созданию новой концепции, которая конкретизирована более, чем общий предок, являющийся непосредственным родителем концепций содержания и оболочки. К большому сожалению, экспериментальное воплощение данного подхода в книге не обсуждается.

Попытка найти адекватную форму представления знаний для понимания и создания метафор, являющихся продуктом творчества, делается в работе [42]. Предлагается модель «жидкостного представления знаний» (Fluid Knowledge Representation). Поскольку метафора искажает семантическое представление, часто привнося в семантику юмористический или поэтический оттенок, компьютерное моделирование метафоры требует более гибкого представления знаний.

Авторы предлагают реализацию такого гибкого представления знаний, которое удовлетворяет условиям семантической неопределенности. Решение получило название «Основные тезисы» («Talking Points»). В статье показывается, как «основные тезисы» могут добываться из WordNet и из интернета. Авторы показывают, что «основные тезисы» могут объединяться гибкой связью для формирования динамической семантической сети - slipnet.

Slipnet может быть представлена в виде долгосрочной памяти у человека. Узлами сети служат концепты, связи между ними – концептуальные отношения. Узлы имеют уровень активации, который повышается с ростом релевантности концепта. Когда программа определяет, что концепт может иметь отношение к конкретной проблеме, уровень активации повышается. Когда достигается определенный порог, узел может достичь предельного уровня активации, и тогда программа фокусируется на нем. Постепенно с

течением времени уровень активации уменьшается, если новые запросы на тот же концепт не поступают. Глубокие концепты более инерционны, чем мелкие: для них требуется большее время для достижения полной активации, зато они «гаснут» также медленнее. Данная модель моделирует работу человеческого внимания и памяти: когда проблема «всплывает» в памяти, внимание на ней удерживается тем дольше, чем более важной она является. Более подробно основы построения динамических семантических сетей рассмотрены в работе [43].

Современные проблемы представления знаний рассматриваются в работе [44]. Авторы корректируют сегодняшнее видение ситуации с представлением знаний при создании систем искусственного интеллекта. Речь идет о построении систем, которые располагают знаниями о том мире, в котором они существуют, и способны действовать в нем на основе получения информации об этом мире подобно тому, как человек действует в своем мире на основе получаемой информации. Критическим положением для организации представления знаний в подобных системах является тот факт, что знание представлено в символической форме, и что процедуры *формирования рассуждений (reasoning)* в состоянии извлекать выводы из такого знания в форме новых символических представлений. Такая способность используется, чтобы принимать на основе получаемой информации взвешенные решения о том, какое действие следует осуществить.

Авторы вслед за Levesque [45] считают эту идею революционной, продолжающей историю логики, начатую Аристотелем и продолженную Лейбницем, который выдвинул тезис о тождестве знания и вычисления. Следующий радикальный шаг был сделан в 1959 году, когда МакКарти (McCarthy) [46] сформулировал тезис о решающей роли представления знаний в искусственном интеллекте, без чего невозможно создание интеллектуальных систем одними лишь средствами программирования и изменения архитектуры системы. Система, действующая разумно, должна уметь определять, что ей нужно знать, а также извлекать выводы из такого знания. Современные основы представления знаний и формирования рассуждений (Knowledge representation and reasoning, KR) строятся именно на ключевой идее МакКарти.

Сегодня представление знаний превратилось в самостоятельную область знаний, по ней читаются курсы, выпускаются учебники, многие из которых признаны классическими. В качестве примера можно привести книгу Брахмана и Левескё [47]. С 1989 года стали проводиться международные конференции по представлению знаний. В настоящее время регулярно проводятся: Международная Объединенная Конференция по Искусственному интеллекту (IJCAI), Конференция AAAI по Искусственному интеллекту (AAAI), и Европейская Конференция по Искусственному интеллекту (ECAI). Одной из основных наиболее представительных конференций в данной области является Конференция по принципам представления знаний и формированию рассуждений (Conference on Principles of Knowledge Representation and Reasoning – KR).

