М. Х. Шульман (shulman@dol.ru, www.timeorigin21.narod.ru)

Информация, смыслы, интеллект

(03.01.2021)

1 Введение

Давно и широко осуществляются попытки сформулировать, что такое информация и знание. Давно и повсеместно признано, что при общем подходе нельзя смешивать понятии “информация” и “количество информации” по Шеннону – последнее эффективно используется в теории связи, но не имеет никакого отношения к содержанию, или смыслу, информации.

В публикации [1] ее автор, отчасти ссылаясь на предшествующие исследования, указывает, что *смысл* сообщения, переданного системе, определяется ее *переходом (детерминированным или вероятностным) из одного состояния в другое* (и/или *изменением ее поведения*). Я считаю здесь ключевым понятие “состояние”.

Хорошо известно, что человеческий мозг “навязывает” владельцу доминирующую реакцию на *изменения* в окружающей среде. Характерный пример приводил несколько раз известный радиоведущий С. Доренко: “мы обычно не вслушиваемся в звук работающего двигателя самолета, но моментально реагируем, когда этот звук реально исчезает”.

Познание начинается с первичного восприятия и попыток расчленения воспринимаемого образа на части, затем – расчленения каждой части на более мелкие части, и т.д. Если такое (устойчивое) структурирование мозг выполнить не может, то он воспринимает наблюдаемое как хаос или как неразличимое *ничто* (например, “Черный квадрат” Малевича неограниченного размера). Если же структурирование хотя бы в какой-либо степени, возможно, наш мозг осуществляет его, и мы воспринимаем наблюдаемое как упорядоченное, структурно организованное, т.е. можем говорить о некотором правиле или законе, которому подчиняется воспринимаемое.

Более того, наш мозг может оперировать не только с материальными объектами, но и с абстрактными понятиями (идеями), которые он изначально представляет как *отдельные* сущности, и эти сущности он также организует тем или иным образом. Таким образом, по моему мнению, *самый главный закон природы состоит в том, что реальность (и материальная ее часть, и абстрактные конструкции) действительно и в высшей степени структурирована (по крайне мере, в той Вселенной, в которой мы живем)*!

Из этого вытекает, что информация, или знание, представляет собой отображение структуры той или иной части Вселенной. Оно и представляет собой описание состояния соответствующей структуры (системы).

Далее, как учит нас философия, есть понятия, имеющие одинаковое значение для любой науки. Это категории вещи, свойства и отношения. Всякая наука, каков бы ни был ее предмет, изучает вещи, их свойства и отношения [2]. Рассматривая эти категории совместно с представлениями о пространстве и времени, мы получаем необходимый инструментарий для описания структур в окружающем нас мире.

2 Информация

Получение информации системой связано с воздействием на нее внешней среды. Это воздействие может быть простым или сколь угодно сложным, одномоментным или сколь угодно протяженным во времени. Часто в физике или в практике противопоставляют получение информации и получение энергии/массы. Я считаю, однако, что *любой акт передачи энергии следует рассматривать как частный случай передачи информации, поскольку он неизбежно влияет на состояние/поведение системы-получателя* (например, сообщение газовому объему некоторого количества тепла изменяет распределение частиц газа по скоростям).

Обратимся теперь к очень важному вопросу – как происходит передача информации и что при этом происходит. Рассмотрим известную историческую ситуацию:

По радио сказали – «Над всей Испанией безоблачное небо». Один испанец после этого выбросил зонтик, а другой откопал в огороде пулемёт. Почему? Ведь оба получили одинаковый набор звуков…

С моей точки зрения, здесь допускается одна общепринятая ошибка. Этот “испанский” текст рассматривается как нечто самостоятельное и обособленное. Думаю, что именно здесь корень проблемы. Проблема решается тем, что кроме самого сообщения **обязательно** подразумевается его **контекст**, его окружение. И приписать **смысл** сообщению можно, только рассматривая сообщение с учетом контекста.

Когда тебе в руки попадает книга, ты не можешь ее прочитать, не зная алфавита и языка, на котором она написана. Когда ты требуешь от своего студента решить задачу по химии, он не сможет даже приступить к ее решению, не прочитав написанного тобой же учебника по химии. Когда ты пишешь свой учебник, ты опираешься на все предшествующие свои знания об электронах и атомах, об агрегатных состояниях вещества и пр.

Когда ты передаешь сообщение с помощью точек и тире, получатель должен знать, используешь ли ты азбуку Морзе или код Грея или еще какой-либо двоичный код. Когда ты шифруешь текст, используя некую книгу, получатель текста должен обладать этой книгой, чтобы расшифровать текст. Мы чаще всего игнорируем контекст, разговаривая о теории Шеннона применительно к свойствам отдельного сообщения, но это справедливо лишь в узких пределах!

Когда ребенок покидает материнскую утробу, несколько самых насыщенных первых лет его жизни посвящены накоплению и систематизации его опыта и знаний, и этот процесс продолжается в последующие годы. При этом не только наращиваются знания, но накапливаются также и варианты “перевода” этих знаний в аналогичные формы – зрительная/слуховая, вербальная/письменная, английская/французская. Только постепенно накапливая знания, человек получает доступ к анализу и пониманию частных сообщений. Вспомним про Розеттский камень…

Теперь мы можем попытаться уточнить, что же такое “смысл”. С моей точки зрения смысл – это отображение подмножества точек “фазового пространства” состояний и траекторий во Вселенной. Например, смысл конкретных событий в жизни конкретного человека в конкретный период времени (например, смысл предреволюционных событий в Испании). Тогда извещение по радио о “безоблачном небе” позволяет чуть-чуть переместиться во множестве смыслов, и это перемещение несет в себе небольшое количество информации (вопрос о ценности этой информации вне обсуждения). Это – как разложение в ряд Тейлора

в малой окрестности точки, но результат-то – новое положение изображающей точки, и он ничем не менее значителен, чем положение исходной точки. Т.е., если

$$f(x) \approx f(x_0) + k\,\Delta x,$$

то $k\Delta x$ – это то самое малое дополнительное приращение информации (сообщение), которое обретает смысл только при дополнительном учете основной составляющей $f(x_0)$.

