

Рекомендованная ссылка

Нейросетевое моделирование когнитивных функций мозга: обзор основных идей [Электронный ресурс] / Терехин А.Т. [и др.] // Психологические исследования: электрон. журн. 2009. N 2(4). URL: <http://psystudy.ru> (дата обращения: чч.мм.20гг).

Нейросетевое моделирование когнитивных функций мозга: обзор основных идей

Терехин А.Т.*, Будилова Е.В.**, Качалова Л.М.***, Карпенко М.П.****
Московский государственный университет им. М.В.Ломоносова, Москва, Россия
Современная гуманитарная академия, Москва, Россия

Дан обзор основных идей нейросетевого моделирования когнитивных функций мозга. Описаны ряд моделей нейрона (пороговый нейрон Мак-Каллока и Питтса, нейрон с сигмоидальной функцией активации, нейрон с немонотонной функцией активации, стохастический нейрон, импульсный нейрон) и ряд нейросетевых архитектур (перцептрон, сеть обратного распространения, сеть Хопфилда, машина Больцмана). Рассмотрены структурные модели, состоящие из нескольких нейронных сетей, и моделирующие функции конкретных систем мозга (гиппокамп, гиппокамп – неокортекс, префронтальная кора – базальные ганглии). Обсуждаются общие проблемы моделирования когнитивных функций мозга.

Ключевые слова: нейросетевое моделирование, нейронные сети, модель нейрона, перцептрон, нейронная сеть, сеть обратного распространения, машина Больцмана, когнитивные функции, мозг, гиппокамп, неокортекс, префронтальная кора, базальные ганглии

Neural network modeling of brain cognitive functions: review of basic ideas

Terekhin A.T., Budilova E.V., Kachalova L.M., Karpenko M.P.
Lomonosov Moscow State University, Moscow, Russia
Modern University for the Humanities, Moscow, Russia

A review of basic ideas of neural network modeling of brain cognitive functions is given. Several models of the neuron are described: threshold neuron of McCulloch and Pitts, neuron with sigmoid activation function, neuron with non-monotone activation function, stochastic neuron, impulse neuron. As well different types of network architectures are described: perceptron, back-propagation network, Hopfield network, Boltzman machine. Structured models composed of more than one neural networks and used for modeling particular brain systems (hippocampus, hippocampus – neocortex system, prefrontal cortex – basal ganglia system) are considered. In conclusion, general principles of brain modeling are discussed.

Keywords: neural network modeling, neural network, neuron model, perceptron, Hopfield network, back-propagation network, Boltzman machine, cognitive functions, brain, hippocampus, neocortex, prefrontal cortex, basal ganglia

Оглавление

1 Модели нейрона

- 1.1 Пороговый нейрон Мак-Каллока и Питтса
- 1.2 Нейрон с сигмоидной функцией активации
- 1.3 Нейрон с немонотонной функцией активации
- 1.4 Стохастическая модель нейрона
- 1.5 Импульсная модель нейрона

2 Персептроны и сети обратного распространения

- 2.1 Персептроны
- 2.2 Сети обратного распространения ошибки

3 Сети Хопфилда

- 3.1 Пороговая функция активации
- 3.2 Сигмоидная функция активации
- 3.3 Машина Больцмана и моделированный отжиг
- 3.4 Запоминание последовательностей

4 Импульсные сети

- 4.1 Обучение
- 4.2 Стохастичность
- 4.3 Запоминание последовательностей

5 Структурные сетевые модели

- 5.1 Гиппокамп
- 5.2 Гиппокамп – неокортекс
- 5.3 Префронтальная кора – базальные ганглии
- 5.4 Основные принципы моделирования мозга

Обсуждение и перспективы исследований

Литература

Математическое моделирование работы мозга позволяет получить научные результаты, к которым трудно прийти иным путем. Хотя несомненно, что на интуитивно-вербальном уровне формируются первоначальные гипотезы и в вербальной форме формулируются окончательные результаты исследования, только после построения строгой математической модели можно быть уверенным, что вербальная модель не имеет скрытых противоречий, логических пропусков и неявных нестыковок. В процессе построения математической модели выявляются пробелы в имеющихся знаниях о некоторых частях моделируемого объекта и формируются рабочие гипотезы о возможных механизмах функционирования этих частей, стимулирующие проведение эффективных экспериментальных исследований. Во многих случаях, когда реальный эксперимент затруднен или невозможен, он может быть заменен модельным экспериментом. Часто результаты, полученные в результате моделирования, радикально отличаются от тех, которые основаны на чисто вербальных теориях [O'Reilly, Munakata, 2003]. Поэтому нейросетевое моделирование играет важную роль в когнитивных исследованиях, как при планировании экспериментальных работ, так и при теоретическом осмыслении их результатов.

Предметом нашего рассмотрения будут так называемые нейронные сети. Удобно различать реальные биологические, математические и технические нейронные сети.

Биологические нейронные сети, состоящие из анатомически и информационно связанных между собой клеток-нейронов, являются основой нервной системы животных и человека и служат предметом изучения нейроанатомии и нейрофизиологии.

Математические нейронные сети исторически возникли как модели биологических нейронных сетей. Первым наиболее заметным достижением в этом направлении явилась работа Мак-Каллока и Питтса [McCulloch, Pitts, 1943], в которой в качестве модели реальной нейронной сети была предложена абстрактная сеть из формальных нейронов, функционирующих по строго определенным математическим правилам. Дальнейшие продвижения в этой области были связаны с созданием так называемых персептронов – сетей, преобразующих сигналы, поступающие на входной слой сенсоров, в нужные сигналы на выходном слое, состоящем из формальных нейронов Мак-Каллока и Питтса [Rosenblatt, 1962]. Существенным развитием персептронов стали сети обратного распространения ошибки (error back-propagation networks), фактически представляющие собой многослойные персептроны с алгоритмом обучения, основанным на обратном (от выходного слоя к входному) распространении ошибки (отклонения выходных сигналов от желаемых) [Rumelhart, Hinton, Williams, 1986]. Сильное влияние на развитие нейросетевых исследований оказала работа Хопфилда [Hopfield, 1982], в которой исследовалось поведение полносвязных сетей из формальных нейронов и было показано, что если проводить обучение такой сети по правилу Хебба [Hebb, 1949], то есть при предъявлении на входе сети некоторого образа усиливать связи между одновременно возбужденными нейронами, то в последующем достаточно предъявлять сети только части этого образа для того, чтобы он восстанавливался полностью, то есть сеть Хопфилда обладает свойством ассоциативного вспоминания. Высокую степень воздействия этой работы на ускорение исследований в области нейронных сетей демонстрирует экспоненциальный рост числа ссылок на нее во второй половине 80-х и начале 90-х годов [Дунин-Барковский, Терехин, 1990; Budilova, Teriokhin, 1992]. Это стимулирующее влияние сохраняется и в настоящее время – библиографическая поисковая система Scholar Google выдает 700 ссылок на документы 2007 года, содержащие набор ключевых слов «Hopfield neural networks».

Следует отметить, что хотя традиционно математические нейронные сети воспринимаются как модели биологических сетей, большая часть нейросетевых исследований в настоящее время посвящена разработке и анализу нейронных сетей, предназначенных для решения вычислительных задач и задач управления техническими системами. Основным критерием в этом случае становится эффективность решения сетью поставленной задачи, а ее биологическая правдоподобность вообще не принимается во внимание. Реализованные в виде компьютерных программ или микропроцессорных схем такие математические сети фактически становятся столь же реальными, как и биологические сети, – можно назвать их техническими нейронными сетями. Размах прикладного технического направления в нейросетевых исследованиях можно, в частности, оценить по большому числу ссылок на документы, содержащие набор ключевых слов «error back-propagation networks» в поисковой системе Scholar Google, – около 1400 ссылок на работы 2007 года (сети обратного распространения редко рассматриваются как адекватные модели реальных биологических сетей и используются, в основном, как формальный метод для решения вычислительных задач).

В этом обзоре мы сосредоточимся на рассмотрении математических нейронных сетей исключительно в качестве моделей биологических нейронных сетей как инструмента, позволяющего теоретически объяснить имеющиеся экспериментальные результаты нейроанатомических, нейрофизиологических и нейропсихологических исследований и, возможно, наметить перспективные направления дальнейших экспериментальных работ.

1 Модели нейрона

1.1 Пороговый нейрон Мак-Каллока и Питтса

Пороговый нейрон Мак-Каллока и Питтса [McCulloch, Pitts, 1943] представляет собой, по-видимому, самую простую модель биологического нейрона, которая, тем не менее, отражает наиболее существенные функции последнего и поэтому вплоть до настоящего времени служит базовым элементом при построении многих нейросетевых моделей.

Этот модельный нейрон представляет собой элемент с несколькими входами и одним выходом, являясь устройством, преобразующим набор входных сигналов в выходной сигнал по строго определенным правилам (рис. 1).

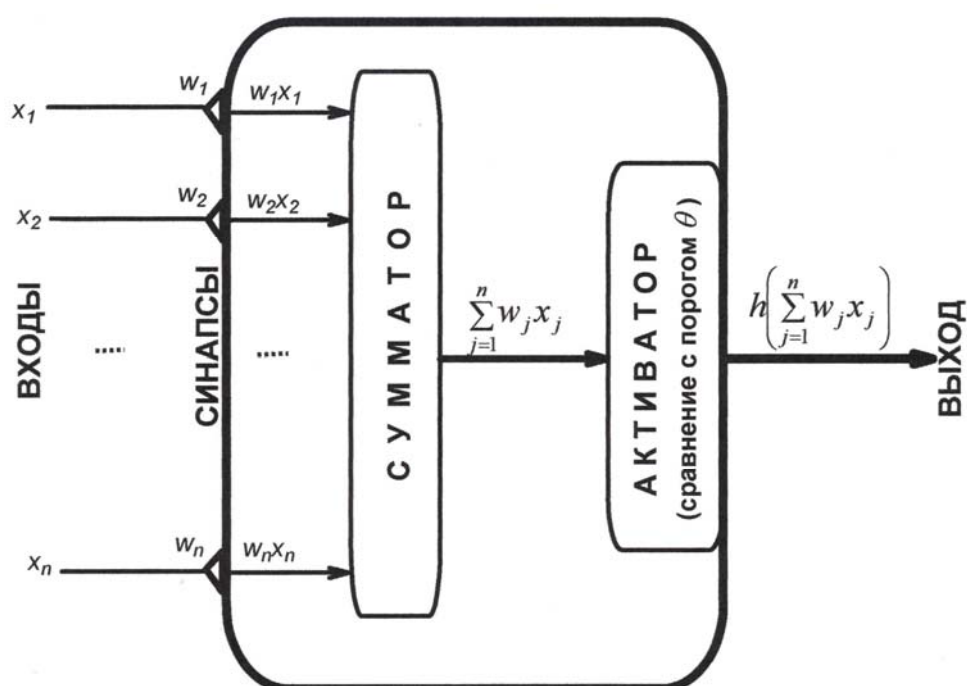


Рис. 1. Формальный нейрон Мак-Каллока и Питтса.

Предполагается, что если на входы (синапсы) нейрона i подаются сигналы x_1, x_2, \dots, x_n от выходов других нейронов или внешние сигналы, то они умножаются на весовые множители $w_{i1}, w_{i2}, \dots, w_{in}$, суммируются и сравниваются с пороговым значением θ_i . В зависимости от результата сравнения выход i -го нейрона u_i устанавливается равным либо значению 0, если взвешенная сумма входных сигналов меньше порога (невозбужденное состояние нейрона, соответствующее минимальной фоновой интенсивности генерации спайков), либо значению 1, если взвешенная сумма входных сигналов больше порога или равна ему (возбужденное состояние, соответствующее максимальной интенсивности генерации спайков), то есть

$$u_i = \begin{cases} 0, & \text{если } \sum_{j=1}^n w_{ij} x_j \leq \theta_i \\ 1, & \text{если } \sum_{j=1}^n w_{ij} x_j > \theta_i. \end{cases}$$

Другими словами, нейрон Мак-Каллока и Питтса после умножения каждого входа на соответствующий ему синаптический вес и суммирования полученных произведений преобразует полученную сумму в выходной сигнал с помощью пороговой функции

$$h(x; \theta) = \begin{cases} 0, & \text{если } x \leq \theta \\ 1, & \text{если } x > \theta \end{cases}$$

(θ – параметр этой функции, задающий ее порог). На рис. 2 приведен пример графика пороговой функции для $\theta = 0,5$.