Среди современных приоритетных направлений в сфере представления знаний можно выделить следующие: дескрипционная (описательная) логика (Description Logics, DL); обоснование действий и процессов (Reasoning about Actions and Processes); обновление представлений и немонотонность (Belief Revision and Nonmonotonicity); Представление общих знаний и формирование рассуждений (General Knowledge Representation and Reasoning); Планирование, стратегии и диагноз (Planning, Strategies, and Diagnosis); программирование, основанное на поиске стабильной модели и логическое программирование (Answer Set Programming and Logic Programming); аргументация (Argumentation); Автоматическое формирование рассуждений и машинное вычисление (Automated Reasoning and Computation); причинная зависимость (Causality); разумность и неопределенность (Rationality and Uncertainty). В отдельную область выделилось Понимание естественного языка (Natural Language Understanding (NLU) [48].

Дескрипционные логики (DL) представляют собой семейство языков представления знаний для формализованного описания понятий конкретных предметных областей. Язык DL по синтаксису напоминает язык модальных логик. DL наследовали наработки логики концептов. В настоящее время DL применяются в семантической сети (Semantic Web) и при создании онтологий. Основы DL изложены в [49], на русском языке их описание вкратце дается в монографии С. Рассела и

П. Норвига “Искусственный Интеллект, Современный Подход” [50].

Reasoning about Actions - обоснование действий и процессов. Это направление имеет уже богатую историю в разработке методов искусственного интеллекта. *Ситуационное исчисление (situation calculus)* было введено Джоном Маккарти в 1963 как способ логичного определения динамических систем. В 1998 году Alessandro Artale и Enrico Franconi предложили темпоральную дескрипционную (описательную) логику для обоснования действий и планирования [51]. Ими был введен класс темпоральных языков, основанных на интервалах времени, для *формирования рассуждений* по поводу действий и планирования. Действия были представлены с помощью описания, что именно является истинным в то время, когда возникают сами эти действия. Планы создавались временными действиями, привязанными к конкретному временному отрезку, а также к состоянию мира. Темпоральные языки были разработаны в рамках семьи дескрипционных логик. Дальнейшее развитие направления изложено в работе [52].

Ситуационное исчисление продолжает расширяться, в этой области работает большое число исследователей. Их цель – научиться описывать и формализовывать множество явлений, связанных с обоснованием действий (т.е. *формирование рассуждений* в отношении действий). Концентрирование усилий в данной области оказывает существенную помощь в разработке спецификаций и попытках внедрения разнообразных систем машинного поиска логического вывода (другое название - автоматизированное формирование рассуждений). Данные системы находят применение в диагностике, при создании и обслуживании веб-сервисов.

Появились альтернативные подходы к планированию, которые, в частности, получили распространение под именем неклассических систем автоматического планирования (non-classical automated planning). В классической ветви исследований по ситуационному исчислению в последнее время интерес смещается от формирования рассуждений в отношении какого-то одного действия к формированию рассуждений в отношении процессов и программ высокого уровня.

8 Приложения с машинным пониманием

Актуальность представления знаний при разработке современных приложений, рассчитанных на крайне широкую аудиторию пользователей, рассматривается в статье [53]. Приложение «Помощник-организатор времени» (Personal Time Assistant, PTA), которым в настоящее время занимается компания Google, состоит из трех частей: первая отвечает за ввод информации: пользователь должен иметь возможность представить системе все варианты использования времени, вторая – это обучение программы, и третья должна обрабатывать варианты поведения при моделировании окружающей среды. При этом должны учитываться разнородные варианты, выбор которых зависит от разнородных параметров: деловые встречи, события, поручения, проекты, хобби, семейные дела, состояние здоровья и его поддержание, занятия спортом, потребность просто отдохнуть и собраться с мыслями.