3 Статистическая интерпретация семантики

Существует интересный подход к интерпретации смыслов, точнее – к установлению их эквивалентности и различия в языковых конструкциях. Как можно установить близость семантических элементов языка?

Ответ может быть получен с помощью статистики семантических конструкций. Например [3], откуда мы знаем, что у слов “лампа” и “светильник” схожее значение? Внешне эти слова совсем не похожи.

Будем придерживаться “дистрибутивной гипотезы”: Значение лингвистической единицы складывается только из ее использования. В мозгу хранится сумма всех тех контекстов, в рамках которых мы слышали или видели то или иное слово. *Это и есть его смысл*. Без знания *типичных соседей* никакой семантики нет. Отсюда вывод: слова с похожими типичными контекстами имеют схожее значение.

Теперь поступим следующим образом:

- Возьмем достаточно большой *корпус* (множество текстов) данного языка.
- Создадим *лексикон* (словарь) этого корпуса, т.е. множество из N слов, которые в нем встречаются хотя бы один раз.
- Для каждого слова из лексикона запоминаем, какие слова стояли с ним рядом, т.е. создавали его контекст.
- Охарактеризуем каждое слово лексикона (в нижеследующем примере N=5) набором частот совместной встречаемости исходного слова с другими словами из этого же лексикона:

Исходное слово	Встречаемость совместно с другими словами				
	“вектор”	“значение”	“хомяк”	“семантика”	“суслик”
“вектор”	0	10	0	8	0
“значение”	10	0	1	15	0
“хомяк”	0	1	0	0	20
“семантика”	8	15	0	0	0
“суслик”	0	0	20	0	0

Это означает, что, например, слово “вектор” 10 раз встречается в корпусе текстов совместно со словом “значение”, 8 раз – совместно со словом “семантика”, и ни разу – с другими словами этого списка.

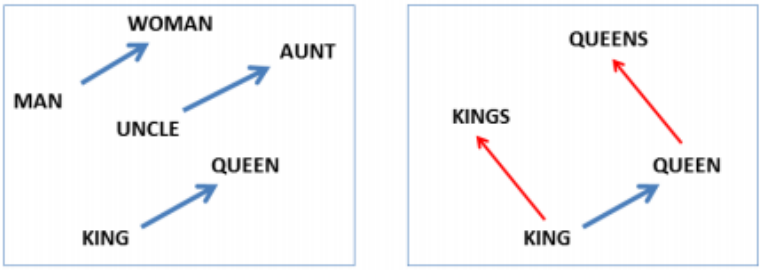
Теперь возьмем N-мерную систему прямоугольных координат и для каждого слова отложим точку в N-мерном пространстве. В данном примере[1] слово “вектор” будет иметь координаты <0, 10, 0, 8, 0> по соответствующим осям.

Что произойдет, если некоторые исходные слова будут иметь одни и те же или близкие координаты для каждого из других слов? Точки, изображающие эти слова, совпадут или окажутся рядом в таком N-мерном пространстве. Это и будет

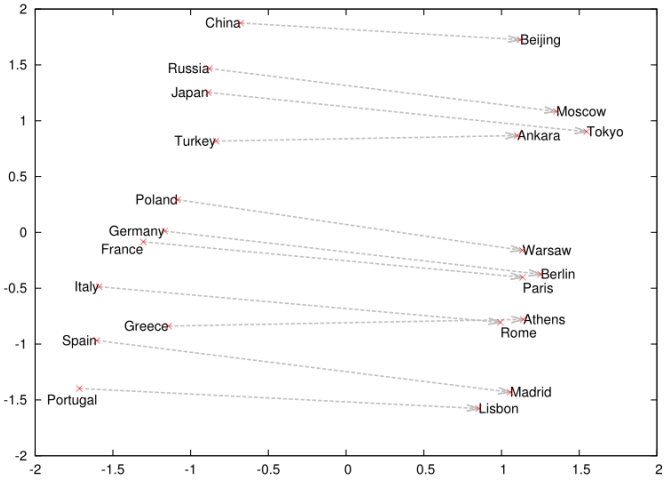
[1] В реальных случаях значение N может составлять десятки тысяч слов и более.

означать совпадение или близость их смысла. Классический способ определения семантической близости слов в векторном пространстве – через косинусную близость[2] векторов (схожесть сильнее по мере уменьшения угла между векторами слов, и наоборот).

Это открывает интересные перспективы. Фактически мы видим семантические отношения в системе языка, можем их “потрогать” [4].

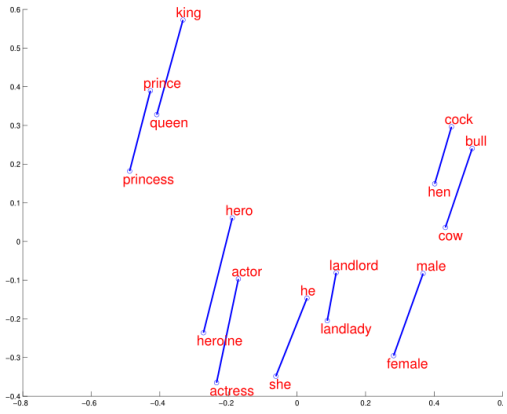


Например, “женщина” относится к “мужчине”, как “королева” к “королю”, и т.д.

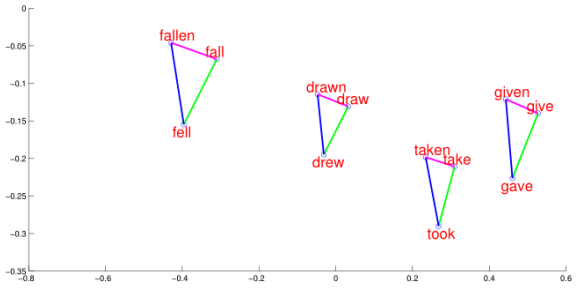


Географические отношения

Модели, натренированные на огромных корпусах (миллиарды слов), демонстрируют удивительные свойства.