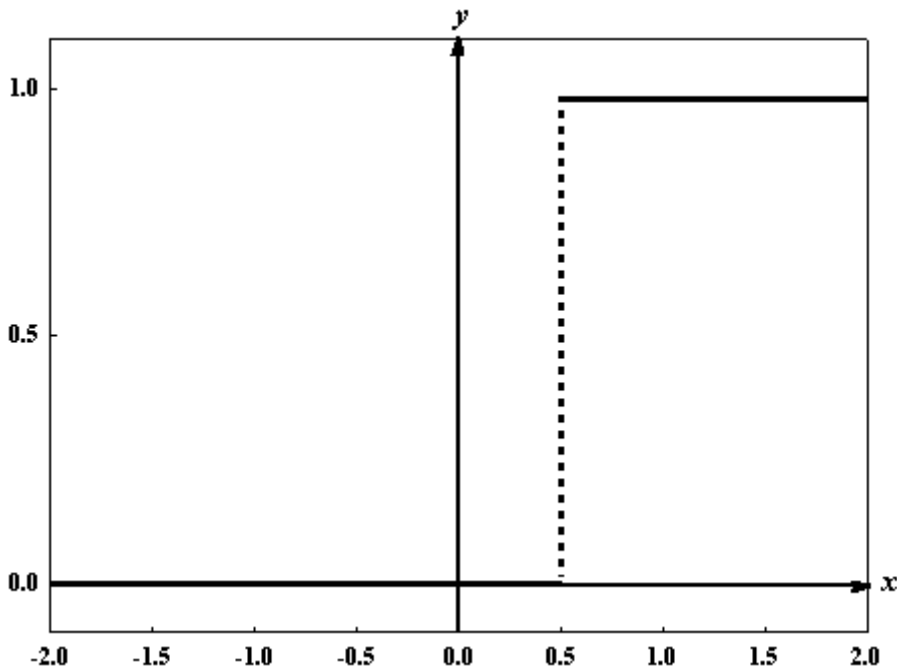


Рис. 2. Пример пороговой функции $y = h(x; \theta)$ для $\theta = 0,5$.

Функция, преобразующая взвешенную сумму входов в выходной сигнал нейрона, называется его функцией активации.

1.2 Нейрон с сигмоидной функцией активации

Наряду с пороговой функцией в нейросетевом моделировании используются и другие функции активации, в частности так называемая сигмоидная, или логистическая, функция

$$s(x; \sigma, \theta) = \frac{1}{1 + e^{-\sigma(x-\theta)}}, \text{ пример которой для } \theta = 0 \text{ и } \sigma = 5 \text{ представлен на рис. 3.}$$

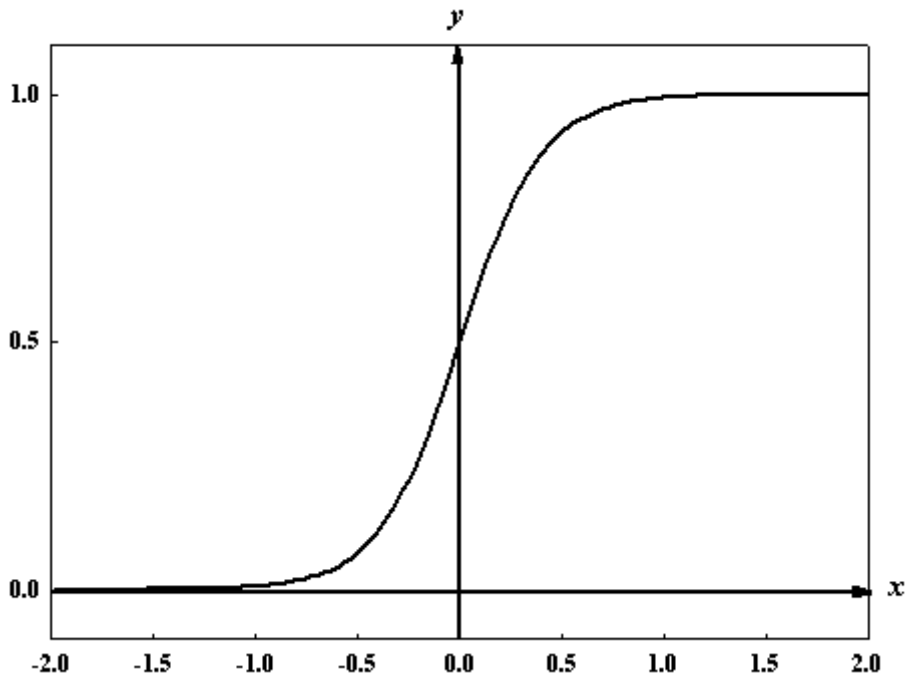


Рис. 3. Пример сигмоидной функции для $\theta = 0$, $\sigma = 5$.

Параметр сдвига θ этой функции аналогичен параметру порога θ пороговой функции, а параметр σ определяет тангенс угла наклона кривой в точке перегиба: чем больше σ , тем круче кривая, приближаясь к пороговой функции при неограниченном увеличении σ . В отличие от пороговой функции, принимающей только два значения 0 и 1, сигмоидная функция может принимать любое значение в промежутке между 0 и 1, что более адекватно описывает реальный нейрон, который может генерировать спайки и с промежуточными между минимальной и максимальной интенсивностями.

1.3 Нейрон с немонотонной функцией активации

Пороговая и сигмоидная функции активации являются монотонно неубывающими функциями. Однако определенными преимуществами обладают и немонотонные функции активации [Morita, 1996] типа функции, заданной формулой

$$n(x; c, d, k, h) = \frac{1 - e^{-cx}}{1 + e^{-cx}} \times \frac{1 - ke^{-d(|x|-h)}}{1 + e^{-d(|x|-h)}}$$

и представленной на рис. 4 для следующих значений ее параметров: $c = 50$, $d = 10$, $k = 0,5$ и $h = 1$ (параметр k определяет уровни асимптот, а h – ширину диапазонов, где y близок к 0 или 1. Характерным свойством такого нейрона является снижение уровня выходного сигнала в предпороговой области и его повышение (с последующим снижением) – в послепороговой).

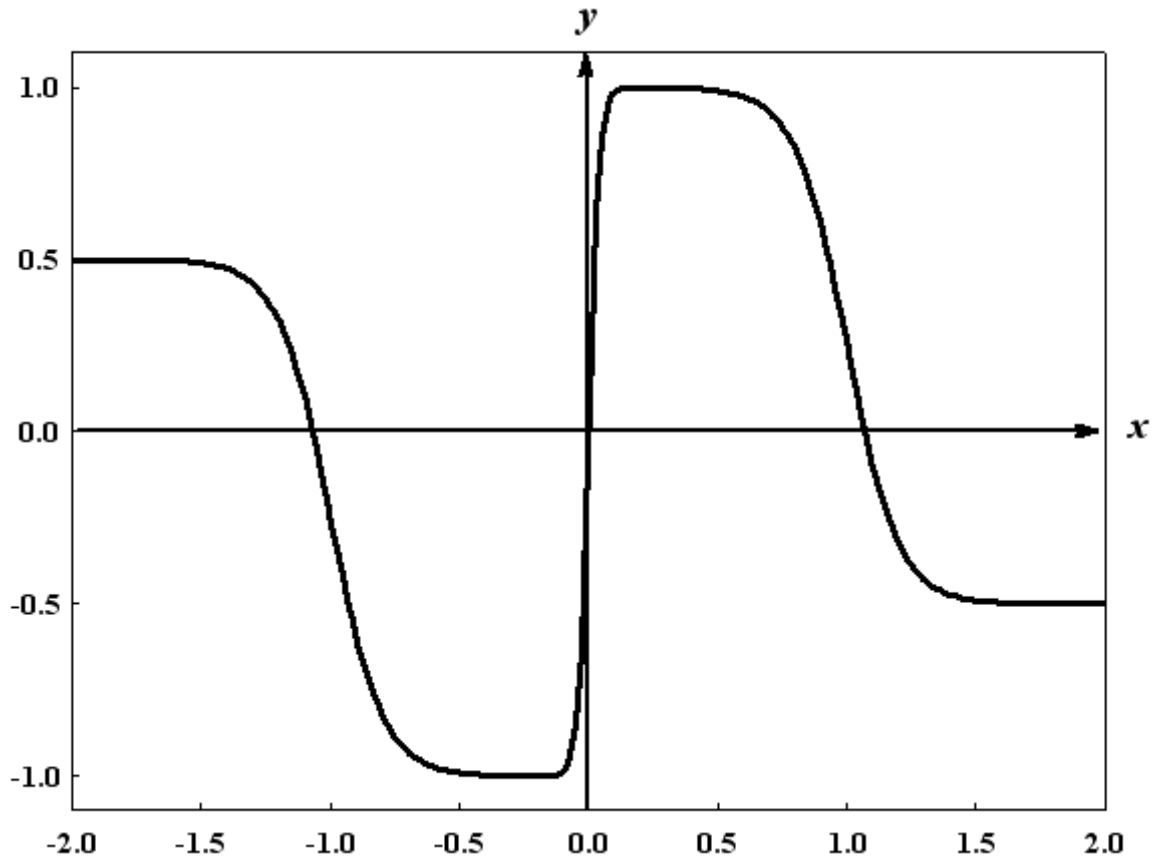


Рис. 4. Пример немонотонной функции активации для $c = 50$, $d = 10$, $k = 0,5$ и $h = 1$.

1.4 Стохастическая модель нейрона

Стохастический нейрон, так же как и нейрон Мак-Каллока и Питтса, может принимать только два выходных значения – 0 или 1, однако эти значения не определяются однозначно взвешенной суммой входов, а зависят от нее стохастически, а именно вероятность появления значения 1 на выходе нейрона сигмоидально (с нулевым сдвигом) зависит от этой суммы [Hinton, Sejnowski, 1986]:

$$P(u_i = 1) = \frac{1}{1 + e^{-\frac{1}{T} \sum_{j=1}^n w_{ij} x_j}}$$

Параметр T , входящий в эту формулу и называемый температурой нейрона, определяет степень его стохастичности. При $T \rightarrow 0$ степень стохастичности уменьшается, так как крутизна сигмоидной кривой неограниченно растет и формула начинает работать детерминистически просто как пороговая функция активации:

значение 1 принимается с вероятностью 1 при $\sum_{j=1}^n w_{ij} x_j \geq 0$, а значение 0 – с

вероятностью 1 при $\sum_{j=1}^n w_{ij} x_j < 0$. И, наоборот, при $T \rightarrow \infty$ крутизна сигмоидной

кривой уменьшается и она стремится к горизонтальной прямой, проходящей на уровне $P(u_i = 1) = 0,5$, то есть при очень больших температурах независимо от значения взвешенной суммы входных сигналов выходной сигнал с равными

вероятностями может принять значения 0 или 1, и, таким образом, стохастичность (неопределенность) достигает своего максимума.

1.5 Импульсная модель нейрона

В рассмотренных выше моделях нейронов значение выходного сигнала фактически ассоциируется просто со средней частотой следования спайков в пакете. Напротив, так называемые импульсные (*integrate-and-fire*) модели нейрона [Gerstner, 1995; Gerstner, Kistler, 2002] основаны на учете информации о моментах появления отдельных спайков и в этом смысле более детально описывают функционирование нейрона [Hodgkin, Huxley, 1952].

Суть работы импульсного нейрона полностью отражена в его английском названии *integrate and fire* – интегрировать и возбуждаться: каждый поступающий на нейрон входной спайк увеличивает мембранный потенциал нейрона, и как только этот потенциал превзойдет заданный порог активности θ , нейрон генерирует выходной спайк. После генерации спайка мембранный потенциал падает до фонового уровня, а нейрон становится на некоторое время (период рефрактерности) неспособным генерировать следующий спайк.

Также предполагается, что параллельно с увеличением мембранного потенциала нейрона под воздействием спайков, поступающих на его синаптические входы, происходит самопроизвольное (за счет утечек) уменьшение этого потенциала во времени, то есть динамика изменения мембранного потенциала $u_i(t)$ нейрона i может быть описана следующим дифференциальным уравнением:

$$\frac{du_i(t)}{dt} = \frac{1}{C} I_i(t) - \frac{1}{RC} u_i(t),$$

где $I_i(t)$ – синаптический приток, C – электрическая емкость нейрона, а R – его сопротивление утечки. Чтобы отобразить эффект самопроизвольного падения мембранного потенциала в импульсном нейроне, в его англоязычное название иногда добавляют еще один эпитет – *leaky* – имеющий протечки): *leaky integrate and fire neuron*.

Форму зависимости выходного импульса (спайка) от времени t можно задать в простейшем случае просто дельта-функцией $\delta(t)$, то есть функцией, значение которой равно бесконечности в точке нуль и нулю вне этой точки и интеграл от которой в нуле равен 1. В этом случае спайк, сгенерированный некоторым нейроном j в момент τ_j , будет описываться дельта-функцией, сдвинутой на время τ_j вправо от нуля, то есть как $\delta(t - \tau_j)$, а на синаптический вход нейрона i этот спайк придет с некоторым запаздыванием Δ_{ij} и будет равен $\delta(t - \tau_j - \Delta_{ij})$.

Соответственно, общий синаптический приток нейрона i будет равен взвешенной сумме притоков от отдельных синапсов

$$I_i(t) = \sum_j w_{ij} \delta(t - \tau_j - \Delta_{ij}),$$

где w_{ij} – эффективность (вес) синаптической связи от нейрона j к нейрону i .

Более реалистичная, но достаточно простая для моделирования модель импульсного нейрона была предложена в [Izhikevich, 2003]. Она задается системой двух дифференциальных уравнений:

$$\frac{du_i(t)}{dt} = I_i(t) + mu_i^2(t) + nu_i(t) + p - v(t),$$

$$\frac{dv_i(t)}{dt} = a(bu_i(t) - v_i(t)),$$

где переменная $u_i(t)$ по-прежнему обозначает мембранный потенциал нейрона i , тогда как потенциал восстановления $v_i(t)$ связан с активацией калиевого и дезактивацией натриевого ионных токов и обеспечивает обратную связь для $u_i(t)$. Константы m , n , p , a и b являются параметрами этих уравнений. Еще три параметра c , d и θ характеризуют процесс образования спайка: он генерируется, как только мембранный потенциал превысит порог θ , после чего немедленно мембранный потенциал опускается до уровня c , а потенциал $v_i(t)$ устанавливается на уровне $v_i(t) + d$.