В результате была создана модель данных, которая получила название «объект намерения» (intention object, IO). IO - вектор характеристик (вектор признаков - feature vector), который включает в себя текстовое описание, временные признаки (когда действие может быть выполнено, когда оно должно быть выполнено, его продолжительность — при этом все параметры задаются с разной степенью точности), условия для выполнения намерения (такие как местоположение или необходимые инструменты), а также другие виды параметров.

9 Распознавание речи

Машинное понимание речи в целом определяется технологиями понимания текста, однако имеет свои особенности. Как правило, обработка речи начинается с ее распознавания (Speech Recognition, SR), и далее происходит по той же схеме, по которой происходит обработка текста.

Долгое время на разговорную речь смотрели как на устную версию упрощенного письменного языка. Однако с развитием методов SR стало понятно, что разговорная речь имеет ряд принципиальных отличий, которые становятся определяющими при машинной обработке. Если текст является одинаправ-

ленной коммуникацией, то разговорная речь подразумевает диалог, ориентированный на ожидание обратной связи. При этом ответ может быть сформирован в невербальной форме с помощью жеста, мимики или даже молчания, которое в каких-то случаях может обладать семантической составляющей. К этому нужно добавить, что частотный словарь, устойчивые словосочетания и весь грамматический строй разговорной речи сильно отличается от письменного языка. Требования к грамматике при общении несравненно ниже, а грамматические правила в диалоге очень часто нарушаются: фразы не договариваются до конца, с середины предложения может быть подхвачено продолжение предыдущей фразы и т.п.

В настоящее время стадии распознавания речи и семантическая обработка в большинстве подходов машинной обработки речи разделены. На этапе преобразования речи в текст теряется много семантически значимой информации. С решением проблем машинного понимания текста, очевидно, потребуется другая организация архитектуры нейронных сетей и другие решения, позволяющие учитывать семантику интонаций, временную составляющую речи (скорость, паузы), звуки, относящиеся к речи, но не входящие в состав фонем (таких как «гм-гм», пощелкивание языком) и пр.

Задачи машинного распознавания речи возродили актуальность тех разделов лингвистики естественных языков, которые занимаются звуковыми аспектами языка. Любая речь представляет собой звуковые колебания, которые с помощью микрофона оцифровываются машиной. Согласные и гласные звуки составляют множество *элементарных речевых звуков* (phones). В звонких согласных присутствуют как тоны, так и шумы. Шумы и тоны речи обладают специфическим тембром, высотой тона, относительным уровнем громкости и другими параметрами. Звуковой строй языка изучает самостоятельный раздел лингвистики - *фонология*. Базовой единицей фонологии является *фонема* (элементарная единица языка, с изменением которой меняется смысл слова: «тот» состоит из фонем «т» и «о»). С изменением первой фонемы на «д» мы получаем другое слово: «дот»). Фонология занимается исследованием фонем, из которых складывается фонологическая система языка. В речи в каких-то случаях различные

фонемы могут иметь идентичное произношение.

В русской фонологии различаются две школы, отличающиеся своими подходами: Ленинградская фонологическая школа и Московская фонологическая школа.

При обработке речи задачей машины на начальном этапе является выделение фонем из звукового ряда. При этом сразу встает проблема фильтрации шумов. Сигнал, поступающий в компьютер, кроме речи содержит посторонние шумы, которые необходимо отфильтровывать от элементарных речевых звуков.

Когда такой фильтр настроен и человеческая речь выделяется из общего звукового потока, элементарные речевые звуки требуется объединить в морфемы и разделить слова, которые произносятся слитно. Произношение в естественном языке изобилует редукцией гласных, оглушением согласных, и в результате знакомое нам слово представляет собой вовсе не последовательную сумму известных фонем аналогично тому, как последовательность букв складывается в слово. Например, слово произношение «сколько» практически неотличимо от последовательности двух слов: «с Колькой». Звук «й» в окончании машины вполне может не отличить от шумового фона. Омфония наблюдается и у последовательности одних и тех же фонем, где семантическую функцию в тексте берет на себя пробел: «в месте» — «вместе».