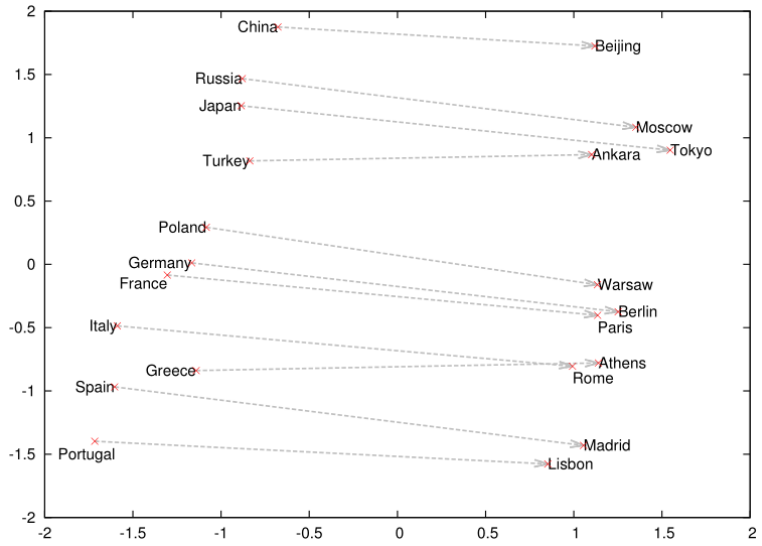
Однотипные семантические отношения



Однотипные отглагольные отношения

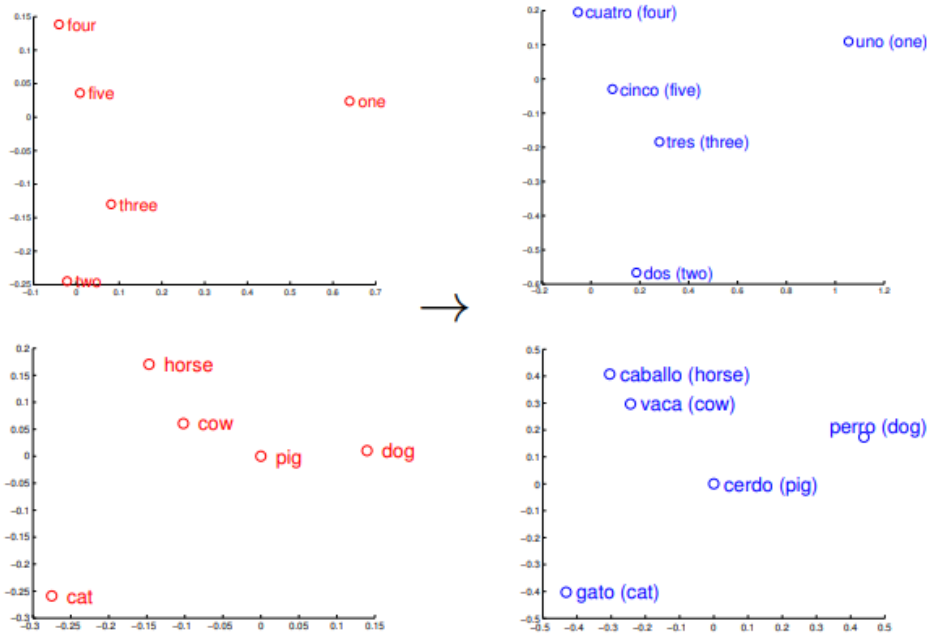
[2] При соответствующей переномировке компонент.

Алгебраические операции на векторах отражают операции семантические. Например, если мы вычтем из вектора слова “Париж” вектор слова “Франция” и прибавим слово “Германия”, получится вектор слова “Берлин” (точнее, он будет очень близким к получившемуся вектору).



Отношения страна - столица

Семантические структуры воспроизводятся даже в разных языках:



Распределение векторов слов (числительные вверху и названия животных внизу) в английском (слева) и испанском (справа) языках.

Аналогичные идеи закладываются поэтому в системы машинного перевода с языка на язык.

В работе [5] описывается программный комплекс «семантический процессор *Голем*», основанный на сходной идеологии и способный выявлять иерархии языковых паттернов при обучении на больших текстовых массивах. Значения двух слов считаются совпадающими, если они употребляются одинаково (могут заменять друг друга в синтагмах). Соответственно,

семантическая карта русского языка будет обобщать, в основном, статистику взаимного употребления основ Приведенная ниже таблица иллюстрирует содержание нескольких из 4000 ячеек такой семантической карты. Обучение проводилось на текстовом массиве объемом 6 ГБ, состоящем из материалов русскоязычных интернет-СМИ. Объем обучающей выборки примерно соответствует языковому опыту 20-летнего человека (при восприятии ~10^5 слов в день).

алексей	иванов	грузия	париж	вылетают	заявил
олег	васильев	армения турция	москва	отправляются	рассказал
				приезжают	говорит
александра	орлов	украина	лондон петербург	вышли	заявляет
николай	алексеев	эстония белоруссия		прилетели	
			город	посетили	
валерий игорь евгений татьяна	андреев николаев	узбекистан	берлин нью	выехали	говорил подтвердил
		азербайджан		уехали	напомнил
	александров		йорк		рассказал
	жуков				

Интересно, что для разбора предложений авторам не понадобились ни грамматические правила, ни определение частей речи, падежей, склонений, спряжений, времен и т.д. Вся необходимая информация о слове задается отнесением его к той или иной «букве синтаксического алфавита», а структура фразы определяется путем конкуренции между собой ~10^4 синтагм (языковых штампов).