На рис. 5 показаны разные типы последовательностей спайков, которые могут быть сгенерированы этой моделью [Izhikevich, 2003].

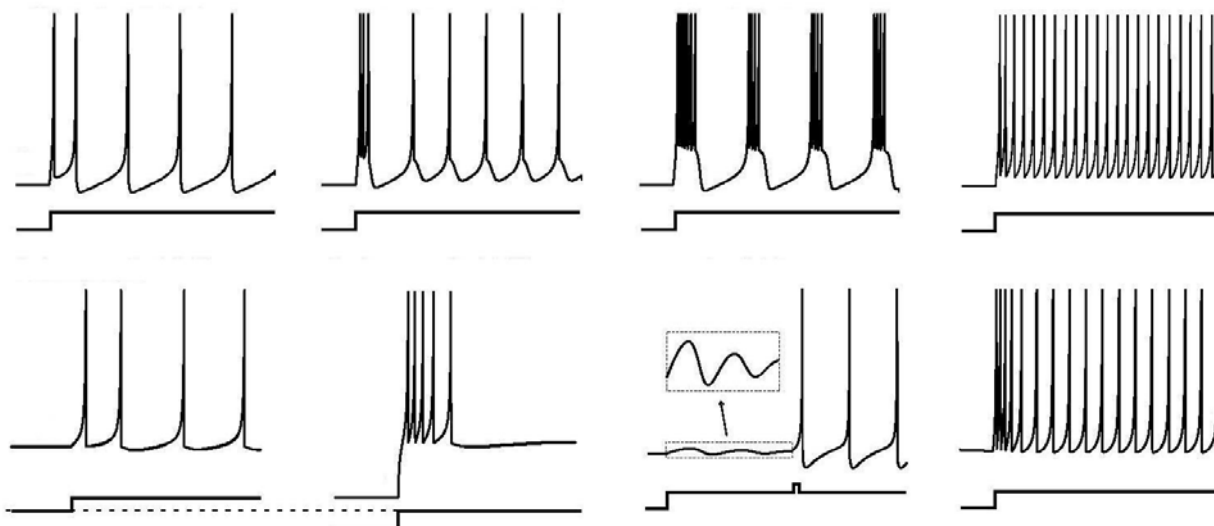


Рис. 5. Различные типы последовательностей спайков, генерируемых импульсной моделью нейрона Ижикевича [Izhikevich, 2003].

2 Перцептроны и сети обратного распространения

2.1 Перцептроны

Появление перцептронов в начале 60-х годов прошлого века [Rosenblatt, 1962; Розенблатт, 1965] вызвало первый бум нейросетевого моделирования, поскольку было показано, что даже совсем простая сеть из нейронов Мак-Каллока и Питтса может демонстрировать элементы интеллектуального поведения.

Перцептрон состоит из входного слоя сенсоров, соединенных синаптическими связями с нейронами выходного слоя по принципу «все со всеми». На рис. 6 приведен пример перцептрона с четырьмя входами и двумя выходами.

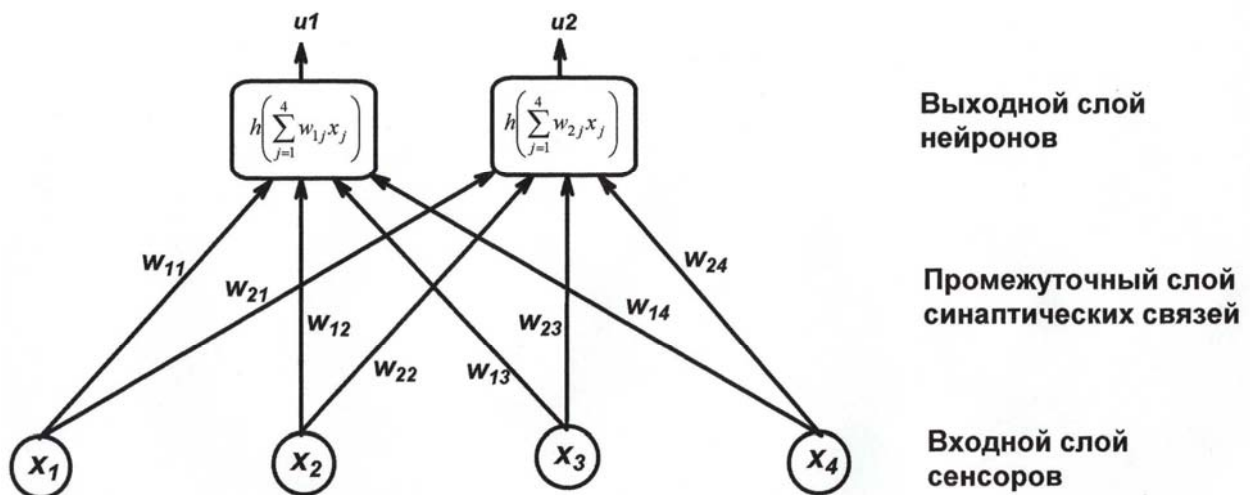


Рис. 6. Пример персептрона с четырьмя входами и двумя выходами.

Если выходные нейроны персептрона – это командные нейроны, запускающие различные программы действий организма (нападение на жертву, бегство от хищника и т.п.), то можно сказать, что персептрон на основании получаемой от сенсоров информации вырабатывает решение о принятии тех или иных действий. Какие именно действия будут предприняты в ответ на те или иные ситуации, зависит от значений синаптических весов нейронов и их порогов. Впрочем, пороги можно считать фиксированными, например нулевыми, если ввести дополнительный вход с постоянным сигналом, равным -1. Влияние этого входа будет равноценно наличию у нейрона порога, равного его синаптическому весу. Соответственно, в дальнейшем мы не будем упоминать о влиянии порога на выходное значение нейрона.

Нужные значения синаптических весов могут сформироваться двумя путями.

Если речь идет о безусловном рефлекс на ситуацию, то можно предположить, что веса сформировались генетически: мутации порождали индивидуальные колебания в значениях весов, а естественный отбор выбраковывал индивидов, неадекватно реагирующих на ситуацию из-за неправильных значений весов (бегство от жертвы, нападение на хищника и т.п.).

Что касается условного рефлекса, то он может сформироваться вследствие изменения весов в соответствии с правилом Хебба, согласно которому веса синаптических связей между одновременно возбужденными нейронами увеличиваются. Рис. 7 иллюстрирует образование условного рефлекса выделения слюны в ответ на звонок на примере простейшей сети из трех входных нейронов и одного выходного.

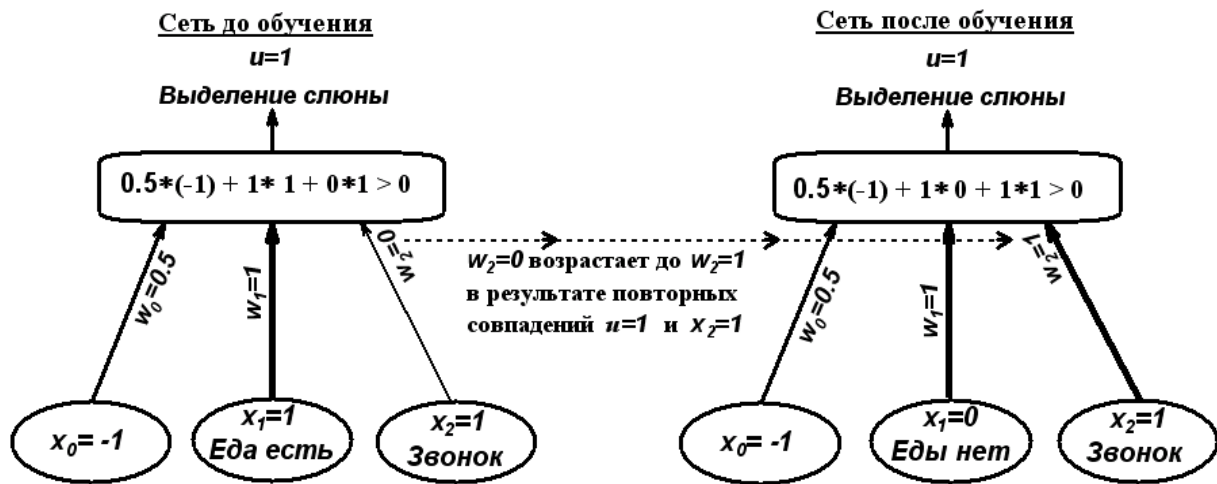


Рис. 7. Иллюстрация хеббовского обучения персептрона с тремя входами и одним выходом условному рефлексу выделения слюны в ответ на звонок.

Еще один способ обучения (подстройки синаптических весов) персептрона состоит в том, что после каждого предъявления персептрону некоторой ситуации (образа) полученный выход сравнивается с правильным решением (которое каждый раз предполагается известным), после чего веса корректируются в направлении уменьшения расхождения (дельты) между правильным решением и полученным выходом: если отклонение произошло в большую сторону, то надо уменьшить каждый синаптический вес на величину, пропорциональную его входу, а если в меньшую, то увеличить. Этот метод (называемый дельта-правилом) применим к персептронам из нейронов как с пороговой, так с сигмоидной функцией активации. Непосредственной биологической интерпретации способу обучения с помощью дельта-правила не просматривается (хотя попытки ее найти предпринимаются до сих пор), однако наличие регулярного алгоритма обучения персептрона немало способствовало его популярности, по крайней мере среди специалистов в области технической кибернетики.

Энтузиазм, однако, сменился глубоким разочарованием, когда было показано, что круг задач, решаемых однослойным персептроном Розенблатта, довольно ограничен [Minsky, Papert, 1969; Минский, Пейперт, 1971]. Точнее, его можно научить различать лишь ситуации, линейно разделимые в пространстве признаков. Например, невозможно обучить персептрон относить к одному классу ситуации, когда сигналы на двух входах имеют либо оба низкие, либо оба высокие значения, а к другому классу – ситуации, когда сигналы имеют разные значения. Хотя было известно, что многослойные персептроны, то есть персептроны, имеющие промежуточные слои нейронов (см. рис. 8), в принципе могут решать задачи любой сложности, в то время не имелось регулярного алгоритма обучения многослойных персептронов. Результатом было резкое снижение исследовательской активности в области нейронных сетей в семидесятые и начале восьмидесятых годов.

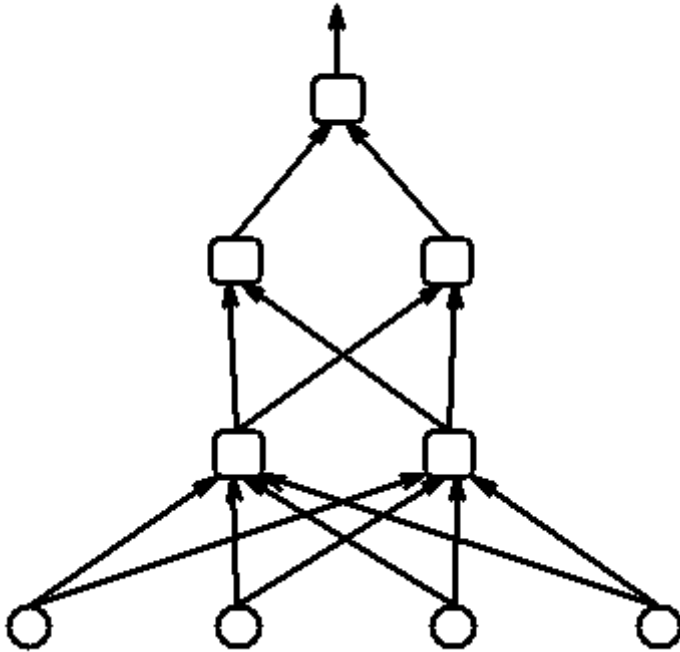


Рис. 8. Пример трехслойного персептрона (четыре входа, два нейрона в первом слое, два – во втором и один – в третьем).

2.1 Сети обратного распространения ошибки

Однако во второй половине восьмидесятых годов интерес к персептронам, теперь уже главным образом многослойным, вновь возродился. Это было связано с разработкой эффективного алгоритма обучения многослойных персептронов – метода обратного распространения ошибки [Rumelhart, Hinton, Williams, 1986]. Многослойные персептроны, обучаемые с помощью этого метода, называют сетями обратного распространения.

Метод обратного распространения ошибки является обобщением дельта-правила и применяется, в основном, к сетям, образованным из нейронов с сигмоидной функцией активации. Суть его состоит в следующем.

Сначала вычисляются выходные значения сети. Это делается последовательно: на первом шаге по входам сети вычисляются выходные значения нейронов первого слоя, после чего полученные значения принимаются за входные и по ним вычисляются значения второго слоя и т.д. Это прямой проход сети, в процессе которого меняются выходные значения нейронов. После него следует обратный проход от выходного слоя к входному, во время которого, однако, меняются не выходные значения нейронов, а веса связей. Для корректировки весов последнего слоя можно фактически применить дельта-правило, поскольку есть входные значения предпоследнего слоя, вычисленные выходные значения последнего слоя и известные правильные выходные значения. Сложнее дело обстоит с предпоследним слоем, поскольку для него неизвестны правильные значения и поэтому нельзя вычислить ошибку. Авторы алгоритма предложили ее вычислять по выходной ошибке, используя те же веса, которые применялись для пересчета значений нейронов выходного слоя, но уже в обратном направлении. А зная ошибку предпоследнего слоя и входные значения с предпредпоследнего слоя, можно,

пользуясь дельта-правилом, модифицировать входные веса предпоследнего слоя. Действуя далее аналогично, можно скорректировать веса всех слоев вплоть до первого.

