дежавю, нейросети и спам

крис касперски, ака мыщъх, no-email

что такое дежавю наверняка знает каждый, а кто не знает — тот может прочитать на вике, выбрав из кучи гипотез наиболее симпатичную и привлекательную (какая разница что выбирать? все они не состоятельны), поэтому представляет интерес поговорить о том, чего на вике нет, рассмотрев феномен дежавю с точки зрения последних достижений кибернетики.

чувство нереальности происходящего, зачастую сопровождаемое беспричинным страхом и ощущение, что все это уже когда-то было, а теперь повторяется вновь, и мы даже знаем, что в следующее мгновение произойдет это дежавю — бесценный дар природы, мощнейший механизм предсказания вероятностных событий, которого нужно не бояться, а учиться управлять им!

введение

Дежавю — французское слово, точное даже целых два! Deja — "уже", vu – "видеть", соединив их вместе получаем "уже (как бы) увиденное". Причем, "как бы" это не мышиная отсебятина! Deja активно используется в значении "как же это…", например, "comment vous appelez-vous, deja?" — переводится "блин, как же вас зовут? я помнил, но забыл".

Термин "дежавю" ввел в обиход французский психолог Эмиль Буарак (Emile Boirac), опубликовавший на последнем курсе обучения в университете эссе "L'Avenir des sciences psychiques" ("Будущее психологии") и впервые описавший психологическое состояние, при котором возникает устойчивое ощущение, что это уже было, человек как бы заново переживает уже пережитое, не только вспоминая прошлое, но и предугадывая будущее (причем, в большинстве случаев довольно успешно). В той же книге описывались и другие состояния: deja-vecu (уже пережитое), deja entendu (уже слышанное) и jamais vu (никогда не виденное), однако, они не прижились, а вот дежавю просочилось в массовую культуру, попав на страницы литературных романов и экраны телевизоров, впрочем, в довольно искаженном и перевранном виде, не имеющего ничего общего с реальным феноменом, который по разным оценкам хотя бы однажды испытывали свыше 70% человек.