Для машины будет по-разному выглядеть одна и та же фраза, произнесенная одним и тем же человеком в разное время или в сопровождении разного шумового фона. Звуковой сигнал будет сильно зависеть от конкретных характеристик микрофона и соотношения уровней громкости «фон - речевой сигнал». На него сильно влияет стиль, эмоциональное состояние говорящего, его индивидуальные биологические характеристики (у всех людей резонансные полости, где образуются звуки речи, разные), скорость речи.

В итоге, как мы видим, распознавание речи превращается в отдельную большую задачу. На сегодняшний день она решена на удовлетворительном уровне. Смартфоны вполне справляются с задачей распознавания речи пользователей. Применяемые технологии активно используют методы глубокого обучения (Deep learning) Иногда можно встретить перевод термина как глубинное обучение.

Глубокое обучение использует архитектуру из нескольких слоев нейронной сети, где абстрактное значение высокого уровня собирается из абстрактных значений более низких уровней. Так, например, человек сразу понимает эмоции по выражению лица. Состояние радости можно разложить на признаки более конкретные: в данном случае важно отслеживать положение уголков губ. Для машины изображение лица складывается из отдельных пикселей, и методика глубокого обучения позволяет связать положение конкретных пикселей на изображении с эмоцией «радость».

Подробнее про методы глубокого обучения можно посмотреть в [54], [55], а также [56]. Эффективность использования методов глубокого обучения сильно зависит от объема вычислений. Например, технология перевода слов в векторную форму word2vec, используемая Google в семантических приложениях, в частности, при обработке запросов и исправлении опечаток при вводе запроса, собирает статистику по совместному появлению слов в запросах или фразах из корпусов текстов большого объема. После этого с помощью нейронных сетей решается задача уменьшения размерности, и на выходе получаются векторные представления слов, по которым легко выявляются слова с близкими семантическими значениями.

Оптимальное решение, естественно, находится с помощью соответствующего подбора архитектуры и необходимого объема статистической выборки. В некоторых случаях решение задачи может в большей степени определяться архитектурой модели.

Современные методы хорошо справляются с обработкой одного голоса. При переходе на другой голос система распознавания делает намного больше ошибок. Однако исследователи в IBM заявили о создании алгоритма, который отлично работает при распознавании речи двух голосов, беседующих по телефонной линии. Уровень ошибок при этом составляет всего 8%, это уже очень близко к человеческому восприятию: уровень ошибок восприятия в аналогичных условиях у человека может достигать 4%. Исследователи надеются выйти на уровень человека в распознавании речи в самое ближайшее время. [57]

Улучшение было достигнуто в результате применения сильных моделей языка с обучением на разнообразных источниках данных; объединенному обучению рекуррентной

нейронной сети (Recurrent Neural Network, RNN) и свёрточной нейронной сети Convolutional Neural Network, CNN) с 32000 выходами на протяжении 2000 часов аудио, а также сетей maxout с исключением разрядов метод имитации отжига (annealed dropout); Достижение повышения точности ожидалось обучением сетей maxout и свёрточных нейронных сетей большего объема. Сети maxout рассматриваются в работе [58].

Сегодня технология распознавания речи успешно используется в приложениях Apple и Google, существуют продукты по распознаванию речи и автоматическому преобразованию текста в речь (TTS), которые активно используются банками и авиакомпаниями. В этой области хорошо известны продукты Nuance.

Недавно появилось сообщение, что алгоритм Baidu опередил людей в распознавании речи на английском и китайском языках. Лаборатории искусственного интеллекта SVAIL китайской компании Baidu удалось улучшить алгоритм распознавания речи Deep Speech настолько, что в некоторых случаях программа распознает речь быстрее людей. Препринт статьи опубликован на arXiv.org [59].