Основная область применения сетей обратного распространения – технические приложения. Кроме того, они стали широко использоваться в задачах анализа данных вместо классических методов регрессионного и дискриминантного анализа. Это допустимо, потому что многослойный персептрон – это просто сложное уравнение нелинейной регрессии, но вряд ли оправдано в случаях, когда из-за зашумленности трудно выявить в данных зависимости более сложные, чем линейные.

Как модель реальной обработки информации в мозге метод обратного распространения ошибки в настоящее время имеет ограниченное распространение – более широко для этого применяются рассматриваемые в следующем разделе сети Хопфилда. Тем не менее идеи этого метода используются для моделирования некоторых механизмов работы головного мозга [Hinton, McClelland, 1988; O'Reilly, 1996, 1998; O'Reilly, Frank, 2006].

3 Сети Хопфилда

Если открытие метода обратного распространения ошибки стимулировало, главным образом, развитие технических применений формальных нейронных сетей, то появление статьи Хопфилда [Hopfield, 1982], анализирующей свойства полносвязной сети нейронов Мак-Каллока и Питтса, сыграло ключевую роль в развитии нейросетевого моделирования биологических когнитивных процессов (в одной из аннотированных библиографий после этой статьи была лишь одна короткая фраза – «это то, с чего все началось»).

3.1 Пороговая функция активации

На рис. 9 приведен пример сети Хопфилда из четырех нейронов. Она характеризуется тем, что все нейроны связаны со всеми и есть только два ограничения на веса синаптических связей: веса прямых и обратных связей равны между собой, а вес связи нейрона с самим собой равен нулю (условие симметричности связей не является биологически мотивированным, а, скорее, является допущением, упрощающим математический анализ свойств сети). В принципе, каждый нейрон i , $i = 1, 2, \dots, n$ может иметь как вход, так и выход, а его состояние задается переменной состояния u_i , принимающей значение -1 или 1 (использование значений -1 и 1, а не 0 и 1 также мотивировано стремлением к упрощению анализа свойств сети).

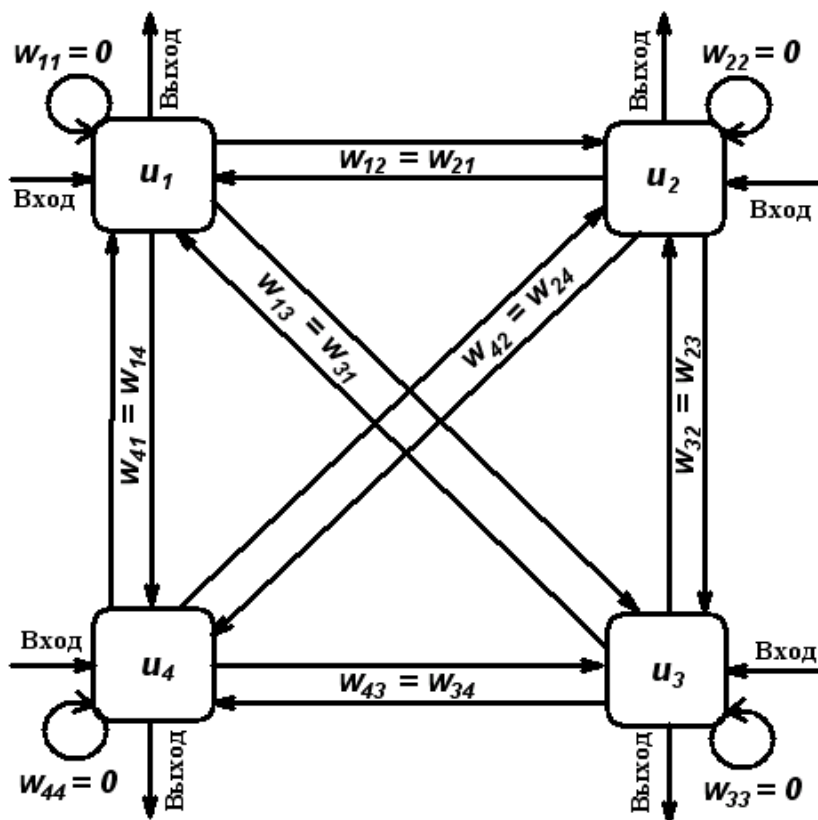


Рис. 9. Пример симметричной сети Хопфилда из четырех нейронов.

Динамика сети определяется правилами изменения состояний ее нейронов и весов соединяющих их синаптических связей.

Предположим, что веса сети $w_{ij}(t)$ и состояния $u_i(t)$ всех ее нейронов в момент времени t каким-то образом заданы, например, просто случайным образом. Тогда в следующий момент $t + 1$ состояние нейрона i изменится в соответствии с правилом Мак-Каллока и Питтса:

$$u_i(t + 1) = \text{sgn} \left(\sum_{j=1}^n w_{ij}(t) u_j(t) + I_i(t) \right),$$

где $I_i(t)$ – внешний входной сигнал, а $\text{sgn}(x)$ – функция сигнум (знак), равная -1 для отрицательных значений аргумента и 1 – для положительных (то есть просто пороговая функция).

Что касается синаптических весов, то для всех $j \neq i$ они изменяются в соответствии с правилом Хебба:

$$w_{ij}(t + 1) = w_{ij}(t) + u_i(t) u_j(t),$$

то есть вес увеличится на единицу, если нейроны i и j находятся в одинаковых состояниях (оба равны -1 или оба равны 1) и уменьшится на единицу – если в разных (один в состоянии -1, а другой – в состоянии 1).

Определенная таким образом сеть обладает рядом интересных свойств с точки зрения ее использования как модели биологических когнитивных процессов. Рассмотрим эти свойства. Для наглядности представим, что нейроны образуют

прямоугольную решетку и мы можем наблюдать их состояния в разные моменты времени, как на рис. 10, где неактивные нейроны показаны мелкими точками, а активные – крупными.

Пусть сначала все синаптические веса сети равны нулю, а на входы ее нейронов поступает некоторая конфигурация сигналов – образ. В результате нейроны установятся в состояния, соответствующие этому образу, например букве «Е», после чего, в соответствии с правилом Хебба, сформируются новые веса сети. Можно повторно предъявить образ несколько раз – тогда значения весов кратно увеличатся. Если после этого предъявить сети любое другое сочетание сигналов и разрешить ей свободно эволюционировать во времени в соответствии с правилом Мак-Каллока и Питтса, то через небольшое число шагов состояние сети стабилизируется на ранее многократно предъявленном образе «Е». Можно сказать, что сеть «запомнила» этот образ!

Но что произойдет, если сети предъявить другие образы? Предъявление каждого из них внесет свой вклад в изменение синаптических весов, и оказывается, что они тоже будут «запомнены» сетью.

На рис. 10 демонстрируется динамика сети Хопфилда из 56 нейронов, запомнившей четыре образа – «Е», «Н», «Х» и «У», в моменты времени $t = 1, 2, 3, 4$, после того как в момент $t = 0$ ей был предъявлен некий новый образ [Терехин, Будилова, 1995]. Показано несколько типичных примеров такой динамики. Так, первые четыре тестовых образа представляют собой искаженные варианты ранее предъявляемых букв «Е», «Н», «Х» и «У». Мы видим, что всякий раз от искаженного образа сеть переходит к запомненному неискаженному образу и потом постоянно остается в этом состоянии. Таким образом, сеть может вспомнить запомненный образ, если ей предъявить похожий образ, то есть некоторую «ассоциацию», связанную с этим образом. Другими словами, сеть Хопфилда обладает свойством ассоциативного воспоминания образов!

Последние три примера динамики на рис. 10 получены при задании случайных начальных образов, не похожих или не очень похожих на образы, предъявленные во время обучения. Тем не менее в одном случае система переходит в состояние, соответствующее одному из этих образов, – «Х». Однако в двух других случаях мы получаем новые образы, которые не предъявлялись сети при ее обучении. Это говорит о том, что в числе устойчивых образов, к которым сеть стремится в режиме распознавания и которые называются ее аттракторами, кроме предъявленных при обучении образов имеются и другие. Часто их называют «ложными» и рассматривают как помехи осуществлению сетью функций ассоциативной памяти. Однако появление «ложных» образов имеет и положительный аспект, поскольку резко расширяет когнитивные свойства сетей. Действительно, можно заметить, что последние два аттрактора на рисунке не случайны, а как бы составлены из частей образов, предъявленных при обучении, то есть представляют собой своего рода химеры. Возможность появления таких химер наделяет сеть в некотором смысле способностью к «творческой фантазии» и «обобщению». Например, предпоследний химерический образ на рисунке можно рассматривать как обобщающий символ для «Е» и «Н», а последний – для «Х» и «У».

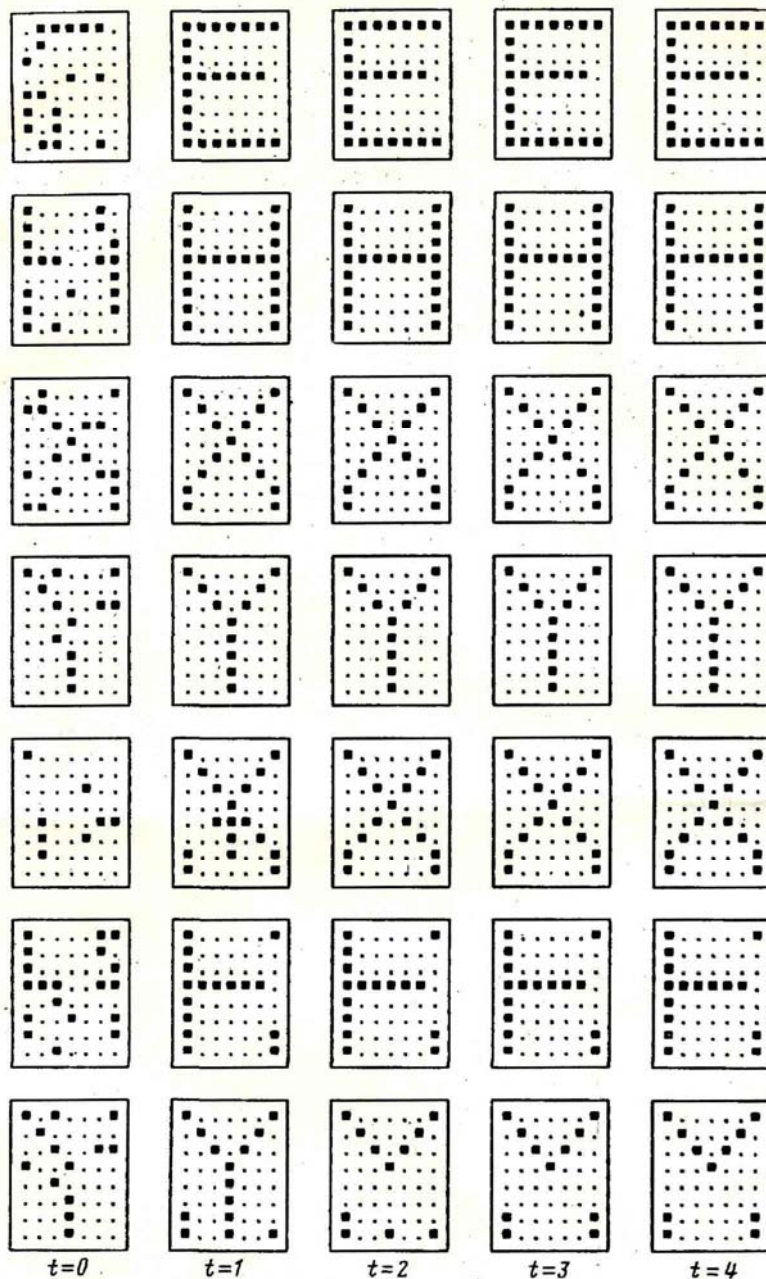


Рис. 10. Динамика состояния сети, обученной распознаванию образов «Е», «Н», «Х» и «У» при предъявлении различных начальных образов в момент $t = 0$.

Поведение сети Хопфилда в режиме распознавания очень удобно интерпретировать в терминах так называемой функции энергии (функции Ляпунова), вид которой в момент t определяется значениями весов связей и которая задается формулой

$$E(t) = -0,5 \sum_{i=1}^n \sum_{j=1}^n w_{ij}(t) u_i(t) u_j(t).$$

Аргументом этой функции является совокупность состояний нейронов, называемая состоянием сети, которое можно наглядно представлять в виде точки в n -мерном пространстве состояний нейронов.

Можно показать, что определяемое правилом Мак-Каллока и Питтса изменение состояния сети может происходить только в направлении уменьшения функции

энергии. Чтобы это показать, допустим, например, что некоторый нейрон k находится в состоянии -1 и $\sum_{j=1}^n w_{kj}(t)u_j(t) > 0$. В этом случае он должен на

следующем шаге перейти в состояние 1 . Соответствующее приращение энергии при этом переходе будет равно

$$dE = E(u_k(t+1) = 1) - E(u_k(t) = -1) = -0,5 \sum_{j=1}^n w_{kj}(t)u_j(t) - 0,5 \sum_{j=1}^n w_{kj}(t)u_j(t) =$$

$$= -\sum_{j=1}^n w_{kj}(t)u_j(t),$$

поскольку ниже все члены в сумме по i , кроме k -го, в выражении для энергии в момент $t + 1$ остаются такими же, как и в момент t и поэтому взаимно уничтожаются при вычислении приращения. Так как, по предположению,

$$\sum_{j=1}^n w_{kj}(t)u_j(t) > 0, \text{ а } dE = -\sum_{j=1}^n w_{kj}(t)u_j(t), \text{ то, очевидно, } dE < 0, \text{ то есть,}$$

действительно, при изменениях состояний сети ее функция энергии может только уменьшаться.