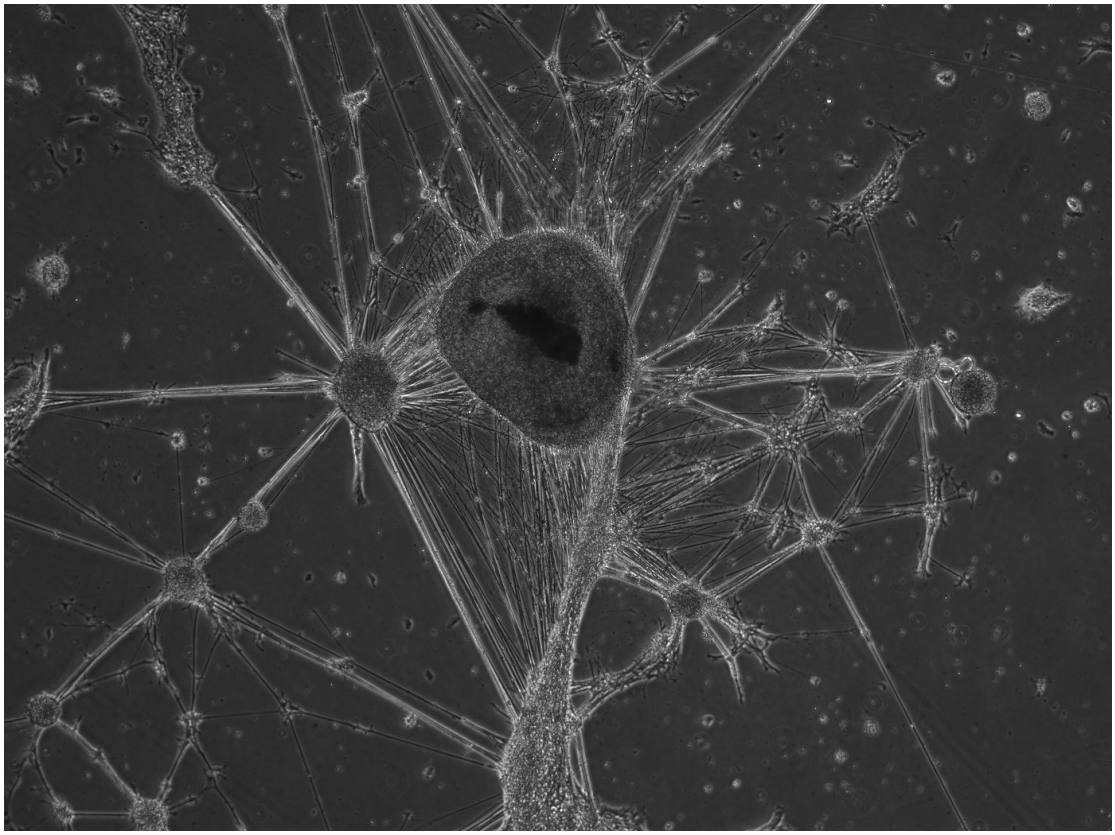

Рисунок 1 мы знаем как выглядит нейрон, но не имеем никакого представления о том, как работает человеческий мозг в целом

наука и схоластика

Реакция ученых мужей пришла с большим запозданием. Сначала на дежавю вообще не обращали никакого внимания, потом попытались списать его на психические расстройства (мол, с больной головой еще и не такое привидится), когда же количество "больных" превысило всякие разумные пределы, ученым пришлось пересмотреть свои позиции, перестать махать руками, топать ногами и трясти бородой с усами, а тщательно и кропотливо изучать с чем они имеют дело, а дело это оказалось намного более сложным и запутанным, чем можно было предположить поначалу.

С момента публикации "L'Avenir des sciences psychiques" прошло свыше ста лет, а гипотезы, объясняющие природу дежавю, продолжают плодиться как австралийские кролики в брачный период сезона дождей. Тем временем, отрывается все больше и больше книг, описывающих дежавю задолго до Бурака — от научных трудов до литературных произведений, убедительно доказывающих, что дежавю — вовсе не выдумка, а реальность, с которой приходится считаться вопреки сложившимся стереотипам.

Ощущение "уже пережитого" — само по себе всего лишь чувство, легко объяснимое с научной точки зрения без привлечения дополнительных теорий. Кому-то мерещатся черти, кому-то зеленые человечки, а кто-то уверен, что он здесь уже был и все видел. Вот только… в отличии от чертей, в существовании которых трудно убедить окружающих (особенно скептиков), дежавю позволяет человеку предсказывать будущее, в том числе и "нелогичные" события, не объявляющие о своем приходе и наступающие совершенно внезапно. Поскольку, вероятность (и точность!) подобных предсказаний далеко выходит за рамки простой проницательности, то экспериментаторы всеми силами стараются игнорировать подобные факты, опасаясь, что в противном случае научное сообщество обвинит их в грубой фальсификации.

Еще бы! Ведь будущее предсказать невозможно! Это — аксиома! От которой ученые (не все, конечно, но большинство) и начинают "плясать", натягивая экспериментальные данные на уже построенную кривую ожидаемого результата. Другими словами, в "подлинной" науке намного больше схоластики, чем в самой схоластике, поскольку, вместо изучения феномена

дежавю, ученые занимаются сбором фактов, подтверждающих его отсутствие, автоматически отбраковывая все остальные.

На самом деле, тезис о непредсказуемости будущего мягко говоря несостоятелен, вернее, применим лишь к сравнительно небольшому количеству ситуаций. Предсказать сколько очков выпадает на честной игровой кости не возьмется даже продвинутый астролог, поскольку, кость не имеет памяти — последующий бросок никак не связан с предыдущим и описывается исключительно теорией вероятности, в которой все грани равны. Аналогичным образом дела обстоят с рулеткой, спортлото и другими азартными играми, но вот в реальной жизни… абсолютная вероятность наблюдается разве что в специальных физических экспериментах типа распада атомов. Подавляющее большинство событий возникают не сами по себе, а в соответствии с причинно-следственной связью, что делает их вполне прогнозируемыми, особенно, если система проходит через серию состояний, уже наблюдавшуюся в прошлом. Достаточно очевидно, что если череда событий A -> B -> C ранее привела нас к состоянию D, то и сейчас состояние D (которому предшествовали события C, B, A) более вероятно, чем любое другое. Насколько более вероятно? Это зависит от степени детерминированности системы, ее контекстной чувствительности (так же называемой локальной памятью) и полноты наших представлений о ней.