Таким образом, даже в минимальном варианте семантический процессор можно использовать для автоматического анализа больших массивов текстовой информации – выявления и индексации различного рода фактов. Например: поездки и встречи публичных персон, кадровые перестановки, параметры сделок между собственниками и компаниями. Ключевой поиск плохо подходит для таких задач, т.к. один и тот же факт можно представить слишком большим числом способов. Семантическая индексация во многом облегчает задачу.

В работе [6] вводится мера *расстояния между двумя текстовыми документами*, позволяющая эффективно обнаруживать близкие по смыслу и тематике документы. Эта мера является минимумом расстояния между векторами соответствующих слов в обоих документах. Подобный подход позволяет поэтому решать задачи автоматической рубрикации и группировки реферативных документов.

A group of people shopping at an outdoor market. There are many vegetables at the fruit stand.

A stop sign is on a road with a mountain in the background.

A woman is throwing a Frisbee in a park

Нейронные сети умеют самостоятельно аннотировать картинки [7], т.е. распознавать изображение, устанавливать его смысл и генерировать аннотацию.

4 Что такое интеллект?

В живых и искусственных системах полученная и накопленная информация обычно аккумулируется в особых структурах – *памяти*. Состояния частей памяти, их взаимодействие и являются результатом (предварительного или окончательного) изменения состояния/поведения системы в целом.

Автор публикации [8] А.Д. Панов предлагает под системами искусственного интеллекта (ИИ) понимать автономные искусственные устройства, способные выполнять интеллектуальные функции. Он противопоставляет автономный ИИ инструментальным системам, которые являются лишь средствами усиления естественного человеческого интеллекта (его инструментами) - компьютерные системы автоматического проектирования (САПР), различные программы компьютерной графики, редакторы текстов, базы данных и т. д. Далее, Панов под *сильным* ИИ понимает такой ИИ, который превосходит человека или, по крайней мере, не уступает ему по всем интеллектуальным функциям, во всех отношениях. Он даже усиливает этот вывод: под сильным ИИ следует понимать такой ИИ, который превосходит *совокупную интеллектуальную мощность всего человечества по всем параметрам и даже во все времена*. Более того, сильный ИИ может начать саморазвитие со столь высокой скоростью, что люди не только окажутся лишними в этом процессе, но и принципиально не смогут за ним уследить и понять происходящее. Будущее для людей становится полностью непонятным и непредсказуемым. Эта ситуация и называется *технологической сингулярностью*.

В статье Панова изложены его соображения о проблеме искусственного интеллекта. Он считает, что линия аргументов, связанная с теоремой Геделя-Тьюринга и с теоремой Пенроуза об искусственном интеллекте, приводит к выводу, согласно которому модель сознания человека не может быть реализована в архитектуре *конечного автомата*. Поскольку же нейронная сеть, понятая как система пороговых переключателей, является именно конечным автоматом, то сознание, по мнению Панова, не может быть полностью реализовано в архитектуре нейросети.

Не оспаривая категорически этого вывода, я бы хотел, однако, привести такой аргумент. Указанные Пановым соображения базируются на представлении об *идеальном* интеллекте, полностью соответствующем концепции конечного автомата и однозначно – раз и навсегда – отвечающем на *любой* поставленный вопрос. Однако кто сказал, что каждый человек в отдельности и даже все человечество в совокупности (за все время своего существования) демонстрирует идеальный образец интеллекта? Более того, теорема Геделя имеет не только “запретительный” смысл, но и указывает путь к тому, как преодолеть запрет – для этого следует каждый раз (при очередном обнаружении неполноты исходной системы аксиом) расширить эту систему, притом необязательно единственным образом (например, в одной системе аксиом может не существовать множества с мощностью, лежащей между мощностью счетного множества и мощностью континуума, а в другой системе аксиом такое имеет место).

Как отмечал В.И. Арнольд в публикации [9], содиректор Математического института им. М. Планка в Бонне Ю.И. Манин определял математику как раздел филологии или лингвистики: это наука о формальных преобразованиях одних наборов символов некоторого конечного алфавита в другие при помощи конечного числа специальных "грамматических правил". Отличие математики от

живых языков состоит, по Манину, лишь в том, что в ней больше грамматических правил. Заметим, однако, что Ю.И. Манин имеет в виду чисто грамматический аспект языков, тогда на самом деле речь должна идти об их семантике. Как естественные языки, так и математика семантически представляют собой *открытые* системы в том смысле, что они могут бесконечно пополняться новыми смысловыми понятиями и конструкциями. Ровно так же строятся алгоритмические объектно-ориентированные языки программирования.

Что касается интеллекта животных, человека и человечества, то весь ход эволюции говорит о *фундаментальной роли процесса обучения в неуклонном расширении его границ* (т.е. как раз о регулярных переходах от одной – более узкой – системы аксиом к новой – более широкой). Т.е. мы, на языке теоремы Геделя, должны говорить не о *единственной* реализации конечного автомата, а о *семействе* все более мощных таких автоматов, открытом для развития[3]!

5 *Накопление информационного ресурса*

Если принять такую “открытую” точку зрения, то можно согласиться с позицией, подытоженной автором книги [10]: организмы суть алгоритмы (которые в общем случае могут определять не только *само поведение* организма и его популяции, но и формирование *правил* поведения за счет обучения). Каждое животное, включая Homo Sapiens, – это комплекс *органических* алгоритмов, сформированных естественным отбором за миллионы лет эволюции. Их функционирование не зависит от материалов конструкции, реализующей организмы-алгоритмы. Следовательно, нет оснований полагать, что алгоритмы органические могут что-то такое, чего неорганические алгоритмы никогда не смогут повторить или превзойти. Если вычисления правильны, то какая разница, где работают алгоритмы – в углеродной или кремниевой среде?

Генетическая схема едина у всех живых организмов на Земле. Все мы являемся потомками единственной клетки, существовавшей около 3.5 млрд лет назад. Она получила наименование LUCA и содержала порядка 300 генов с информационной емкостью порядка 1000 бит каждый [11].

