

Модель формирования восприятия на основе многослойной нейронной сети

Зайнутдинов М.Р., Горбунов И.А.

Санкт-Петербургский государственный университет

Теория сложных систем расширила понятие научной модели. Если раньше только строгие математические модели считались научными, то сейчас таковыми считаются и вычислительные (Dawney, 2012). Хорошим примером такой модели служит проблема оваций (Miller, Page, 2004). Представим, что после окончания великолепной лекции по методологии психологии некоторые слушатели встают и начинают аплодировать. Увидев аплодирующих, другие тоже начинают вставать и аплодировать. В итоге овация либо затухает, либо заражает весь зал.

Чтобы объяснить динамику роста оваций, была предложена следующая модель. Каждый участник имеет порог: если отношение аплодирующих людей ко всему залу больше этого порога, то участник тоже встает и начинает аплодировать. Динамика этого процесса заключается в том, что сначала встают участники с низкими порогами, а затем к овации начинают подключаться участники с высокими порогами. В ходе симуляции такой модели на компьютере были сделаны некоторые контринтуитивные выводы. Например, если среди участников увеличить средний порог и дисперсию, то, вопреки ожиданиям, вероятность овации повышается (Miller, Page, 2004). Возникает вопрос о ценности среднего значения, которое так любят считать психологи. Такая модель с минимальными изменениями может быть перенесена на другие социальные явления: совершение преступления, мародерства, потребление наркотиков, выбор школы, предпочтение партии на выборах и проч. Важное отличие вычислительной модели от математической заключается в том, что симулировать можно множество объектов, и таким образом появляется возможность изучать их взаимодействие, эмерджентное свойство изучаемой системы.

Для математических моделей это чрезвычайно затруднено. Идея об интеллекте как эмерджентном свойстве системы прозвучала в 80-е годы одновременно в нейронауке, в исследованиях искусственного интеллекта, в когнитивной науке и в информатике (McClelland, 2010). Хопфилдом (Hopfield, 1982) была предложена модель ассоциативной памяти на основе множества связанных с друг другом агентов (нейронов). Другой вид нейронных сетей называется многослойным перцептроном. В такой сети нейроны объединены в слои. Входная информация поступает на нейроны первого слоя. Нейроны каждого слоя передают возбуждение нейронам следующего слоя. Таким образом, сигнал добирается до высшего слоя. Предполагается, что система уже знает, какова должна быть активация нейронов в последнем слое. Если активация отличается от эталона, то связи между нейронами (синапсы) перестраиваются таким образом, чтобы на последний слой приходил

правильный ответ. Постепенное изменение связей между нейронами приводит к тому, что нейронная сеть находит закономерность в данных и начинает распознавать в поступающей информации необходимые паттерны (Lee Cun, 1998; Осовский, 2002).

Такая архитектура нейронной сети хорошо справляется с прикладными задачами, но не подходит для описания живой системы, так как не представляется возможным объяснить, откуда система заранее знает о том, какая информация должна приходить на последний слой. Правдоподобная архитектура для описания когнитивной системы была представлена И. А. Горбуновым. Он предложил, что нейронную сеть можно свернуть так, чтобы последний слой являлся также и первым. Первым слоем и последним оказываются рецепторы. Информация, поступившая на рецепторы в момент времени t , передается по НС на следующие слои и, в конечном счете, возвращается назад к рецепторам, но уже в момент времени $t+1$, отражая ожидаемый образ. Такая система постоянно перестраивается так, чтобы находить структуру в потоке поступающей информации для того, чтобы точнее предсказывать будущее. Чем больше рассогласование предсказанного и случившегося, тем сильнее изменения в синапсах. Такая модель хорошо объясняет эффект мигания внимания. При паузе между картинками возникает ошибка прогноза, которая производит модификацию синапсов и не дает нервной системе объединить две картинки в единый временной континуум. Исходя из этой модели, можно ожидать, что отличие в ожидаемом образе от реально поступившего, отраженное в вызванных потенциалах мозга, будет наблюдаться практически с первых миллисекунд после появления сигнала. Такое следствие противоречит исследованиям (Naatanen, 1978; Иванницкий, 1997; Эдельман, Маунткастел, 1981), в котором регистрировалась негативность рассогласования только через 150-250 мс. Однако, возможно, такая негативность может быть зарегистрирована на более ранних этапах процесса восприятия.