10 Диалекты

Особую проблему в распознавании речи составляет наличие диалектов. Произношение в разных диалектах может отличаться настолько сильно, что носители разных диалектов с трудом понимают друг друга. Эта проблема остро стоит в Азии. Индонезийский и малайский языки имеют много заимствований из санскрита, тамильского, арабского, португальского, нидерландского и английского языков. Единая система орфографии была введена для индонезийского языка только в 1972 году. Китайцы из разных регионов вообще могут не понимать устную речь другого региона, хотя письменный язык остается одним и тем же.

Хотя эта проблема имеет в большей степени «человеческий» аспект, нежели компьютерный, в фонологии, тем не менее, вводится количественная оценка несхожести диалектов между собой. Обычно степень различия определяют с помощью Расстояния Левенштейна. Оно определяется как минимальное количество шагов, необходимых для превращения одного слова в другое. Шагом здесь является операция вставки одного символа, удаления одного символа и замены одного

символа на другой. При фонологических исследованиях роль символов играют фонемы. Реализация использования расстояния Левенштейна в фонологии обсуждается в работе [60].

Пример другого подхода излагается в статье [61]. Метод, применяемый авторами, основан на классификаторе языков Даннинга, использующем максимальную схожесть языков между собой. В русской традиции этот метод получил название «Идентификация языка» (англ. language identification), или «определение языка» (встречается в машинной лингвистике по методам обработке естественного языка – см. [62]). С помощью «наивной транскрипции» (naïve transcriptions) схожесть диалектов проверялась на материале усвоения речи детьми и пользователями слуховых аппаратов. Было сделан вывод о высокой корреляции метода «Идентификация языка» с методами, основанными на расстоянии Левенштейна.

Понимание зависит не только от восприятия слов, но и от грамматики. Различие в грамматике начинает проявляться с расстоянием. Отношения между фонологическим и географическим расстоянием на примере исследования голландских диалектов предлагается в работе [63].

Понимание на уровне коммуникации между людьми, принадлежащими разным культурам, представляет также проблему. Разные народы используют различные структуры «представления знаний» - если смотреть на человека с точки зрения машины. Тем не менее, неречевые семантические знаки человек использует с помощью речевого представления. В статье [64] обсуждается использование фонологического представления при обработке сложных не речевых сигналов.

11 Заключение

Итак, мы попытались окинуть взглядом область проблем, которые лежат перед разработчиками систем искусственного интеллекта при попытках моделирования понимания. Часть проблем уже решена на хорошем уровне, где-то системы ИИ работают с большей эффективностью, чем человек, где-то едва делаются первые шаги. Важно то, что на сегодня, в отличие от конца XX столетия, практически невозможно найти области, к которым хотя бы не намечались путей подхода для внедрения систем искусственного ин-

теллекта. При этом необходимо отметить, что на новом уровне встает вопрос о принципиальной возможности мышления у искусственно созданной системы. Для определения интеллекта, а также связанных с ним понятий разумности и мышления, потребовались новые критерии. На смену тесту Тьюринга, где от искусственного интеллекта требовалось определенное поведение, пришла постановка вопроса в формулировке Джона Серла (John Searle), где ключевым моментом является не действие, а понимание. В мысленном эксперименте «Китайская комната» Серл разграничивает понятия ИИ в слабом и сильном смыслах. Интеллект в слабом смысле отличается от ИИ в сильном смысле тем, что лишен сущностного критерия: он есть лишь моделирование. О наличии искусственного интеллекта в сильном смысле можно будет говорить лишь в том случае, если мышление возникает как неотъемлемое свойство созданной системы независимо от материалов и способа ее изготовления.

Мы постарались показать, что проблема машинного понимания чрезвычайно сложна и связана с крайне широким спектром проблем. Создается впечатление, что для достижения уровня машинного понимания, сравнимого с человеческим, посредством ИИ в слабом смысле потребуются мощности, которые на сегодняшний день представляются фантастическими. С другой стороны, мы обратили внимание на прогресс по многим направлениям, что позволило сделать сильнейший рывок в развитии ИИ, особенно за последние годы. Похоже, что проблема машинного понимания, в решении которой также заметно существенное продвижение, в ближайшее время будет по-прежнему оставаться в числе наиболее актуальных проблем для систем ИИ.