По мере эволюции появились и развивались одноклеточные бактерии, в их геномах накопилось по несколько тысяч генов (до 10^7 бит). Суммарный информационный ресурс жизни за первые несколько млн лет существования жизни (при числе геномов ~10^7 геномов) достиг, таким образом, примерно 10^{14} бит информации (это несколько десятков терабайтовых дисков) и оставался таковым в течение 3 млрд лет.

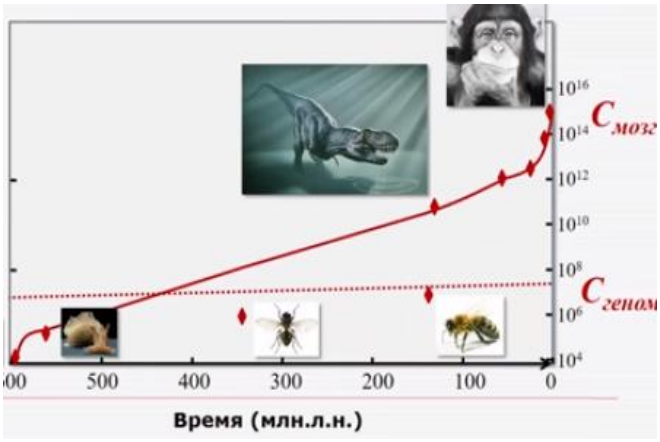
Но у одноклеточных бактерий есть предел сложности, у них по несколько тысяч генов, а у более сложных организмов – несколько десятков тысяч генов. Одноклеточные бактерии не могут больше расти, т.к. все их энергетически ресурсы пропорциональны поверхности. А расходуется энергия быстрее.

Но однажды одна бактерия проглотила другую, и та оказалась в объеме, поэтому ее энергетические ресурсы многократно возросли. Период “застоя эволюции” закончился. Позже, вместо того, чтобы программировать поведение, как у насекомых, жизнь, стала программировать способность к обучению.

Мозг возник, когда появились животные (около 800 млн лет назад). Растениям мозг не нужен, они добывают энергию из воздуха. Животному надо быть быстрым, чтобы искать пищу (и преследовать ее). Генотип изобрел машину, которая умеет учиться гораздо быстрее, чем сам генотип. 80% генов работают именно в мозге. Ниже на диаграмме показана эволюция информационного

[3] По-видимому, регулярное усложнение “автомата” обусловлено мутациями и естественным отбором.

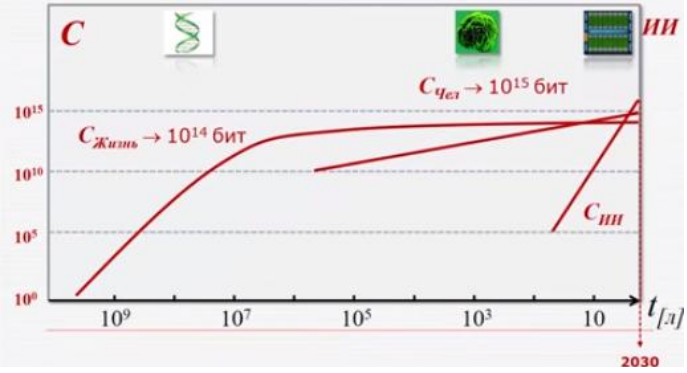
ресурса *C* суммарного мозга человечества за последние 600 млн лет эволюции [11].



Таким образом, мозг – это орган для обучения, он возник в результате генетического отбора. Алгоритмы *обучения* пришли на смену алгоритмам *поведения*. Люди качественно отличаются от животных тем, что у животных опыт накапливается только в течение их индивидуальной жизни, а у людей появился язык, и коллективный опыт наследуется. Скорость обучения человека составляет до 30 бит/с, накопленные за 10 лет знания одного человека – 10^{10} бит, накопленные знания всего человечества ~ 10^{15} бит. Человечество накопило уже больше информации, чем Земля за 3.5 млрд лет. Потому что мозг – это специфическая машина, которая работает гораздо быстрее, чем все генетические механизмы.

Кстати, известно, что мозг реализует еще один (как минимум) вид активности, который отвечает не за работу нейронной сети мозга, а за модификацию структуры этой нейронной сети. В ходе познавательных процессов мозг меняется. Хотя новых нейронов в нём не образуется, но изменяются связи между ними. В отличие от компьютера мозг способен заново прокладывать нервные пути всякий раз, когда усваивается новая информация. Поэтому развить или скорректировать когнитивные возможности можно.

Наконец, сейчас наступает эпоха искусственного интеллекта (ИИ). Она характеризуется тем, что скорость обучения в эпоху первоначальной эволюции жизни (которая привела к накоплению 10^{14} бит за ~ 3 млрд лет) и сильно возросшая в эпоху эволюции мозга животных и человека) стремительно возрастет благодаря появлению и эволюции ИИ (см. ниже диаграмму, также заимствованную из [11]). Т.е. искусственный разум – это очередная часть природы, научившаяся учиться.



Кратко остановимся на *термодинамике* эволюционного прогресса жизни на Земле. В соответствии с принципом Бриллюэна – Ландауэра, на производство одного бита информации требуется [12] энергии не менее $\triangle E \approx 2.8 \cdot 10^{-14}$ эрг.

Энергия, обеспечившая накопление информации на Земле, может поступать в конечном счете только от Солнца. В [13] отмечается, что поток мощности, приходящей от Солнца к Земле, составляет 10^{14} кВт, т.е. 10^{21} эрг/с. Период времени ~10^9 лет с учетом того, что год содержит ~ $3 \cdot 10^7$ с, составит $3 \cdot 10^{16}$ с.

В результате получим, что за это время Солнце передало Земле энергию $E \sim 10^{21}$ эрг/с x $3 \cdot 10^{16}$ с = $3 \cdot 10^{37}$ эрг. Разделив E на $\triangle E$, получим максимальную оценку для информации (в битах), накопленную земной биомассой: $C_{макс} < 10^{51}$ бит, т.е. на 35 порядков больше, чем приведенное в [11] значение.