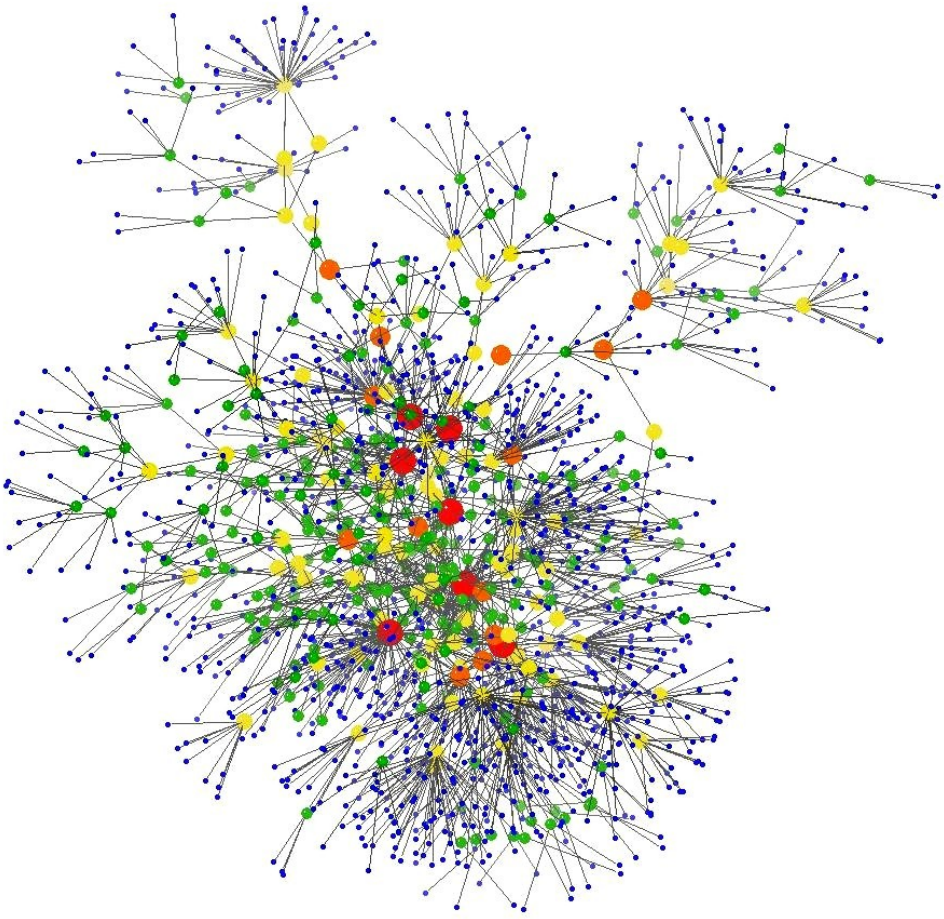

Рисунок 2 причинно-следственная модель событий повседневной жизни в графическом виде

Некоторые системы проходят через каждое состояние лишь однажды, но таких меньшинство (например, движение трех и более гравитирующих тел по незамкнутым и непериодическим орбитам). Реально же нас окружают системы, поддерживающее четкое соотношение между входными и выходными данными. Это и есть причинно-следственная связь в чистом виде (пошла я вчера в парк — изнасиловали, сегодня пошла – изнасиловали, завтра опять пойду…).

За исключением простейших ситуаций, наше сознание не в состоянии обработать огромный массив информации, и на основе накопленного опыта сделать верное и логически

обоснованное предсказание. Подсознание — другое дело! Конечно, это всего лишь гипотеза, однако, во-первых, она очень хорошо согласуется с экспериментальными данными (дети, не имеющие жизненного опыта, с дежавю незнакомы; а вот подростки переживают его достаточно часто, но по мере взросления, жизненный опыт становится все труднее и труднее упорядочивать, подсознание оказывается не в состоянии найти в памяти аналогичную ситуацию, и потому ощущение дежавю посещает нас все реже и реже). Во-вторых, это единственное рациональное объяснение происходящего. Если мы предсказываем будущее не на основе накопленных данных, то… как его можно предсказать еще?!

Но, как бы там ни было, вместо того, чтобы отмахиваться от широко распространенного феномена, лучше научиться использовать его, доверившись своим чувствам. У нас есть все основания утверждать, что они не лгут.