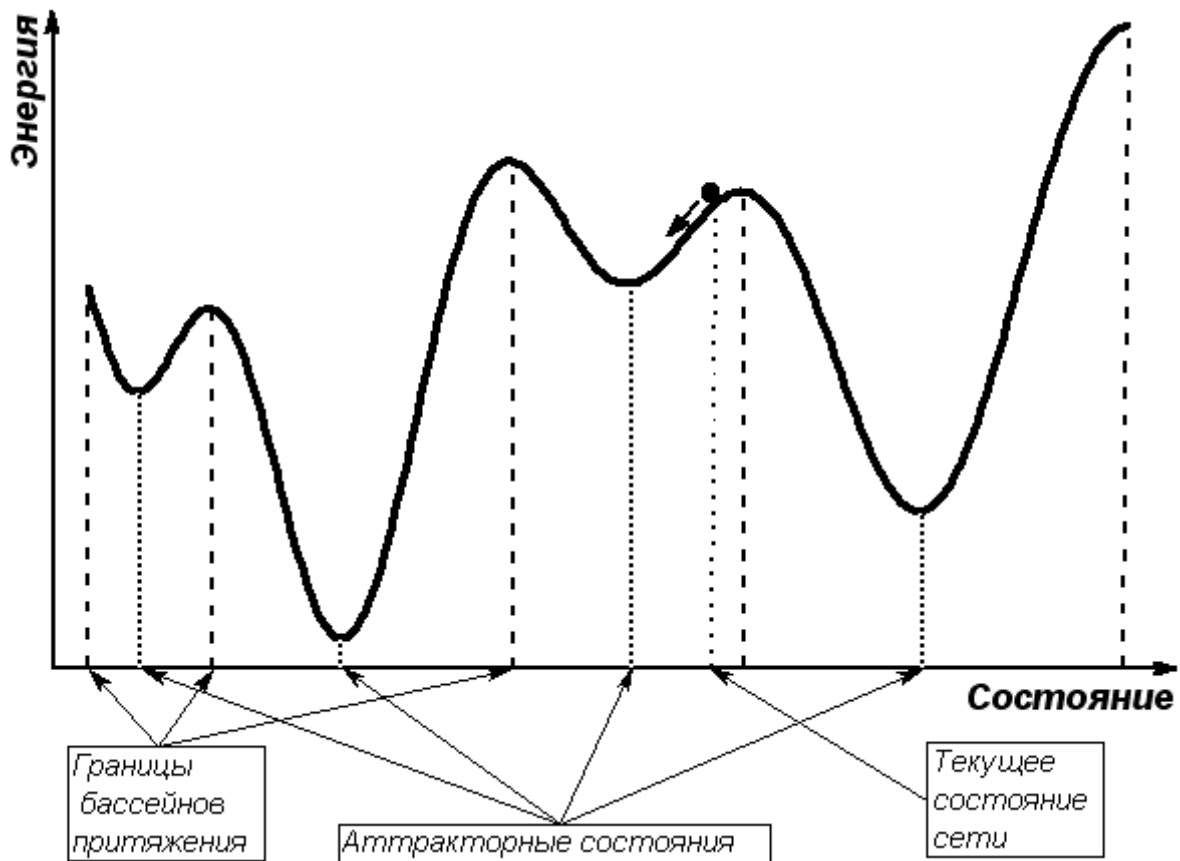


Рис. 11. Иллюстрация аттракторного характера динамики сети Хопфилда в одномерном пространстве состояний.

Этот факт позволяет наглядно представлять изменение состояния сети как движение по поверхности функции энергии, подобно тому как это происходит, например, с помещенным на неровную поверхность шариком, который всегда

движется в направлении уменьшения своей потенциальной энергии (рис. 11). В точках пространства состояний, соответствующих образам-аттракторам (или некоторым их комбинациям), функция энергии имеет локальные минимумы, поэтому сеть, достигнув любого из таких состояний, остается в нем до тех пор, пока не будет принудительно переведена в новое начальное состояние. Все пространство состояний распадается, таким образом, на области, образующие так называемые бассейны притяжения аттракторов, и динамика сети, установленной в некоторое начальное состояние, полностью зависит от того, в бассейн притяжения какого из аттракторов она попала [Веденов, 1988; Amit, 1989].

3.2 Сигмоидная функция активации

В предыдущем пункте мы рассмотрели свойства сетей Хопфилда, построенных из нейронов с пороговой функцией активации. Дополнительный анализ показывает [Hopfield, 1984], что замена пороговых нейронов на сигмоидные (п. 1.2) не меняет принципиально основных свойств сети, то есть ее способности к хеббовскому запоминанию входных образов и их ассоциативному вспоминанию, а также к спонтанному формированию обобщенных и комбинированных образов. Сетевая модель с сигмоидными нейронами представляется более реалистичной по сравнению с сетью из пороговых нейронов, поскольку допускает значения, промежуточные между минимальным и максимальным.

3.3 Машина Больцмана и моделированный отжиг

Поведение сети Хопфилда полностью детерминировано. Однако легко получить стохастический вариант этой сети, называемый машиной Больцмана [Hinton, Sejnowski, 1986], если заменить ее пороговые нейроны на стохастические нейроны (см. 1.4), принимающие значения 1 и -1 с вероятностями

$$P(u_i = 1) = \frac{1}{1 + e^{-\frac{1}{T} \sum_{j=1}^n w_{ij} x_j}}, \quad P(u_i = -1) = 1 - P(u_i = 1) = \frac{1}{1 + e^{\frac{1}{T} \sum_{j=1}^n w_{ij} x_j}}.$$

При очень больших T вероятности переходов в состояния 1 или -1 слабо зависят от соответствующих этим переходам изменений энергии и примерно равны 1/2, что говорит о практической независимости динамики сети от рельефа энергетической поверхности. Наоборот, при $T = 0$ динамика машины Больцмана совпадает с динамикой нестохастической сети Хопфилда, поскольку при $T \rightarrow 0$ и

$$\sum_{j=1}^n w_{kj}(t) u_j(t) > 0 \text{ имеем } P(u_i = 1) \rightarrow 1 \text{ и } P(u_i = -1) \rightarrow 0, \text{ то есть получаем тот же}$$

результат, к которому бы привело и применение детерминированного правила.

При промежуточных температурах сеть с высокой вероятностью остается в состояниях, соответствующих глубоким минимумам энергии, и относительно легко выходит из состояний, соответствующих мелким минимумам. Это означает, что благодаря введению стохастичности сеть получает способность различать локальные минимумы энергии по степени их выраженности, то есть способность отличать надежные, многократно встреченные образы, которым соответствуют глубокие минимумы, от случайных.

Возможность свободного случайного блуждания сети в пространстве состояний расширяет возможности нахождения ею более глубоких минимумов энергии, то есть

способствует более успешному решению задач вспоминания. Эффективная процедура такого блуждания, названная «моделированным отжигом» из-за аналогии с известным способом термической обработки металла, состоит в том, что поиск начинается при высокой температуре сети, которая постепенно снижается до нуля [Kirkpatrick, Gelatt, Vecchi, 1983].

Сеть с элементами стохастичности как бы получает внутренний источник движения. Машина Больцмана, в отличие от нестохастической сети Хопфилда, достигнув ближайшего локального максимума энергии, уже не обречена находиться в нем до тех пор, пока ее состояние не будет изменено извне, а может сама выйти из этого состояния.

3.4 Запоминание последовательностей

При определении архитектуры сети Хопфилда предполагалось, что веса связей между нейронами симметричны, то есть $w_{ij}(t) = w_{ji}(t)$. Что произойдет, если это ограничение снять? Оказывается, что у асимметричной сети появляется новое полезное свойство – такая сеть может иметь не только точечные, но и циклические аттракторы и, соответственно, может запоминать не только отдельные образы, но и последовательности образов.

Для обеспечения возможности запоминания асимметричной сетью последовательности образов достаточно изменить правило Хебба таким образом, чтобы в весах связей фиксировались не только отношения между состояниями нейронов предъявляемого в момент t образа, но и отношения между состояниями элементов образа, предъявленного в момент $t + 1$, и образа, предъявляемого в момент t [Amari, 1972; Hopfield, 1982]. Соответствующее обучающее правило запишется следующим образом

$$w_{ij}(t+1) = w_{ij}(t) + u_i(t) \times u_j(t) + \lambda \times u_i(t+1) \times u_j(t),$$

где λ – константа, определяющая скорость перехода между последовательными образами.

Однако обученная описанным образом сеть может воспроизвести устойчиво лишь очень короткие последовательности. Для преодоления этого недостатка в [Kleinfeld, 1986] и [Sompolinsky, Kanter, 1986] были использованы механизмы, обеспечивающие запаздывание влияния переходной составляющей синаптических весов на динамику сети, а в ряде работ [Morita, 1993, 1994, 1996a, 1996b; Yoshizawa, Morita, Amari, 1993; Suemitsu, Morita, 2000; Morita, Murata, Morokami, 2003] для этой цели были применены нейроны с немонотонной функцией активации (п. 1.3).

Очевидно, что все ранее сказанное относительно целесообразности введения стохастичности в динамику сети в равной степени относится и к асимметричным сетям. Однако интерпретация динамики сети в терминах движения ее состояния только в направлении уменьшения функции энергии на случай асимметричной сети непосредственно не переносится. Действительно, как мы можем возвратиться в прежнее состояние, если на каждом шаге энергия может только уменьшиться? Можно, однако, использовать другой, более общий способ наглядного описания динамических свойств сети – векторное поле скоростей изменения состояния, заданное в пространстве состояний. Это поле легко строится по функции энергии, поскольку в каждой точке пространства состояний оно равно просто антиградиенту функции энергии. Поэтому в случае симметричных сетей мы можем

наряду с функцией энергии использовать для описания динамики также и поле скоростей. Однако в случае несимметричных сетей, которые могут иметь циклическую динамику, годится лишь описание в терминах поля скоростей. Более важно, однако, подчеркнуть, что в обоих случаях пространство состояний делится на непересекающиеся части, соответствующие бассейнам притяжения разных аттракторов, и динамика сети определяется ее движением в соответствии с полем скоростей по направлению к аттрактору, в бассейне притяжения которого находится ее состояние.

Отметим также, что полное поле скоростей сети не исчерпывается полем скоростей в пространстве состояний нейронов. Поскольку веса связей также изменяются в зависимости от состояний нейронов (а через них, следовательно, и в зависимости от значений других весов), то в принципе можно рассматривать поле скоростей в объединенном пространстве состояний элементов и весов связей.

4 Импульсные сети

В перцептронах и сетях Хопфилда, рассмотренных выше, в качестве элементов используются модельные нейроны, функция активации которых ассоциируется с частотой генерации спайков реальным нейроном. Это довольно грубое приближение, хотя, как мы видели, сети из таких нейронов способны моделировать основные когнитивные функции мозга. Тем не менее имеются существенные сомнения в их достаточной биологической адекватности в качестве моделей реальных когнитивных процессов. Поэтому в настоящее время существует нарастающая тенденция использовать в качестве нейробиологических моделей импульсные сети, то есть сети, построенные из импульсных нейронов (п. 1.5). Иногда их даже называют сетями третьего поколения, противопоставляя сетям из нейронов с пороговыми и аналоговыми (например, сигмоидными) функциями активации [Maass, 1996; Florian, 2003]. Тем более что можно показать, что любая сеть Хопфилда (включая сети с асимметричными связями) может быть эмулирована импульсной сетью [Maass, 1997; Maass, Natschläger, 1998].

В качестве примера несоответствия классических активационных моделей нейронов биологической реальности рассматривается, в частности, такая ситуация. Показано, что анализ зрительного образа, включающий не менее 10 синаптических стадий, начиная с сетчатки и кончая височными долями, может быть осуществлен человеком за 100 мс [Perrett, Rolls, Caan, 1982; Thorpe, Imbert, 1989]. В то же время на регистрацию частоты спайков на одной стадии требуется по меньшей мере 20–30 мс [Rolls, Tovee, 1994]. Это приводит к выводу о том, что существенные процессы обработки информации в мозгу человека, по-видимому, происходят уже на уровне отдельных спайков.

4.1 Обучение

Хеббовское обучение в случае импульсных сетей приобретает некоторую специфику по сравнению с сетями, составленными из активационных нейронов, – оно асимметрично во времени. А именно в импульсных сетях более полно учитывается соотношение между временем появления импульса на синаптическом входе нейрона и временем генерации его собственного импульса. Максимальное приращение синаптического веса происходит в случае, когда выходной импульс генерируется немедленно после входного, а при увеличении запаздывания это приращение уменьшается. Если же, наоборот, момент генерации выходного

импульса предшествует появлению сигнала на синапсе, то вес этого синапса уменьшается, причем максимальное уменьшение соответствует минимальному времени предшествования. Зависимость степени изменения синаптического веса от степени запаздывания выходного импульса по отношению к входному, заданная формулой

$$\Delta w = 0,25e^{-0,05|t_{\text{вых}} - t_{\text{вх}}|} (3 \operatorname{sgn}(x) + 1),$$

схематически показана на рис. 12 [Vi, 2002].