Список литературы

1. Silver D. et al. *Mastering the game of Go with deep neural networks and tree search*. Nature 529, 484–489 (28 January 2016). doi:10.1038/nature16961.
2. John Tromp. Number of legal Go positions. URL: <http://tromp.github.io/go/legal.html> (дата обращения 05.03.2016).
3. Lobo D, Levin M. 2015. Inferring Regulatory Networks from Experimental Morphological Phenotypes: A Computational Method Reverse-Engineers Planarian Regeneration. PLoS Comput Biol 11(6): e1004295. doi:10.1371/journal.pcbi.1004295.
4. Riedl, Mark O. and Harrison Brent. 2016. *Using Stories to Teach Human Values to Artificial Agents*. Proceedings of the 2nd International Workshop on AI, Ethics and Society, Phoenix, Arizona.
5. Bloom, B. S. et al. 1956. *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. New York: David McKay Company.
6. McTighe, J. and Wiggins, G. 2004. *Understanding by Design*. p.155 Virginia: ASCD.
7. Nai Ding et al. 2016. Cortical tracking of hierarchical linguistic structures in connected speech. Nature Neuroscience 19, 158–164. doi:10.1038/nn.4186.
8. Remi van Trijp. 2013. A Comparison between Fluid Construction Grammar and Sign-Based Construction Grammar. *Constructions and Frames* 5(1): 88–116.
9. Steels, Luc 2012. *Design Methods for Fluid Construction Grammar*. In Luc Steels (Ed.), *Computational Issues in Fluid Construction Grammar*, 3–36. Berlin: Springer.
10. Wellens, P. et al. 2013. *Fluid Construction Grammar for Historical and Evolutionary Linguistics*. Eds. Steels, Luc and Manfred Hild. Sofia, Bulgaria: Association for Computational Linguistics. 127–132.
11. Steels, L. et al. 2012. *Emergent Action Language on Real Robots*. In Luc Steels and Manfred Hild (Eds.), *Language Grounding in Robot* 255–276. New York: Springer.
12. Holland O. and Knight R. 2006. *The Anthropomorphic Principle*. in Burn, Jeremy and Wilson, Myra (eds.), *Proceedings of the AISB06 Symposium on Biologically Inspired Robotics*.
13. Devereux, D. et al. 2011. *Using the Microsoft Kinect to model the environment of an anthropomorphic robot*. In: The second IASTED international conference on robotics: Robo 2011, 7-9 Nov. 2011, Pittsburgh, USA.
14. Ovchinnikova, E. 2012. *Integration of World Knowledge for Natural Language Understanding*. Atlantis Press.
15. Комиссаров В.Н. Теория перевода М.: Высш. шк., 1990. с. 95.
16. Jurafsky, D and Martin, James H. , 2008. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*, (2nd Ed.), Prentice - Hall.
17. Hutchins, J. 2010. *Machine translation: a concise history*. in *The teaching of computer-aided translation*, ed. Chan Sin Wai. (Chinese University of Hong Kong); pp.29–70.