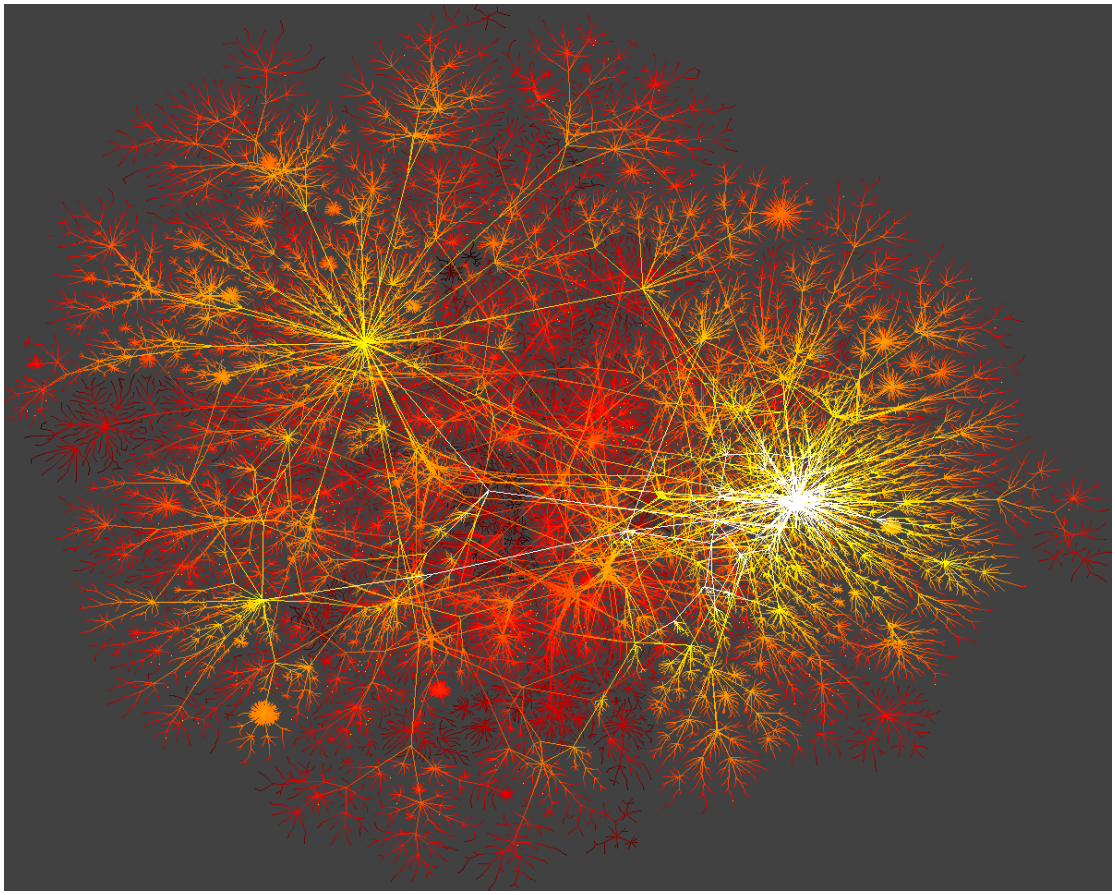

Рисунок 3 фейерверк?! а вот и нет! это модель нейросети, устроенной по образу и подобию той, что находится внутри нашей черепной коробки

нейросети

Нейросети, моделирующие человеческий разум (или то, что мы себе под этим представляем), довольно успешно работают в биржевых программах, системах предсказания погоды и цен на авиабилеты, причем цены намного более переменчивы, чем турбулентная атмосфера нашей планеты, но даже их предугадывают нейросети.

Как они это делают? Обращаются к духам и прочим высшим силам? Конечно же нет! Нейросеть всего лишь запоминает последовательность смены состояний и если аналогичный шаблон повторяется вновь, нейросеть "вспоминает" последующие состояния, возможно, корректируя их с учетом специфики текущей картины. И эта схема работает! Пускай не без ложных выпадов и грубых ошибок, но в целом вероятность наступления полученного результата существенно превышает "фифти-фифти".

Так почему бы не предположить, что человеческий мозг, по образу и подобию которого строятся нейросети, способен прогнозировать наступление событий, опираясь на жизненный опыт? Напротив, было бы очень странно, если бы наш мог этого не умел. Предвидение —

мощный инструмент в борьбе за выживание и потому его возникновение вполне обоснованно. Когда именно он возник сказать сложно. Вероятнее всего на ранних эволюционных стадиях, когда рациональное мышление находилось в зачаточном состоянии и пра-человеком управляли главным образом бессознательные предчувствия и инстинкты.