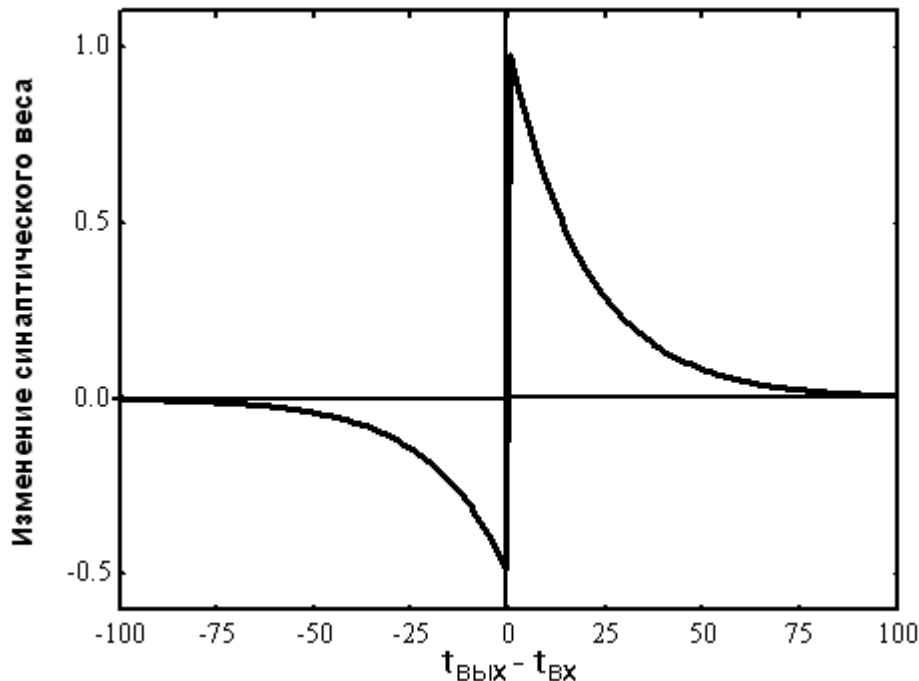


Рис. 12. Относительное изменение синаптического веса в зависимости от степени запаздывания выходного импульса по отношению к входному.

Обучение, зависящее от соотношения времен следования входных и выходных импульсов, имеет то преимущество перед классическим хеббовским обучением, что оно асимметрично во времени и, таким образом, фиксируя возможные причинные связи между событиями, создает базу для прогнозирования будущего.

4.2 Стохастичность

Введение стохастичности (шума) в нейронную сеть целесообразно как с точки зрения ее биологического правдоподобия – реальные нейронные сети подвержены внутренним и внешним случайным влияниям, так и с точки зрения ее эффективной работы – для обеспечения возможности выхода из ложных зацикливаний.

В импульсных сетях стохастичность может быть введена путем добавления шума либо к уровню порога, либо к длине рефрактерного периода, либо к величине мембранного потенциала.

4.3 Запоминание последовательностей

Асимметричное хеббовское правило, описанное выше, хорошо подходит для запоминания временных последовательностей [Gerstner, Ritz, van Hemmen, 1993; Gerstner, Kistler, 2002]. Если один паттерн нейронов активизируется раньше другого

паттерна, то синаптические связи от нейронов первого паттерна к нейронам второго усилятся. Аналогичным образом усилятся связи от нейронов второго паттерна к нейронам третьего паттерна, если нейроны второго паттерна активизируются раньше нейронов третьего паттерна, и т.д. В результате, если после обучения на вход сети подать образ, соответствующий активности нейронов первого паттерна, то он будет стимулировать последующую активность второго паттерна, затем – третьего и т.д. Таким образом сеть воспроизведет всю последовательность.

Часто важно, чтобы сеть воспроизвела последовательность в том же темпе, в каком она была предъявлена, например, при запоминании мелодий. Эта способность может быть, в частности, обеспечена, если пары нейронов связаны не одной связью, а несколькими, различающимися временем аксональной задержки. В этом случае в наибольшей степени будет увеличиваться вес той связи, задержка которой ближе всего к разности моментов активизации пресинаптического и постсинаптического нейронов [Gerstner, Kempter, van Hemmen, Wagner, 1996; Herz, Sulzer, Kuhn, van Hemmen, 1989; Senn, Schneider, Ruf, 2002].

5 Структурные сетевые модели

До сих пор, обсуждая принципиальные когнитивные свойства сетей Хопфилда или импульсных сетей, мы не уточняли их структуру, фактически считая их полносвязными и однородными. Однако нейронная сеть реального мозга неоднородна – она состоит из нескольких, хотя и тесно связанных друг с другом, но анатомически, структурно и функционально различающихся подсетей. В этом разделе мы рассмотрим несколько моделей функциональных систем мозга, учитывающих эту структурированность.

5.1 Гиппокамп

В [Hasselmo, Wyble, 1997; Hasselmo, McClelland, 1999] была предпринята попытка построить многокомпонентную нейросетевую модель гиппокампа. Ее общая схема представлена на рис. 13.

Предполагается, что входные образы из второго слоя энторинальной коры поступают в зубчатую извилину, где происходит процесс самоорганизации, в результате которого формируются представления образов с меньшим взаимным перекрытием. Этот этап очень важен, поскольку гиппокамп должен обладать способностью возможно точно фиксировать огромный объем поступающей информации. Затем активность распространяется на область СА3 гиппокампа, где наличие большого числа возбуждающих рекуррентных связей между нейронами обеспечивает установление ассоциативных связей между различными частями образов и их ассоциативный вызов. Входные образы из энторинальной коры также поступают в область СА1 гиппокампа, формируя новые представления входных образов, которые сравниваются с представлениями этих образов, сформированными в СА3. Между СА3 и СА1 имеется гетероассоциативная связь. Имеются также пути от области СА1 к энторинальной коре, поэтому между этими частями также формируются взаимные ассоциативные связи. Область СА1 через связь с медиальным септумом регулирует холинэргическую модуляцию, обеспечивающую контроль динамики установления ассоциативных связей в гиппокампе.



Рис. 13. Модель гиппокампа.

Описанная модель была использована для выяснения механизма действия на память инъекций блокатора ацетилхолиновых рецепторов скополамина. В частности, на ней был воспроизведен эффект более слабого отрицательного воздействия скополамина на узнавание образов по сравнению с воздействием на их вспоминание.

В реальном эксперименте испытуемым обычно предъявляется список слов, который автоматически сопровождается контекстом – обстановкой эксперимента. Задача вспоминания формулируется как просьба воспроизвести список слов, а задача узнавания – как просьба указать, какие слова из вновь предъявляемых входили в ранее предъявленный список.

В численном эксперименте во время обучения на одну часть входа модельной энторинальной коры подавались сигналы, соответствующие контексту, а на другую – запоминаемым словам. Задача вспоминания имитировалась активизацией в модели энторинальной коры входов, соответствующих контексту, а задача узнавания – активизацией входов, соответствующих словам. При успешном вспоминании активизация контекста приводила к последовательному воспроизведению запомненных слов, а при успешном узнавании активизация слова приводила к активизации контекста.

Воздействие скополамина в моделировании имитировалось уменьшением интенсивности усиления синаптических связей. Это нарушало процесс

вспоминания, но не нарушало успешность узнавания. Причина состоит в том, что контекст предъясняется чаще, чем текст, поэтому связанные с контекстом синапсы усиливаются в большей степени, и, как следствие, для его активизации достаточно даже относительно слабой ассоциации в виде слова из числа ранее предъясвленных.

Детальное описание данного подхода к моделированию гиппокампа можно найти в [Hasselmo, Wyble, Cannon, 2001].

5.2 Гиппокамп – неокортекс

Ранее на примере сети Хопфилда было проиллюстрировано, что нейронная сеть хранит запоминаемые образы распределенными во всей системе синаптических весов. Отмечалось также, что в результате такого совмещенного хранения может происходить своего рода интерференция, порождающая новые химерические образы. Этот процесс может приобрести катастрофический характер, если число запоминаемых образов будет слишком велико [McCloskey, Cohen, 1989; Ratcliff, 1990]. Однако в человеческом мозгу такой катастрофической интерференции не происходит. В [McClelland, McNaughton, O'Reilly, 1995] было выдвинуто предположение, что этого удастся избежать благодаря тому, что функция запоминания осуществляется совместно двумя системами мозга – гиппокампом и новой корой: новые образы сначала запоминаются гиппокампом, после чего постепенно передаются в новую кору.

В ряде работ это предположение проверялась на двухкомпонентных нейросетевых моделях [Ans, Rousset, 1997; French, 1997; French, Ans, Rousset, 2001], в основу которых была положена идея об образовании в мозгу так называемых псевдообразов [Robins, 1995, 1996]. Робинс предположил, что для избежания катастрофической интерференции при запоминании нового образа сеть дополнительно запоминает серию псевдообразов, отражающих ранее запомненные образы, то есть как бы повторяет, репетирует их в разных вариантах. Путем нейромоделирования было показано, что такой режим запоминания действительно снижает шансы появления катастрофической интерференции [Robins, McCallum, 1998].

Двухкомпонентная модель [French, Ans, Rousset, 2001] включает две нейронных сети, одну из которых можно ассоциировать с гиппокампом, а вторую – с новой корой. Новая информация поступает в гиппокамп, но туда приходят также и псевдообразы, сгенерированные на основе образов, полученных ранее новой корой из гиппокампа. Псевдообразы также используются для передачи новой информации из гиппокампа в новую кору. Показано, что такая процедура уменьшает взаимное искажение между старой и новой информацией.

Предполагается также, что обмен псевдообразами между гиппокампом и корой лежит в основе процесса консолидации долговременной памяти во сне [Robins, McCallum, 1998; Stickgold, 1999; Sutherland, McNaughton, 2000]. Кроме того, в рамках этой модели удастся объяснить загадочный феномен избирательного ослабления памяти при анонии [French, 1997; French, Mareschal, 1998], а также механизмы антероградной и ретроградной амнезии [Ans, Rousset, 1997].

5.3 Префронтальная кора – базальные ганглии

Модель взаимодействия гиппокампа и новой коры, рассмотренная в предыдущем пункте, имела целью выяснение механизмов формирования долговременной памяти. Рассматриваемая ниже модель [Frank, Loughry, O'Reilly, 2001; O'Reilly, Frank, 2005], основанная на моделировании взаимодействия префронтальной коры и базальных ганглиев, построена для изучения механизмов функционирования кратковременной, рабочей памяти.

Практически общепризнано, что префронтальная кора играет центральную роль в функционировании рабочей памяти. Также известно, что базальные ганглии тесно взаимодействуют с префронтальной корой и их повреждение вызывает такие же когнитивные нарушения, как и повреждение префронтальной коры. Соответственно, возникает проблема разделения функций этих двух структур, решение которой было одной из целей при построении нейросетевой модели системы «префронтальная кора – базальные ганглии».

Рабочая память используется при решении текущих когнитивных задач, позволяя временно сохранять и в нужный момент предоставлять необходимые информационные компоненты, например, при проведении в уме арифметических вычислений или при анализе смысла предложения. В модели предполагается, что, в отличие от долговременной памяти, где информация хранится в весах связей, в рабочей памяти она хранится в виде паттернов возбужденных нейронов, что обеспечивает оперативность взаимодействия с ней, но ограничивает ее потенциальную емкость и осложняет хранение.

Стандартный тест для рабочей памяти состоит в том, что испытуемому предъявляется произвольная последовательность букв и цифр и он должен положительно реагировать на появление X после A, если последней из цифр 1 и 2 появилась цифра 1, и должен положительно реагировать на появление Y после B, если последней появилась цифра 2. Очевидно, что при выполнении этого теста испытуемый должен попеременно хранить в рабочей памяти последнюю из этих цифр, заменяя ее в нужный момент другой. Таким образом, желательно, чтобы имелся механизм, принимающий решение о пропуске или непропуске очередного стимула в рабочую память. Основным постулат модели состоит в том, что такой «пропускной контроль» (gating в терминологии авторов) существует и осуществляется структурой базальных ганглиев, что имеет биологическое основание [Deniau, Chevalier, 1985].

Общая схема модели представлена на рис. 14. Существенная роль в ней отведена формированию в среднем мозге условного рефлекса (в модели этот процесс осуществляется с помощью биологически обоснованного аналога алгоритма обратного распространения ошибки). Проверка продемонстрировала способность модели успешно выполнять широкий круг тестов рабочей памяти.