18. Anjali, M K., Babu, Anto P. 2014. *Ambiguities in Natural Language Processing* International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 5, Oct. 2014. p. 392–394.
19. Peter, F. et al. 1991. *Word-sense disambiguation using statistical methods*. in: 91 Proceedings of the 29th annual meeting on Association for Computational Linguistics, p.264-270.
20. Márquez, L. *Machine Learning and Natural Language Processing*. Technical Report LSI-00-45-R, Departament de Llenguatgesi Sistemes Informàtics, Universitat Politècnica de Catalunya, 2000. URL: <http://l2r.cs.uiuc.edu/~danr/Teaching/CS546-11/Papers/marquez-LNLP00.pdf> (дата обращения 05.02.2016).
21. Evang, K. & Bos, J. 2013 *Scope Disambiguation as a Tagging Task*. in: Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013) – Short Papers. Erk, K. & Koller, A. (eds.). Potsdam, Germany: Association for Computational Linguistics (ACL), p. 314–320.
22. Adwait Ratnaparkhi. 1998. Maximum entropy models for natural language ambiguity resolution. University of Pennsylvania Philadelphia, PA, USA.
23. Ubeeka Jain, Amandeep Sandhu. 2015. A Review on the Emotion Detection from Text using Machine Learning. International Journal of Current Engineering and Technology. Vol.5, No.4 (Aug-2015). p. 2645–2650.
24. Ameeta Agrawal, Aijun. 2012. *An Unsupervised Emotion Detection from Text Using Semantic and Syntactic Relations*. in: Proceedings of The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology - Volume 01 Pages 346–353.
25. Javier G.Razuri et al. 2015. *Speech emotion recognition in emotional feedback for Human-Robot Interaction (IJARAI)* International Journal of Advanced Research in Artificial Intelligence, Vol. 4, No.2, 2015 p. 20–27.
26. Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis Foundations and Trends in Information Retrieval Vol. 2, No 1-2 1–135.
27. Orestes Appel et al. 2015. Main Concepts, State of the Art and Future Research Questions in Sentiment Analysis Acta Polytechnica Hungarica Vol. 12, No.3. p. 87–108.
28. Whitelaw, C. et al. 2005. Using appraisal groups for sentiment analysis. In: Proc. the ACM SIGIR Conference on Information and Knowledge Management (CIKM), pp. 625-631.
29. Н. Д. Арутюнова. Метафора и дискурс. // Теория метафоры. М., 1990. С. 5–32.
30. Друлак П. Метафора как мост между рациональным и художественным. // Будаев Э.В., Чудинов А.П. Современная политическая лингвистика. Екатеринбург, 2006. С. 136–151.
31. Lakoff G. 1993. *The Contemporary Theory of Metaphor* // Metaphor and Thought. / ed. A. Ortony. - Cambridge: Cambridge University Press, p. 203.
32. Дэвидсон Д. Что означают метафоры // Теория метафоры. М., 1990. С. 172-193.
33. М. Блэк Метафора // Теория метафоры. М., 1990. С. 153 –172.
34. Anguelov D. et al. 2015. *Going Deeper with Convolutions*. in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
35. Composing Music With Recurrent Neural Networks URL: <http://www.hexahedria.com/2015/08/03/composing-music-with-recurrent-neural-networks/> (дата обращения 05.02.2016).
36. Jose David Fernandez Francisco Vico. 2013. *AI Methods in Algorithmic Composition: A Comprehensive Survey* Journal of Artificial Intelligence Research 48 (2013) 513-582.
37. A Corpus-Based Hybrid Approach to Music Analysis and Composition. 2007, Association for the Advancement of Artificial Intelligence URL: <https://www.aaai.org/Papers/AAAI/2007/AAAI07-133.pdf> (дата обращения 05.02.2016).
38. Automatic Composition from Non-musical Inspiration Sources URL: <http://www.hexahedria.com/2015/08/03/composing-music-with-recurrent-neural-networks/> (дата обращения 05.02.2016).
39. Цит. по: Блэк М. Метафора // Теория метафоры. М., 1990. С. 153–172.
40. Дэвидсон Д. Что означают метафоры // Теория метафоры. М., 1990. С. 172–193.
41. Eileen Cornell. 1991 *Knowledge Representation and metaphor*. Kluwer Academic Publishers.
42. Veale T, Hao Y. 2008. *A Fluid Knowledge Representation for Understanding and Generating Creative Metaphors*. Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008), p. 945–952 Manchester.
43. Joaquin Vanschoren, 2CW. June 2004 *Fluid Concept Architecture: A Critical Evaluation* URL: <http://people.cs.kuleuven.be/~joaquin.vanschoren/zsp/fluidconcepts.pdf> (дата обращения 05.02.2016).
44. Chitta Baral, Giuseppe De Giacomo. *Knowledge Representation and Reasoning: What's Hot*. Proceedings of the Twenty-Ninth {AAAI} Conference

