Как нейронные сети объясняют принципы мышления людей, животных и инопланетян.

Перевод Сергея Рачинского. Иллюстратор Игорь Лобанов.

As neural networks explain the principles of thinking people, animals and aliens.

Translation Sergei Racha. Illustrator Igor Lobanov.

Фрэнк Вильчек

Физик-теоретик, лауреат Нобелев­ской премии по физике 2004 года, профессор Массачу­сет­ского технологического института, «За открытие асимптотической свободы в теории сильных взаимодействий», а также медали Лоренца и премии Юлиуса Эдгара Лилиенфельда.

Frank Wilczek

A theoretical physicist, winner of the Nobeltion Prize in Physics in 2004, professor of the Massachusetts set Skog Institute of Technology "for the discovery of asymptotic freedom in the theory of strong interactions", as well as the Lorentz Medal and Prize by Julius Edgar Lilienfeld.

 «Когда я только начинал заниматься фортепиано, каждое извлечение нужной ноты, каждое попадание по нужной клавише требовало моего пристального внимания. Но чем дольше я учился, тем проще мне было играть музыкальными фразами и аккордами. Шло время, я воспроизводил все более сложную музыку, затрачивая на это все меньше осознанных усилий. Очевидно, в моем мозге произошли какие-то мощные изменения.

История, конечно, вполне заурядная. То же самое происходит с каждым из нас, когда мы учим новый язык, осваиваем новую игру или просто приспосабливаемся к новой для нас среде. Вполне вероятно, что во всех этих случаях действует один и тот же механизм, и мне кажется, в общих словах этот механизм можно описать как создание скрытых слоев. Как говорится, лучше один раз увидеть.

На рисунке 1 потоки информации идут сверху вниз. Сенсорные нейроны — глазные яблоки в верхнем слое — получают сигналы из внешнего мира и кодируют их в удобную для прочтения в рамках данной системы форму. В биологических нейронных сетях (человеческом мозге) это означает преобразование сигнала в электрические импульсы, в искусственных нейронных сетях, которые моделируют мозг, — оцифровка сигнала. Кодированные таким образом сигналы передаются нейронам в следующий слой, расположенный ниже. Эффекторные нейроны — звездочки в нижнем слое — посылают свои сигналы «устройствам вывода данных» (для биологических нейронов это обычно мышцы, для искусственных — компьютерный терминал, с помощью которого с устройством взаимодействует пользователь). Посередине же находятся нейроны, которые не связаны напрямую с внешним миром — они не получают из него информацию и никак на него не воздействуют. Они взаимодействуют только с другими нейронами. Это и есть скрытые слои.

В первых искусственных нейронных сетях никаких скрытых слоев не было, и выходные данные были относительно простой функцией входных данных. Такие двухслойные «перцептроны» (от лат. perceptio — «восприятие», математическая и компьютерная модель восприятия информации мозгом. — Esquire), действующие исключительно на «вход-выход», были крайне ограничены в своих возможностях. К примеру, невозможно сконструировать перцептрон, который, столкнувшись с несколькими черными кругами на белом фоне, был бы в состоянии посчитать количество этих кругов. Лишь в 1980-е, спустя многие годы после первых работ в этой области, ученые поняли, что включение хотя бы одного или двух скрытых слоев поразительно расширяет возможности искусственных нейронных сетей. Сейчас такого рода многослойные сети используются, например, для анализа столкновений частиц высоких энергий в Большом адронном коллайдере. Человеку и не снилась та скорость и надежность, с которой они это делают.



Рис. 1

В 1981 году Дэвид Хьюбел и Торстен Визел получили Нобелевскую премию по физиологии и медицине за то, что открыли механизм действия нейронов в зрительной зоне коры головного мозга. Они показали, что в скрытых нейронных слоях последовательно извлекаются наиболее информативные свойства визуальных сигналов (например, резкие смены яркости или цветов, свидетельствующие о границах объекта), а потом складывают их в единое целое (собственно, объекты).

Каждое мгновение нашей взрослой жизни, не занятое сном, мы преобразуем множества фотонов, которые приходят с разных сторон, из множества неупорядоченных источников и попадают на двухмерную сетчатку, в привычный нам трехмерный мир. Поскольку это не требует с нашей стороны никаких осмысленных усилий, мы склонны не замечать это «обыкновенное чудо» и принимать его как должное. Но когда инженеры попытались воспроизвести его, создавая зрение роботов, они потерпели неудачу. По человеческим меркам, зрение роботов до сих пор остается весьма примитивным. Хьюбел и Визел показали архитектурное решение, которое использовала Природа, — это архитектура скрытых слоев. Каждый нейрон в скрытом слое обладает матрицей сравнения, которая активизируется и посылает сигналы в следующий слой, только когда информация, поступающая из предыдущего слоя, соответствует (с некоторой точностью) этой матрице.

В разговоре о скрытых слоях важно делать различие между простой констатацией эффективности и силы хорошей сети и сложной проблемой того, как такую сеть создать. Играть на фортепиано, кататься на велосипеде или плавать легко, когда вы этому уже научились (функционирование готовой системы), но сам процесс обучения труден (создание системы). Одна из величайших нерешенных проблем современной науки: как в нейронной сети появляются и укладываются новые скрытые слои. Я бы даже рискнул назвать эту проблему величайшей.

Если же вывести идею скрытых слоев за рамки нейронных сетей, она становится очень мощной, богатой и многосторонней метафорой, с помощью которой можно многое объяснить. Например, занимаясь физикой, я заметил, насколько большое значение имеет называние тех или иных явлений. Когда Мюррей Гелл-Манн открыл элементарные частицы «кварки», он дал имя некоему набору парадоксальных фактов, объединив их в единое множество. Как только границы этого множества были очерчены, перед физиками встала задача его точного математического описания, но поворотным моментом в решении этой проблемы стало именно ее определение. Точно так же, когда я открыл «энионы» (тип частиц, объединяющий фермионы и бозоны, в двухмерных системах. — Esquire), я знал, что ухватил некое множество связанных между собой идей, но я даже представить не мог, какую потрясающую эволюцию пройдут эти идеи и какое место займут в реальности. Во всех подобных случаях именование явления рождает новые узловые соединения в скрытых слоях.

Убежден, что идея скрытых слоев очень точно объясняет принципы функционирования сознания — и не важно, человеческого, животного или инопланетного, сознания прошлого, настоящего или будущего. Сознание осваивает идеи, интегрируя их в себя как свойства, которые распознаются в скрытых слоях. И как чудесно бы было, если бы такая полезная идея, как «скрытые слои», была включена в скрытые слои повсеместно».

http://esquire.ru/ideas/frank-wilczek