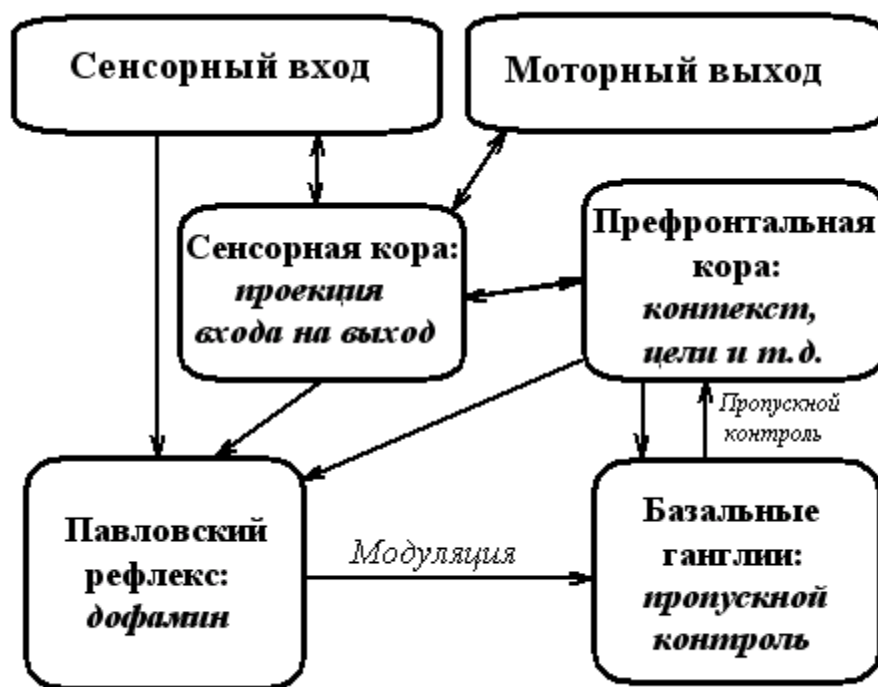


Рис. 14. Модель системы «префронтальная кора – базальные ганглии».

5.4 Основные принципы моделирования мозга

В [O'Reilly, 1998] сформулирован ряд принципов, на которых, по мнению автора, должно базироваться моделирование мозга. Три принципа касаются архитектуры модели и взаимодействия составляющих ее нейроноподобных элементов. И еще два принципа относятся к способам обучения. Но основным является требование биологического правдоподобия модели. По крайней мере первые пять принципов должны ему удовлетворять.

Три архитектурных принципа – это распределенное представление, двустороннее взаимодействие и конкурентное ингибирование. Два принципа обучения состоят в том, что должны быть использованы оба принципиально возможных типа обучения – обучение без учителя (самоорганизация хеббовского типа) и обучение с учителем (обучение на ошибках типа дельта-правила).

Распределенный характер представления информации в мозге подтверждается электрофизиологическими исследованиями. При таком представлении информация хранится во многих нейронах и один нейрон участвует в хранении различной информации. С функциональной точки зрения оно обеспечивает эффективность использования информации, а также ее точность и надежность. Кроме того, распределенное представление дает возможность отражать сходство между разными единицами хранимой информации – через общность используемых нейронов.

Двустороннее взаимодействие нейронов можно противопоставить однонаправленному потоку информации, проходящему через них, как это имело место в перцептронах. Такое взаимодействие объединяет в один процесс разные уровни обработки информации, благодаря чему, например, анализ на уровне слов помогает правильному восприятию букв, то есть анализу на более низком уровне.

Конкурентное ингибирование приводит к ограничению числа одновременно активных нейронов. Кора содержит около 20% ингибирующих интернейронов, которые предотвращают взрыв активности связанных стимулирующими связями пирамидальных нейронов (что происходит, например, при эпилепсии). Благодаря ингибированию создается так называемое разреженное представление информации в нейронной сети мозга (своего рода «контрастирование»), что способствует ее быстрой и эффективной обработке, например снижает риск попадания состояния сети в ложные локальные минимумы.

Обучение сети требует решения двух различных, хотя и взаимодополняющих, проблем: построения модели окружающей среды и принятия правильного решения в конкретной ситуации. Эти задачи решаются двумя различными способами: первая – с помощью «хеббовского обучения», вторая – с помощью «обучения на ошибках».

Биологическая правдоподобность хеббовского обучения общепризнана – оно может реализоваться через НМДА-опосредованный механизм долговременной потенциации синаптических связей. Хеббовское обучение фиксирует коррелированность, частоту встречаемости различных характеристик внешнего мира, что позволяет создать не только сетевое представление его структуры, но и отразить существующие в нем причинно-следственные связи.

Обучение на ошибках, например метод обратного распространения ошибки, доказало свою эффективность в вычислительных применениях нейронных сетей. Его цель – научиться выдавать нужный ответ на предъявленный вход на основе обучения методом проб и ошибок на серии ситуаций с известными ответами. Хотя с психологической точки зрения такая постановка задачи вполне естественна, трудно было найти нейробиологическое обоснование для процесса «обратного распространения ошибки». Было также неясно, откуда может прийти сигнал о правильном ответе. Однако все же удалось предложить такую модификацию метода обратного распространения ошибки, которая может иметь биологическое обоснование [O'Reilly, 1996]. Идея состоит в том, что вместо обратного распространения ошибки распространяются входной и выходной сигналы и сравниваются модификации весов, соответствующие этим двум сигналам. Следует также отметить, что два типа обучения не просто дополняют друг друга, но и синергетически взаимодействуют: при обучении решению задач методом проб и ошибок активно используются знания, ранее полученные методом корреляционного обучения.

Обсуждение и перспективы исследований

В этом обзоре мы попытались отразить как классические (до конца 80-х годов), так и более современные основные подходы к нейросетевому моделированию когнитивной деятельности мозга. Если с классикой все более или менее понятно, то с анализом современного состояния этой области научных исследований дело обстоит сложнее. Уже просто потому, что за последние 15–20 лет ее развитие носило взрывной характер и общее число публикаций по нейронным сетям измеряется многими тысячами. В качестве характерного примера размаха нейросетевых исследований можно привести монографию В.Герстнера и В.М.Кистлера, посвященную импульсным сетям [Gerstner, Kistler, 2002], то есть относительно новому направлению нейромоделирования, – она содержит более 450 страниц текста и около 500 ссылок на литературу.

К счастью, не было больших проблем с получением полных текстов нужных публикаций – большинство из них удалось найти в Интернете, иногда даже полные тексты книг [напр., Gerstner, Kistler, 2002]. Находились они, главным образом, в свободном доступе на личных сайтах авторов. Как ни странно, научные журналы, некогда созданные для распространения научных достижений, превратились в тормоз на этом пути – во многих из них доступ платный и довольно дорогой, и часто они запрещают размещать рукопись в Интернете, как условие публикации в журнале, таким образом, ограничивая ее доступность для научного сообщества.

Среди массы работ по нейромоделированию есть, конечно, и публикации обзорного характера, некоторые из которых довольно информативны и легко воспринимаются, как, например, технический отчет Р.В.Флориана [Florian, 2003], посвященный особенностям использования нейронных сетей в робототехнике. Однако, к сожалению, не удалось найти текста, который можно было бы использовать как основу для данной работы. Некоторые из обзоров носили слишком общий характер, другие были перегружены деталями, третьи касались слишком частных вопросов. Поэтому пришлось, в основном, пытаться понять современное состояние нейросетевого моделирования, обращаясь непосредственно к оригинальным публикациям. Наверное, и этот обзор в чем-то не удовлетворит некоторых читателей, и им придется обращаться к первоисточникам.

На наш взгляд, основная дилемма нейросетевого моделирования на данный момент состоит в выборе между активационными и импульсными сетями. На первый взгляд, ситуация проста. Очевидно, что импульсные сети более детально и биологически более правдоподобно моделируют реальные сети. Однако, как и в случае любой более сложной модели, ее труднее реализовать и труднее интерпретировать полученные с ее помощью результаты. Кроме того, из-за увеличения сложности небольшие ошибки в локальных предпосылках модели могут повлечь катастрофические отклонения от реальности в ее глобальном поведении.

Общее правило состоит в том, что для решения любой задачи следует выбирать самую простую модель среди тех, которые обеспечивают решение поставленной задачи. Поэтому стандартная рекомендация могла бы быть такой: если активационная модель обеспечивает на уровне экспериментальных наблюдений адекватное описание явления и дает приемлемые и полезные предсказания поведения исследуемого объекта в новых ситуациях, то ее и следует использовать. Если же некоторые наблюдаемые эффекты ею не воспроизводятся, то следует изменить модель, – например, воспользоваться импульсной моделью.

Однако ситуация не так однозначна – некоторые новые техники программирования могут упростить компьютерную реализацию импульсной нейросетевой модели до такой степени, что она станет проще компьютерной реализации активационной модели, а это, в свою очередь, может, например, позволить на порядок увеличить число нейронов в модельной сети [Gerstner, 1999; Mattia, Del Giudice, 2000; Mariani, Reilly R.; Maskey, 2002]. В конечном результате может оказаться, что импульсная модель в каком-то смысле проще активационной. В конце концов, спайки, если абстрагироваться от их точной формы и кодировать как бинарные переменные (0 или 1), реализуются проще, чем сигмоидная функция $1/(1 + e^{-\sigma(x-\theta)})$.

Из сказанного следует, что правильный выбор между активационными и импульсными сетями можно сделать, только проинвентаризовав и

проанализировав созданное на сегодняшний день богатое программное обеспечение нейросетевого моделирования (мы не относим сюда нейросетевые программы обработки данных и нейросетевые системы технического назначения). Вопросы эффективной программной реализации нейросетевых моделей остались за рамками данного обзора ввиду их технического характера, но, как мы видим, они выходят на первый план.

С содержательной точки зрения проведенный анализ состояния исследований в области нейросетевого моделирования приводит к следующему выводу. Имеется огромное количество работ, решающих частные проблемы – либо биологического, либо математического характера. И есть дефицит работ, которые, используя эту базу, получали бы новые научные результаты более высокого уровня, такие, которые могли бы уже иметь практическое применение в психологии и педагогике. Компьютерное моделирование и экспериментальные исследования полезны и даже необходимы и в этом случае, но они отступают на второй план, служат для подтверждения или опровержения выводов и догадок, полученных на основании комплексного анализа работ других авторов.

Более конкретно – представляется перспективным в таком плане изучить феномен воспоминания, точнее, применить нейросетевое моделирование для выяснения нейробиологических механизмов различных типов сбоев при решении задачи воспоминания здоровым мозгом и разработки на этой основе методов их преодоления.

Модель аттракторных нейронных сетей описывает сущность этого процесса как попадание в бассейн притяжения вспоминаемого образа и движения в направлении образа-аттрактора в пространстве состояний сети. Процедура моделированного отжига (повышение температуры сети с последующим ее постепенным понижением) может быть положена в основу практической методики повышения эффективности решения задачи воспоминания. Имеется и ряд других полученных при анализе модельных нейронных сетей результатов, раскрывающих особенности механизмов работы реального мозга, которые могли бы быть переработаны в комплекс методик по «гигиене памяти». Необходимы, однако, большие усилия по синтезу и развитию достижений, полученных в нейросетевом моделировании, нейрофизиологии и психологии, чтобы получить новые научные результаты, которые могли стать основой для практических рекомендаций. О.С.Виноградова формулирует такой подход следующим образом [Виноградова, 2000]: «Нейробиология – многоуровневая наука, исследования которой ведутся на молекулярном – субклеточном – клеточном – структурном – межструктурном (мозговом) – поведенческом (когнитивном) уровнях. Каждый уровень использует синтетические данные предыдущего и представляет аналитические данные для последующего уровня; перепрыгивание через уровень в логических и модельных построениях недопустимо».

Литература

Веденов А.А. Моделирование элементов мышления. М.: Наука, 1988. 159 с.

Виноградова О.С. Нейронаука конца второго тысячелетия: смена парадигм // Журнал высш. нервн. деятел. 2000. Т. 50. С. 743–774.

Дунин-Барковский В.Л., Терехин А.Т. Нейронные сети и нейрокомпьютеры: тенденции развития исследований и разработок // Микропроцессорные средства и системы. 1990. N 2. С. 12–14.

Мак-Каллок У.С., Питтс У. Логическое исчисление идей, относящихся к нервной активности // Автоматы / пер. с англ. / ред. К.Шеннон, Дж.Маккарти. М.: ИЛ, 1956. С. 362–384.
(McCulloch W.S., Pitts W. A logical calculus of the ideas immanent in neurons activity. Bull. Math. Biophys. 1943. Vol. 5. P. 115–133. [In English])

Минский М., Пейперт С. Перцептроны / пер. с англ. М.: Мир, 1971. 261 с. (Minsky M., Papert S. Perceptrons: An Introduction to Computational Geometry. Cambridge, MA: MIT Press, 1969. [In English])

Розенблат Ф. Принципы нейродинамики. Перцептроны и теория механизмов мозга / пер. с англ. М.: Мир, 1965. 175 с. (Rosenblatt F. Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. Washington, D.C.: Spartan Books, 1962. [In English])

Терехин А.Т., Будилова Е.В. Сетевые механизмы биологической регуляции // Успехи физиологических наук. 1995. Т. 26. N 4. С. 75–97.

Amit S. Learning patterns and pattern sequences by self-organizing nets of threshold elements // IEEE Trans. Comput. 1972. Vol. 21. P. 1197–1206.

Amit D.J. Modeling brain function. The world of attractor neural networks. New York: Cambridge University Press, 1989.

Ans B., Rousset S. Avoiding catastrophic forgetting by coupling two reverberating neural network // Comptes Rendus de l'Academie des Sciences – Series III – Sciences de la Vie. 1997. Vol. 320. P. 989–997.

Budilova E.V., Teriokhin A.T. A bibliographic data base on neural networks and neurocomputers // The RNS/IEEE Symposium on Neuroinformatics and Neurocomputers. Rostov-on-Don, 1992. New York: IEEE, 1992. Vol. 2. P. 1125–1126.