Сознание появилось лишь когда выживание стало в значительной степени зависеть от изобретательных навыков (добыча огня, изготовление орудий труда, etc), а жизненный опыт превратился в бесполезный балласт. Именно отказ от шаблонов позволил древним совершить качественный скачок вперед, потому что, даже матерым волкам известно, что "нельзя за флажки" только молодым и зеленым, а опытным не только можно, но и нужно, потому как за флажками – свобода, а там где флажков нет — стоят нехорошие люди с дробовиками, берданками, а кое-кто даже с BFG, выдранным из дума.

Какое это отношение имеет к дежавю? Да самое прямое! Современный человек представляет собой рационально мыслящее существо, управляемое сознанием и нервным импульсам океана бессознательного очень трудно пробиться наверх, да и те что пробились отметаются рациональным мышлением, как нечто нелогичное, а потому неверное. С другой стороны, если у человека развита эмоциональная память, подсознание может воздействовать на нее, накладывая события давно минувших дней на текущую ситуацию, в результате чего у нас возникает чувство (между прочим, совершенно не подложное), что все это уже когда-то было и вот в душе уже шевелиться предчувствие, что произойдет в следующий момент. На самом деле, это не предчувствие, а результат деятельности естественной нейросети, которая проходит через уже знакомое ей состояние, и в памяти тут же вспыхивают цепочки ассоциаций, ведущие нас к ранее пережитым чувствам, которые мы, проецируя на окружающую действительность, относим к будущему времени, хотя в действительности, это некоторая комбинация прошлой памяти с экстраполяцией поправки на текущую ситуацию.

Другими словами, дежавю это не просто "магазинная" память в чистом виде, это прогноз событий, учитывающих как пережитые события, так и накопленный за это время опыт (даже компьютерные нейросети способны обучению и по мере "взросления" ошибаются все реже и реже, чего уж говорить за человеческий мозг, который в этом плане вообще непревзойденный монстр).

Почему же тогда чувство дежавю так редко возникает? Теоретически, нейросеть должна совершать большое количество предсказаний, ведь через большинство событий, разворачивающихся на жизненной арене, мы проходим неоднократно, снова и снова наступая на те же самые грабли. Где наше подсознание?! Оно что, совсем уснуло?! На самом деле, нейросеть снабжает нас предсказательной информацией постоянно и мы используем ее в процессе принятия решений, сами того не замечая. Доказать это предельно просто. Достаточно попробовать хотя бы неделю жить, опираясь только на рациональное мышление и посмотреть насколько больше промахов мы совершим. Да что там реальная жизнь! Взять хотя бы компьютерные игры! Где-то нужно прыгнуть, а где-то наоборот — притормозить, причем, _зачем_ это делать (при прохождении лабиринта в первый раз) с рациональной точки зрения объяснить невозможно, даже если тянуть объяснение за уши со страшной силой. А вот если вспомнить, что лабиринт, составленный человеком, несет в себе отпечаток его натуры и потому обладает определенной степенью предсказуемости, все сразу становится на свои места. Нейросеть выявляет скрытые закономерности, распознавая шаблоны и выдавая довольно достоверные предсказания, помогающие игроку добиться намного лучшего результата, чем следует из простых логических (умо)заключений.

Ну и чем это не дежавю? Ведь каждый игрок испытывает бессознательные ощущения, что здесь лучше повернуть налево, а не направо, эту аптечку лучше не брать, так как за ней наверняка скрывается монстр, от которого потом так просто не отстреляешься и т.д. Просто, погруженные в игровой процесс мы не придаем этим чувствам особого значения, точно так, как мы не слышим тиканье часов. То есть еще как слышим! Временами… во внезапно наступившей тишине, когда, например, любимая девушка говорит, что уходит к другому…

С другой стороны, нейросеть на низшем бессознательном уровне не способна эффективно отделять мух от котлет. То есть, если хмурым осеним деньком, шли мы в школу мимо высокого кирпичного забора, вдыхая "аромат" тлеющих листьев, и вдруг в забое обнаружилась дыра, а за дырой — очень большая и совсем недружелюбная собака, которая нас не по детски напугала (покусала), то нейросеть будет помнить все: и осень, и дым от кострищ, и высокий забор, срабатывая только при полном совпадении шаблона, хотя с рациональной точки зрения дым тут совсем не причем и главное — высокий забор, каким ограждают себя злобные новые русские, заводящие таких же злобных собак, стерегущих награбленное и вероятность появления собаки из-за высокого забора намного выше, чем из-за низкого и покосившего, за

которым находится хижина старой бабушки, у которой и сторожить-то нечего. Только ведь подсознанию этого не объяснишь… Именно потому, собственно говоря, чувство дежавю иногда бывает столь пугающим. Идем мы себе по тротуару, никого не трогаем, вокруг все спокойно, никаких источников угрозы. Ну осень, ну забор, ну дым. Так почему же нас внезапно охватывает странный панический иррациональный страх?! Откуда это устойчивое чувство тревоги?!