6 Мозг и обучение

Нейросети с большим количество слоев помогают нам понять, как формируются образы высокого уровня в мозгу на бессознательном уровне, а сознательные процессы частично моделируются сетями с рекуррентными связями. Стоит задача построения сильного искусственного интеллекта – это роботы. Она не ограничивается решением конкретных задач, а предусматривает построение *психики* роботов.

Мы хотим понять архитектуру реального мозга. У всех позвоночных задняя часть мозга решает задачу “что происходит”, затем информация поступает в головной мозг, отражается, и в передней части решается задача “что делать”.

У рептилий есть древний мозг (базальные ганглии), там формируются цели, воля. У млекопитающих появилась кора (cerebral cortex / neocortex), в которой строятся модели мира и вырабатываются стратегии поведения. Когда мы рождаемся, кора “пустая”, в дальнейшем она постепенно заполняется знаниями. В коре происходит обучение “без учителя” (образы раскладываются на классы). Кора однородна, она вся работает по единому алгоритму. В ней происходит запоминание корреляций.

Кора организована иерархически – из первичных признаков последовательно формируются все более сложные признаки, например – признаки конкретного человека.

Кора формирует возможные варианты действия, а через положительную обратную связь с древним мозгом (базальные ганглии, таламус) происходит выбор лучшего варианта действия.

Все сигналы из внешнего мира (кроме запаха, т.к. кора возникла из области обоняния) идут в таламус из коры. И обратно в то же место посылается сигнал. Это есть усилительный контур, в котором возникают длительные автоколебания. Благодаря этому мы можем сосредоточиться на этом сигнале. Кроме того, через таламус возникает эффект сознания, синхронизация, понимание того, что происходит. Например, почему мы видим этот треугольник красным? Один участок фиксирует цвет, другие – понятие треугольности и формы, они связались, и возникло слово “треугольник”.

Итак, кора предлагает различные действия, в базальных ганглиях возникают оценки, а таламус разрешает или запрещает эти действия. Возникает окончательное решение. Остается мозжечок, второй контур управления.

Мозжечок тоже очень просто устроен. Он тоже однороден, значит, там есть единый алгоритм, он реализует “обучение с учителем”, мозжечок запоминает правильные решения, найденные корой (“решебник”, а не “задачник”). В коре поиск ответа может занимать полсекунды, а мозжечок находит запомненный ранее ответ (фактически это приобретенный рефлекс на уровне подсознания).

Как происходит обучение в мозжечке? Мозжечок состоит из множества очень мелких гранулярных клеток. В них собирается вся сенсорная информация через 200 млн Mossy fiber (мшистые волокна). Мозжечок получает совершенно случайные комбинации сигналов (хеширование) из коры – любой рисунок возбуждения в ней переходит в любой рисунок возбуждения в мозжечке. Дальше кора выработала решение что делать, а копия решения идет в мозжечок. Клетки Пуркинье, которые получают сигнал “как надо”, запоминают, как надо действовать в такой ситуации. Они загоняются в мозжечок как будто это рефлекторные решения [14].

7 Достижения нейросетевого ИИ

Еще недавно распознавание лиц было излюбленным примером того, что легко дается даже младенцам, но не по силам самым мощным компьютерам. Сегодня программы распознавания лиц способны идентифицировать людей гораздо быстрее и эффективнее, чем люди. Разведывательные и полицейские службы используют такие программы повседневно – для сканирования километров видеопленок с камер наблюдения и поиска подозреваемых и преступников [10].

Когда в 1980-х годах обсуждалась уникальность природы человека, в качестве главного доказательства его превосходства обычно приводились шахматы. Существовало убеждение, что роботу никогда не обыграть живого шахматиста. 10 февраля 1996 года суперкомпьютер Deep Blue разработки IBM взял верх над чемпионом мира Гарри Каспаровым, отправив на свалку и эту заявку на человеческое превосходство. Создатели хорошо подготовили Deep Blue, начинив его не только правилами игры, но и подробнейшими инструкциями относительно шахматных стратегий.

Искусственный интеллект нового поколения предпочитает советам людей машинное обучение. В феврале 2015 года компьютерная программа, разработанная подразделением Deep Mind компании Google, самостоятельно освоила 49 классических игр Atari. Один из ее разработчиков, доктор Демис Хассабис, объяснил: «Единственная информация, которая была предоставлена системе, – это необработанные пиксели на экране и идея, что нужно набирать очки. Все остальное она должна была просчитать сама». Программа ухитрилась выучить правила всех предложенных ей игр от «Пакмана» и «Космических захватчиков» до автогонок и тенниса. И играла в них не хуже, а иногда и лучше людей, подчас применяя стратегии, которые никогда не приходили в голову игроку-человеку.

Немного погодя искусственный интеллект отметился еще более сенсационным достижением: созданная DeepMind программа AlphaGo обучилась играть в древнюю китайскую настольную игру го, значительно более сложную, чем шахматы. Го долго считалась слишком сложной для искусственного интеллекта. В марте 2016 года в Сеуле состоялся матч между AlphaGo и южнокорейским мастером игры в го Ли Седолем. AlphaGo обыграла Ли со счетом 4:1, применив оригинальные стратегии, ошеломившие экспертов. Если до матча профессиональные гоисты почти не сомневались, что Ли победит, то проанализировав ходы AlphaGo, они пришли к выводу, что игра окончена: люди могут попрощаться с надеждой выиграть у AlphaGo и тем более ее потомков [10].

5 декабря 2017 года, команда поразила шахматный мир, объявив о своём алгоритме машинного обучения AlphaZero, который сумел овладеть не только обычными шахматами, но и японскими шахматами и игрой го. Алгоритм начал работу без какого бы то ни было понятия об играх, кроме базовых правил. Затем

он начал играть сам с собой несколько миллионов раз и учиться на своих ошибках. Всего за несколько часов алгоритм стал наилучшим игроком, как среди людей, так и компьютеров, из всех, что видел мир [15].