- on Artificial Intelligence, January 25-30, 2015, Austin, Texas, {USA.} p. 4316—4317 URL: <http://www.dis.uniroma1.it/~degiacom/papers/2015/AAAI15kr.pdf> (дата обращения 05.02.2016).
45. Levesque, H. 2014. *On our best behavior*. Artificial Intelligence 212:27–35.
46. McCarthy, J. 1959. *Programs with common sense*. In Proceedings of the Teddington Conference on the Mechanization of Thought Processes, 756–791.
47. Brachman, Levesque. Knowledge Representation and Reasoning Morgan Kaufman, 2004 URL: <http://www.cin.ufpe.br/~mtcfa/files/in1122/Knowledge%20Representation%20and%20Reasoning.pdf> (дата обращения 05.02.2016).
48. Ovchinnikova Ekaterina. 2012. *Integration of World Knowledge for Natural Language Understanding* Atlantis Thinking Machines Volume 3.
49. Baader F. et al. The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press, 2003.
50. Рассел С., Норвиг П. Искусственный интеллект. Современный подход. 2-е изд. М.: Вильямс, 2007. с. 482–483.
51. Alessandro Artale , Enrico Franconi. 1998. *A Temporal Description Logic for Reasoning about Actions and Plans*. Journal of Artificial Intelligence Research 9. p. 463-506 URL: <http://arxiv.org/pdf/1105.5446.pdf> (дата обращения 05.02.2016).
52. Rosati, R. 2014. Query answering and rewriting in ontology-based data access. In KR. Tutorial.
53. Yoav Shoham. 2016. *Why Knowledge Representation Matters*. Communications of the ACM, Vol. 59 No. 1, p. 47-49.
54. Li Deng and Dong Yu. 2014. *Deep Learning: Methods and Applications*, Foundations and Trends® in Signal Processing: Vol. 7: No. 3–4, pp 197-387.
55. Dahl, G. et al. 2012. *Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition*. IEEE Trans. Audio, Speech, and Language Processing, Vol. 20(1): 30–42.
56. LeCun Y. et al. 2015. *Deep Learning*, Nature, Vol. 521.
57. Saon G. et al. 2015. English Conversational Telephone Speech Recognition System The IBM.
58. IJ Goodfellow, D Warde-Farley, M Mirza, A Courville, Y Bengio *Maxout Networks* arXiv preprint arXiv:1302.4389 URL: <http://arxiv.org/pdf/1302.4389.pdf> (дата обращения 05.02.2016).
- 59 Dario Amodei, Rishita Anubhai et al. 2015. *Deep Speech 2: End-to-End Speech Recognition in English and Mandarin* arXiv:1512.02595 (8 Dec 2015) [cs.CL]
60. John Nerbonne , Wilbert Heeringa 1997. *Measuring Dialect Distance Phonetically* Proceedings of the Third Meeting of the ACL Special Interest Group in Computational Phonology p. 11–18.
61. Nathan C. Sanders & Steven B. Chin. 2009. *Phonological Distance Measures*. Journal of Quantitative Linguistics Volume 16, Issue 1, p. 96–114.
62. Dunning, T. 1994. *Statistical Identification of Language*. Technical Report MCCS 94-273, New Mexico State University.
63. Sloos, Marjoleine and Marc van Oostendorp. 2012. The relationship between phonological and geographical distance: Umlaut on the diminutive in Dutch dialects. *Taal en Tongval* 62(2): p. 204—250.
64. Mahan Azadpour, Evan Balaban. 2008. *Phonological Representations Are Unconsciously Used when Processing Complex, Non-Speech Signals* PLoS ONE. 2008; 3(4): e1966. Published online 2008 April 16. doi: 10.1371/journal.pone.0001966 PMID: PMC2292097 URL: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2292097/citedby/> (дата обращения 05.02.2016).