Нелогично. Да, нелогично. Но иногда подобные предсказания все-таки сбываются. В заборе обнаруживается дыра, а за дырой — собака. Дым, осень — откуда нам знать, какое значение это имеет для атаки? Быть может, совершенно никакого, а быть может — осенью собаки особенно злы и раздражительны. Дым? Так ведь это не просто дым! Это признак изменения атмосферного давления, а собаки к нему чувствительны. Откуда такие заключения? При устойчивом антициклоне дым поднимается вверх, костры горят ярко и совсем не воняют. Если же дым стелется по земле, а угли быстро покрываются золой, значит, надвигается циклон или что-то похуже того. Следовательно, даже на рациональном уровне анализа мы не можем точно сказать какие факторы в наибольшей степени управляют поведением собаки, а раз так, то ни один из них нельзя отбрасывать. Именно так подсознание и поступает, что несет в себе не плюсы, но и минусы, причем не известно чего больше. Чем больше деталей включается в текущий контекст тем выше вероятность "промахов" (непредсказанных ситуаций), но иметь ложные позитивные срабатывания по сто раз на дню — еще хуже. Человек тогда просто не будет обращать никакого внимания на свои предчувствия.

К счастью, нейросеть способна обучаться и не только обучаться, но и использовать эффективные методики оценки вероятности наступления (или не наступления) заданной ситуации. И вот тут мы плавно переходим к теорема Байеса и проблеме фильтрации спама.

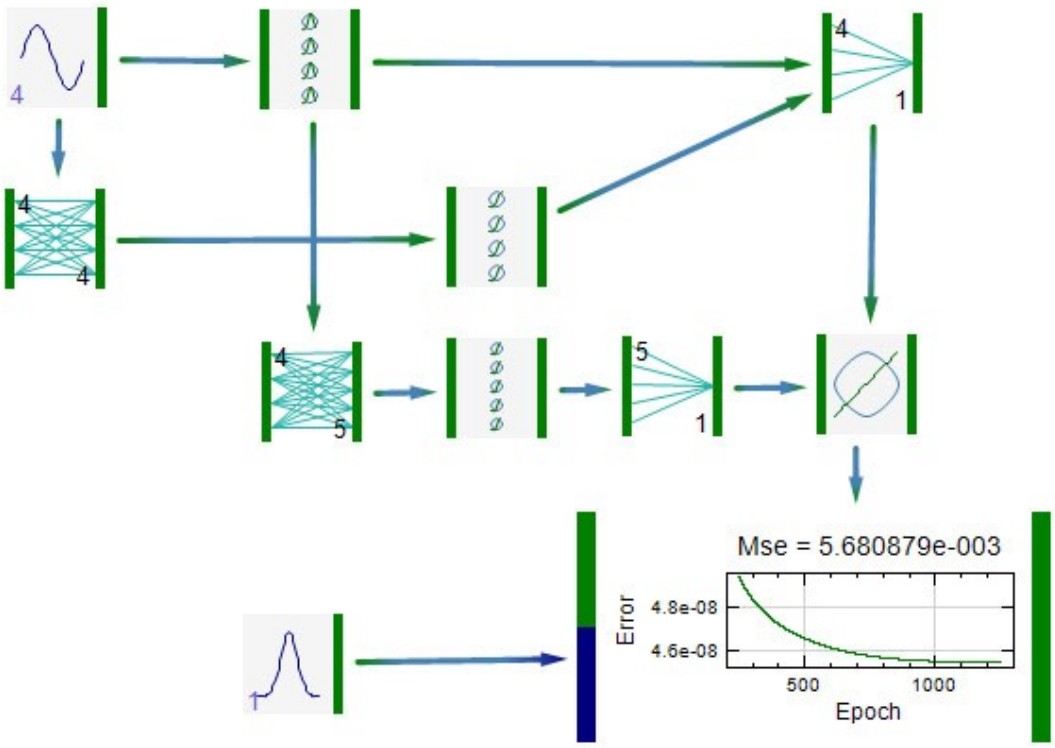

Рисунок 4 обученная нейросеть способна прогнозировать будущее, оценивая вероятность наступления тех или иных событий

математика, гармония и интуиция

Вика сообщает, что "Теорема Байеса — одна из основных теорем элементарной теории вероятностей, которая определяет вероятность наступления события в условиях, когда на основе наблюдений известна лишь некоторая частичная информация о событиях. По формуле

Байеса можно более точно пересчитывать вероятность, беря в учет как ранее известную информацию, так и данные новых наблюдений".