Даже те, кто управляет коллективами и различными видами деятельности, могут быть заменены. Благодаря своим мощным алгоритмам Uber координирует передвижения миллионов таксистов силами небольшого числа операторов. Основная часть распоряжений отдается алгоритмами без контроля со стороны человека. В мае 2014 года Deep Knowledge Ventures – гонконгский венчурный фонд, специализирующийся на регенеративной медицине, – удивил мир, сделав алгоритм по имени VITAL членом своего совета директоров. VITAL анализирует огромную базу данных, касающихся текущей финансовой ситуации, клинических испытаний и интеллектуальной собственности различных компаний, на предмет вложений. Наравне с другими пятью членами совета алгоритм голосует в поддержку или против инвестиций в то или иное предприятие.

Глядя на то, как проявил себя VITAL, понимаешь, что он уже успел перенять по крайней мере один порок управленцев: непотизм. Он неизменно голосовал за инвестирование в компании, предоставляющие алгоритмам значительную свободу действий. Например, с его благословения Deep Knowledge Ventures вложил средства в компанию Pathway Pharmaceuticals, использующую алгоритм OncoFinder для персонализированного подбора онкотерапии…

В сентябре 2013 года оксфордские ученые Карл Бенедикт Фрей и Майкл А. Осборн опубликовали статью «Будущее занятости», в которой исследуют вероятность замещения разных видов деятельности компьютерными алгоритмами в ближайшие двадцать лет. Алгоритм, разработанный Фреем и Осборном для расчета степеней этой вероятности, вычислил, что в США в зоне высокого риска находится 47 процентов профессий. Например, существует 99-процентная вероятность, что к 2033 году свои места алгоритмам уступят специалисты по телефонному маркетингу и страховые агенты, с 98-процентной вероятностью та же участь постигнет спортивных рефери, с 97-процентной – кассиров, с 96-процентной – шеф-поваров. Вероятность, с которой потеряют работу официанты, составляет 94 процента. Секретари юридических контор – 94 процента. Экскурсоводы – 91 процент. Пекари – 89 процентов. Водители автобусов – 89 процентов. Строительные рабочие – 88 процентов. Ветеринарные фельдшеры – 86 процентов. Охранники – 84 процента. Моряки – 83 процента. Бармены – 77 процентов. Архивариусы – 76 процентов. Плотники – 72 процента. Спасатели на воде – 67 процентов. И так далее, и так далее, и так далее…

Даже программирование системы на добро может привести к беде. По одному из популярных сценариев корпорация создает первый искусственный суперинтеллект и дает ему безобидное задание – рассчитать число π. Никто не успевает ничего понять, как искусственный интеллект захватывает всю планету, изничтожает весь человеческий род, распространяет свою власть до границ галактики и превращает Вселенную в гигантский суперкомпьютер, который миллиардолетие за миллиардолетием педантично, со все бóльшей точностью рассчитывает число π. Ведь такова была священная миссия, возложенная на него Создателем…

Не исключено, что технологии XXI века позволят внешним алгоритмам «взломать человеческую сущность» и узнать меня лучше, чем знаю себя я сам. Как только это случится, вера в индивидуализм рухнет и полномочия перейдут от отдельных личностей к сетевым алгоритмам. Люди больше не будут автономными единицами, устраивающими свою жизнь в соответствии со своими желаниями, а привыкнут воспринимать себя как совокупность биохимических механизмов, которые находятся под постоянным наблюдением и контролем сети алгоритмов.

Для того чтобы это произошло, не нужен алгоритм, который знает меня в совершенстве и никогда не ошибается; ему достаточно знать меня лучше, чем знаю себя я сам, и ошибаться реже, чем ошибаюсь я. Тогда у меня будет резон доверять этому алгоритму все больше моих решений – как мелких повседневных, так и жизненно важных.

В медицине мы уже перешли эту черту. В больницах мы больше не индивиды. Уже не вызывает сомнения, что вскоре основную часть самых важных решений о состоянии здоровья человека будут принимать алгоритмы, такие как Watson. И это совсем не обязательно плохо. Диабетики уже носят сенсоры, автоматически несколько раз в день измеряющие уровень сахара и при критических показателях подающие тревожный сигнал. В 2014 году исследователи из Йельского университета объявили о первых успешных испытаниях «искусственной поджелудочной железы», контролируемой айфоном. В эксперименте приняли участие пятьдесят два диабетика. В живот каждого пациента были имплантированы крошечный сенсор и миниатюрная помпа, соединенная с маленькими емкостями с инсулином и глюкагоном – гормонами, совместно регулирующими уровень сахара в крови. Сенсор непрерывно мониторил уровень сахара, передавая данные на айфон. В нем было установлено анализирующее информацию приложение, которое при необходимости посылало приказ помпе ввести точно рассчитанную дозу либо инсулина, либо глюкагона. Все это без какого-либо участия человека…

Если мы обеспечим Google и его конкурентам беспрепятственный доступ в наши биометрические устройства, к нашим ДНК-тестам и к нашим историям болезней, то в итоге получим всезнающий медицинский сервис, который не только поведет борьбу с эпидемиями, но и защитит нас от рака, инфарктов и Альцгеймера. Однако, располагая такой базой данных, Google сможет сделать еще больше. Представьте себе систему, от которой, как в известной песне группы The Police, не укроется ни один твой шаг, ни один твой вздох, ни одно твое слово; систему, которая мониторит ваш банковский счет и ваш пульс, ваш сахар и ваши сексуальные приключения. Разумеется, эта система будет знать вас гораздо лучше, чем вы сами знаете себя. Самообманы и самообольщения, удерживающие людей в плену дурных связей, нелюбимых профессий и вредных привычек, не введут Google в заблуждение. В отличие от комментирующего «я», которое управляет нами сегодня, Google не будет принимать решения, ориентируясь на мифы, и не попадется в ловушку когнитивных упрощений и правила «пика – финала». Он на самом деле будет помнить каждый наш шаг, каждое слово и каждое рукопожатие.