Сама же формула (записанная в сокращенном виде) выглядит следующим образом:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Рисунок 5 сокращенная формула Байеса

где:

- P(A) - априорная вероятность гипотезы A;
- P(B) - вероятность наступления события B;
- P(A | B) - вероятность гипотезы A при наступлении события B;
- P(B | A) - вероятность наступления события B при истинности гипотезы A;

В переводе на хакерский язык это означает, что мы получаем мощный инструмент обратной трассировки, той, которую поддерживал Soft-Ice под MS-DOS, а сейчас разве что некоторые эмулирующие отладчики, позволяющие обратить ход исполнения инструкций вспять, определив какой именно причиной вызвано срабатывание заданного условного перехода, что существенно упрощает взлом программ, делая его тривиальной задачей даже для начинающих. Но вернемся к Байесу, формула которого преследует аналогичные цели. С ее помощью мы можем оценить вероятность того, что событие B действительно вызвано причиной A. Другими словами, вместо того, чтобы идти от причины к следствию, по известному следствию мы выбираем наиболее вероятную причину. Подобный математический аппарат замечательно работает как в искусственных, так и в естественных нейросетях, реализуя механизм самообучения.

Полная формула Байеса — намного более мощная штука, дающая вероятностную оценку наступления заданного события, зависящего от нескольких независимых причин (как, например, в случае с забором, дымом и собакой).

$$P(B) = \sum_{i=1}^{N} P(A_i)P(B|A_i)$$

Рисунок 6 полная формула Байеса

Таким образом, обученная нейросеть способна выявить совокупность причин, приводящих к наступлению некоторого события, не вдаваясь в "физический" смысл происходящего, и оперируя одной лишь вероятностью. При достижении некоторого порогового уровня, в черепной коробке врубается защитный механизм, сигнализирующий о том, что сейчас должно произойти то-то и то-то.

Чем выше вероятность наступления B, тем сильнее наше предчувствие, подвергаемое "цензуре" рационального анализа. Если с его точки зрения все ОК, то мы даже не замечаем работы, проделанной подсознанием, поскольку, предчувствие опирается на привычный для нас логический аппарат. А вот если обосновать причины наступления B с позиции "здравого смысла" никак не удается — возникает устойчивое чувство нереальности происходящего, словно мы переживаем уже пережитое. Логика тихо курит в сторонке, оставляя нас наедине со своими чувствами. Дежавю!

Рисунок 7 Байес давно уже как почил, а его формула до сих пор живет!

Кстати, формула Байеса нашла применение в спам-фильтрах, позволяющих оценить вероятность принадлежности письма к спаму без всякого лексического и семантического анализа текста, то есть без анализа "физического" смысла послания. Поразительно, но качество "тупых" Байесовских фильтров вплотную приближается к "человеку мыслящему" и сквозь них проскальзывают только те послания, которые ставят в тупик не только машину, но и получателя, вынужденного прочитать весь текст письма, прежде чем до него дойдет, что это спам.

заключение

Дежавю — вовсе не психологическое расстройство! И нет ничего мистического в удачных предсказаниях еще не наступивших событий. Логика не единственный (и, вероятно, не самый удачный) инструмент познания мира. Вероятностная оценка наступления событий намного надежнее логического анализа, особенно если физическая природа происходящего ясна не до конца или же вообще неизвестна.

К сожалению, повальное увлечение логикой и отказ от веры в собственные чувства привели к тому, что современный человек выбирает далеко не лучшую стратегию поведения из всех, предлагаемых ему (под)сознанием. Отказ от логики позволил нейросетям существенно повысить степень своей "проницательности". Так чем же мы, люди, хуже?!

Рисунок 8 Xiao Xiao Li — очаровательная сотрудница Microsoft Research, разрабатывающая Байесовские фильтры против спама