Многие из нас были бы рады переложить основную часть процесса принятия решений на такую систему или хотя бы в особо ответственных случаях консультироваться с ней. Google безошибочно подскажет, какой посмотреть фильм, куда отправиться в отпуск, что изучать в колледже, какое предложение работы принять и даже с кем закрутить роман и вступить в брак…

Исследование, проведенное недавно по заказу Facebook, показало, что уже сегодня фейсбуковский алгоритм разбирается в характерах и настроениях людей лучше, чем их друзья, родители и супруги. В эксперименте участвовало 86 220 волонтеров, которые имели аккаунт в Facebook и ответили на сто вопросов персональной анкеты. Алгоритм предсказывал ответы волонтеров, основываясь на анализе их «лайков» – то есть страниц, изображений и клипов, которые они отметили как понравившиеся. Точность предсказаний зависела от количества лайков. Предсказания алгоритма сравнивались с предсказаниями коллег по работе, друзей, членов семьи и супругов. Поразительно, но алгоритму хватило всего десяти лайков, чтобы превзойти в прозорливости коллег по работе. Ему

понадобилось семьдесят лайков, чтобы превзойти друзей, сто пятьдесят лайков, чтобы превзойти членов семьи, и триста лайков, чтобы превзойти супругов. Другими словами, если вам случится триста раз лайкнуть в Facebook, его алгоритм сможет предсказывать ваши мнения и желания лучше, чем ваши муж или жена [10].

И еще. Астрофизики создали нейросеть, которая делает то, чему её не учили. И делает это невероятно быстро. Подобный подход впервые опробовали астрофизики, и результат оказался не только потрясающим, но и немного пугающим.

Проект назвали Deep Density Displacement Model (D3M). Фактически это нейросеть, которая способна моделировать виртуальную вселенную. Если точнее, учёные сконцентрировались на гравитации во вселенной. Для обучения нейросети ей «скормили» 8000 подобных симуляций высокой точности, на которые обычно уходит несколько дней вычислений. На быстрые симуляции с более низкой точностью зачастую уходят минуты.

Однако астрофизикам каким-то образом удалось создать нейросеть, которая мало того, что способна создавать модель вселенной за 30 миллисекунд, так ещё и делает то, чему её не учили. В частности, учитывает изменения определённых параметров. К примеру, концентрации тёмной материи. Это, как если бы вы обучали программу распознавания изображений, используя фотографии кошек и собак, а она в итоге научилась бы определять слонов. Никто не знает, как это происходит. *Ширли Хо (Shirley Ho) — один из авторов проекта.*

Но и это ещё не всё. Оказалось, что относительная погрешность итоговой модели у нейросети D3M составила 2,8%, тогда как этот же показатель у моделей, созданных посредством метода быстрой симуляции, составляет около 9,3%. И учёные понятия не имеют, каким образом у них получилось то, что получилось.

Уточним, в рамках проекта моделировалась вселенная в виде куба с гранью в 600 млн световых лет. [16]

Литература:

[1] А.Л. Круглый. Модель смысла информационных сообщений. Материалы IV Российской конференции “Основы фундаментальной физики и математики”, РУДН, 2020 г., с. 14 – 19.

[2] А. И. Уемов. Вещи, свойства и отношения. Издательство АН СССР, М., 1963. С. 185.

[3] А. Кутузов. Нейронные языковые модели в дистрибутивной семантике. ВШЭ. 07.03.2015 https://www.youtube.com/watch?v=7k_MOBYbw_w

[4] Tomas Mikolov et al. Learning Representations of Text using Neural Networks. NIPS Deep Learning Workshop 2013. http://www.machinelearning.ru/wiki/images/9/96/MikolovNIPS-DeepLearningWorkshop-NNforText.pdf

[5] С.А. Шумский. Мозг и Язык: Гипотеза о строении “органа языка”. **2001 г.** Физический институт им.П.Н. Лебедева РАН, Москва.

[6] J. Kusner et al. From Word Embeddings To Document Distances. Proceedings of the 32 nd International Conference on Machine Learning, Lille, France, 2015. JMLR: W&CP volume 37. https://mkusner.github.io/publications/WMD.pdf

[7] Григорий Сапунов. Введение в архитектуры нейронных сетей http://ai-news.ru/2017/10/vvedenie_v_arhitektury_nejronnyh_setej.html

[8] А.Д. Панов. Технологическая сингулярность, теорема Пенроуза об искусственном интеллекте и квантовая природа сознания. Приложение к журналу «Информационные технологии» № 5/2014. http://www.intelros.ru/pdf/metafizika/2013_3/7.pdf

[9] В.И. Арнольд. — Математическая дуэль вокруг Бурбаки‖. Вестник РАН, 2002, том 72, № 3, с. 245-250
[10] Ю. Н. Харари. Homo Deus. Краткая история будущего. Издательство “Синдбад”. 2018 г.
[11] С.А. Шумский. Эволюция разума. Лекторий фонда “Траектория”, 08.04.2017. https://elementy.ru/video/316/Evolyutsiya_razuma
[12] L. Herrera. The mass of a bit of information and the Brillouin’s principle. arXiv:1403.4511v1 [gr-qc] 18 Mar 2014
[13] А.А. Матвеева. Томский политехнический институт. Солнечная энергетика. http://portal.tpu.ru/SHARED/n/NASA/Education/NiVIE/Tab/p2.pdf
[14] С.А. Шумский. Моделирование работы мозга. Лекторий фонда “Траектория”, 18.08.2017. https://elementy.ru/video/355/Modelirovanie_raboty_mozga
[15] С. Строгач. Один гигантский шаг для машины, играющей в шахматы. https://habr.com/ru/post/436598/
[16] ScienceDaily, PNAS. https://www.ixbt.com/news/2019/06/29/astrofiziki-sozdali-nejroset-kotoraja-delaet-to-chemu-ejo-ne-uchili.html