Deniau J.M., Chevalier G. Disinhibition as a basic process in the expression of striatal functions: II. The striato-nigral influence on thalamocortical cells of the ventromedial thalamic nucleus // Brain Research, 1985. Vol. 334. P. 227–233.

Florian R.V. Biologically inspired neural networks for the control of embodied agents // Center for Cognitive and Neural Studies (Cluj-Napoca, Romania), Technical Report Coneural-03-03, 2003. URL: <http://www.coneural.org>

Frank M.J., Loughry B., O'Reilly R.C. Interactions between the frontal cortex and basal ganglia in working memory: A computational model // Cognitive, Affective, and Behavioral Neuroscience. 2001. Vol. 1. P. 137–160.
URL: http://psych.colorado.edu/oreilly/paprrs/FrankLoughryOReilly014_fcbg.pdf

French R.M. Pseudo-recurrent connectionist networks: An approach to the “sensitivity–stability” dilemma // Connection Science. 1997. Vol. 9, Issue 4. P. 353–379.
URL: <http://www.ulg.ac.be/cogsci/rfrench/psdnet.pdf>

French R.M. Selective memory loss in aphasics: An insight from pseudorecurrent connectionist networks. Connectionist Representations // Proceedings of the Fourth Neural Computation and Psychology Workshop / eds. J.Bullinaria, G.Houghton, D.Glasspool. Berlin: Springer-Verlag, 1997. P. 183–195. URL: <http://www.ulg.ac.be/cogsci/rfrench/aphasics.pdf>

French R.M. Catastrophic Forgetting in Connectionist Networks // Trends in Cognitive Sciences. 1999. Vol. 3, Issue 4. P. 128–135. URL: http://www.ulg.ac.be/cogsci/rfrench/TICS_cat_forget.pdf

French R.M., Ans B., Rousset S. Pseudopatterns and dual-network memory models: Advantages and shortcomings // *Connectionist Models of Learning, Development and Evolution* / eds. R. French, J. Sougné. London: Springer, 2001. P. 13–22.

URL: http://www.ulg.ac.be/cogsci/rfrench/ncpw6_french_ans_rousset.pdf

French R.M., Mareschal D. Could category-specific semantic deficits reflect differences in the distributions of features within a unified semantic memory? // *Proceedings of the Twentieth Annual Cognitive Science Society Conference*. NJ: LEA, 1998. P. 374–379.

URL: http://www.ulg.ac.be/cogsci/rfrench/anomia_98.pdf

Gerstner W. Time structure of the activity in neural network models // *Phys. Rev. Evol.* 1995. Vol. 51. P. 738–758.

URL: <http://diwww.epfl.ch/~gerstner/PUBLICATIONS/Gerstner95PRE.pdf>

Gerstner W. Spiking neurons // *Pulsed neural networks* / eds. Maass W., Bishop C.M. Cambridge, MA: MIT Press, 1999. P. 3–53.

Gerstner W., Kempter R., van Hemmen J.L., Wagner H. A neuronal learning rule for sub-millisecond temporal coding // *Nature*. 1996. Vol. 383. P. 76–78.

URL: <http://diwww.epfl.ch/~gerstner/PUBLICATIONS/Nature96-text.pdf>

URL: <http://diwww.epfl.ch/~gerstner/PUBLICATIONS/Nature96-figs.pdf>

Gerstner W., Kistler W.M. Spiking Neuron Models: Single Neurons, Populations, Plasticity.

Cambridge, MA: Cambridge Univ. Press, 2002. URL: <http://diwww.epfl.ch/~gerstner/BUCH.html>

Gerstner W., Ritz R., van Hemmen J.L. Why spikes? Hebbian learning and retrieval of time-resolved excitation patterns // *Biol. Cybern.* 1993. Vol. 69. P. 503–515.

Hasselmo M.E., Cannon R.C., Koene R. A simulation of parahippocampal and hippocampal structures guiding spatial navigation of a virtual rat in a virtual environment: A functional framework for theta theory // *The Parahippocampal Region: Organization and Role in Cognitive Function* / eds. M.P. Witter and F. Wouterlood. London: Oxford Univ. Press, 2002.

Hasselmo M.E., McClelland J.L. Neural models of memory // *Current Opinion in Neurobiology*. 1999. Vol. 9. P. 184–188. URL: <http://biomednet.com/elecref/0959438800900184>

Hasselmo M.E., Wyble B.P. Simulation of the effects of scopolamine on free recall and recognition in a network model of the hippocampus // *Behav. Brain Res.* 1997. Vol. 89. P. 1–34.

Hasselmo M.E., Wyble B.P., Cannon R.C. From spike frequency to free recall: How neural circuits perform encoding and retrieval // *The Cognitive Neuroscience of Memory: Encoding and Retrieval* / eds. A. Parker, T.J. Bussey, E. Wilding. London: Psychology Press, 2001.

URL: <http://people.bu.edu/hasselmo/BusseyChapterFinal.pdf>

Hebb D.O. The Organization of Behavior. A Neuropsychological Theory. New York: Wiley, 1949.

Herz A.V.M., Sulzer B., Kühn R., van Hemmen J. L. Hebbian learning reconsidered:

Representation of static and dynamic objects in associative neural nets // *Biol. Cybern.* 1989. Vol. 60. P. 457–467.

Hinton G.E., Sejnowski T.J. Learning and relearning in Boltzman machines // *Parallel Distributed Processing. Explorations in the Microstructure of Cognition*. Cambridge, MA: MIT Press, 1986. Vol. 1. P. 282–317.

Hinton G.E., McClelland J.L. Learning representations by recirculation // *Neural information processing systems* / ed. D.Z. Anderson. New York: American Institute of Physics, 1988. P. 358–366.

Hodgkin A.L., Huxley A.F. A quantitative description of membrane current and its application to conduction and excitation in nerve // *Journal of Physiology*. 1952. Vol. 117. P. 500–544.

Hopfield J.J. Neural networks and physical systems with emergent collective computational abilities // *Proc. Natl. Acad. Sci. USA*. 1982. Vol. 79, Issue 8. P. 2554–2558.
URL: <http://www.pnas.org/cgi/reprint/79/8/2554>

Hopfield J.J. Neurons with graded response have collective computational properties like those of two-state neurons // *Proc. Natl. Acad. Sci. USA*. 1984. Vol. 81. P. 3088–3092.
URL: <http://www.pnas.org>

Izhikevich E.M. Simple model of spiking neurons // *IEEE Transactions on Neural Networks*. 2003. Vol. 14, Issue 6. P. 1569–1572. URL: <http://www.nsi.edu/users/izhikevich/publications/spikes.htm>

Izhikevich E.M. Which model to use for cortical spiking neurons // *IEEE Transactions on Neural Networks*. 2003. Vol. 15. P. 1063–1070.
URL: <http://www.nsi.edu/users/izhikevich/publications/whichmod.htm>

Kirkpatrick S., Gelatt C.D., Vecchi M.P. Optimization by simulated annealing // *Science*. 1983. Vol. 220. P. 671–680.

Kleinfeld D. Sequential state generation by model neural networks, *Proc. Nat. Acad. Sci. USA*. 1986. Vol. 83. P. 9469–9473. URL: <http://www.pnas.org>

Maass W. Networks of Spiking Neurons: The Third Generation of Neural Network Models // *Neural Networks*. 1997. Vol. 10, Issue 9. P. 1659–1671. URL: <http://www.eccc.uni-trier.de/eccc>

Maass W., Natschläger T. Associative memory with networks of spiking neurons in temporal coding // *Neuromorphic Systems: Engineering Silicon from Neurobiology* / eds. L.S. Smith, A. Hamilton. Amsterdam: World Scientific, 1998. P. 21–32.
URL: <http://www.cis.tugraz.at/igi/maass/psfiles/htc-ewns1.ps.gz>

Marian I., Reilly R.G., Mackey D. Efficient event-driven simulation of spiking neural networks // *Proceedings of 3rd WSES International Conference on Neural Networks and Applications*. Interlaken, Switzerland, February 2002.
URL: http://cortex.cs.nuim.ie/papers/IM_RR_DM_EventDrivenSim.pdf

Mattia M., Del Giudice P. Efficient event-driven simulation of large networks of spiking neurons and dynamical synapses // *Neural Comp.* 2000. Vol. 12. P. 2305–2329.
URL: <http://neural.iss.infn.it/Papers/asynsimu.pdf>

McClelland J., McNaughton B., O'Reilly R. Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory // *Psychological Review*. 1995. Vol. 102. P. 419–457.
URL: <http://www.cnbc.cmu.edu/~jlm/papers/McCMcNaughtonOReilly95.pdf>

McCloskey M., Cohen N. Catastrophic interference in connectionist networks: The sequential learning problem // *The Psychology of Learning and Motivation*. 1989. Vol. 24. P. 109–165.

Morita M. Associative memory with nonmonotone dynamics // *Neural Networks*. 1993. Vol. 6. P. 115–126. URL: <http://volga.esys.tsukuba.ac.jp/~mor/paper/>

Morita M. Smooth recollection of a pattern sequence by nonmonotone analog neural networks // *Proceedings of the IEEE International Conference on Neural Networks, Orlando, 1994*. Vol. 2. P. 1032–1037.

Morita M. Memory and Learning of Sequential Patterns by Nonmonotone Neural Networks // *Neural Networks*. 1996. Vol. 9, Issue 8. P. 1477–1489.

URL: <http://volga.esys.tsukuba.ac.jp/~mor/paper/mor1996a.pdf>

Morita M. Computational study on the neural mechanism of sequential pattern memory // *Cognitive Brain Research*. 1996. Vol. 5. P. 137–146.

URL: <http://volga.esys.tsukuba.ac.jp/~mor/paper/mor1996b.pdf>

Morita M., Murata K., Morokami S. Context-dependent sequential recall by a trajectory attractor network with selective desensitization // *Proceedings of the Third International Conference on Neural Networks and Artificial Intelligence, Minsk*. 2003. P. 235–238.

URL: <http://volga.esys.tsukuba.ac.jp/~mor/paper/ICNNAI2003.pdf>

O'Reilly R.C. Biologically Plausible Error-driven Learning using Local Activation Differences: The Generalized Recirculation Algorithm // *Neural Computation*. 1996. Vol. 8. P. 895–938.

URL: http://psych.colorado.edu/~oreilly/papers/OReilly96_generec_nc.pdf

O'Reilly R.C. Six Principles for Biologically-Based Computational Models of Cortical Cognition // *Trends in Cognitive Sciences*. 1998. Vol. 2. P. 455–462.

URL: http://psych.colorado.edu/~oreilly/papers/OReilly98_6princip_tics.pdf

O'Reilly R.C., Frank M.J. Making Working Memory Work: A Computational Model of Learning in the Frontal Cortex and Basal Ganglia // *Neural Computation*. 2006. Vol. 18. P. 283–328.

URL: <http://www.u.arizona.edu/~mfrank/papers/OReillyFrank06.pdf>

O'Reilly R.C., Munakata Y. Computational Neuroscience and Cognitive Modeling // *Encyclopedia of Cognitive Sciences* / ed. L. Nadel. London: Macmillan, 2003.

URL: http://psych.colorado.edu/~oreilly/papers/OReillyMunakata01_ecs_encyc.pdf

Perrett D.I., Rolls E.T., Caan W.C. Visual neurons responsive to faces in the monkey temporal cortex // *Experimental Brain Research*. 1982. Vol. 47. P. 329–342.

Ratcliff R. Connectionist models of recognition memory: Constraints imposed by learning and forgetting functions. *Psychological Review*. 1990. Vol. 97. P. 285–308.

Robins A. Catastrophic forgetting, rehearsal, and pseudorehearsal // *Connection Science*. 1995. Vol. 7. P. 123–146.

Robins A. Consolidation in neural networks and in the sleeping brain // *Connection Science*. 1996. Vol. 8. P. 259–275.

Robins A., McCallum S. Pseudorehearsal and the catastrophic forgetting solution in Hopfield type networks // *Connection Science*. 1998. Vol. 10. P. 121–135.

Rolls E.T., Tovee M.J. Processing speed in the cerebral cortex and the neurophysiology of visual backward masking // *Proc. Roy. Soc. B*. 1994. Vol. 257. P. 9–15.

Rumelhart D.E., Hinton G.E., Williams R.G. Learning representation by back-propagating error // *Nature*. 1986. Vol. 323, Issue 6088. P. 533–536.

Ruppin E. Evolutionary embodied agents: A neuroscience perspective // *Nature Reviews Neuroscience*. 2002. Vol. 3. P. 132–142. URL: http://www.cs.tau.ac.il/~ruppin/evol_age_rev.pdf

Senn W., Schneider M., Ruf B. Activity-dependent development of axonal and dendritic delays or, why synaptic transmission should be unreliable // *Neural Comput.* 2002. Vol. 14, Issue 3, 583–619. URL: http://www.cns.unibe.ch/publications/ftp/paper_DelayNECO.pdf

Sommer F.T., Wennekers T. Associative memory in networks of spiking neurons // *Neural Networks*. 2001. Vol. 14. P. 825–834.

URL: <https://redwood.berkeley.edu/fsommer/papers/sommerwennekers01.pdf>

Sompolinsky H., Kanter I. Temporal association in asymmetric neural networks // *Phys. Rev. Lett.* 1986. Vol. 57. P. 2861–2864.

Stickgold R. Sleep: off-line memory reprocessing // *Trends in Cognitive Sciences*. 1999. Vol. 2, Issue 12. P. 484–492.

Suemitsu A., Morita M. A model of memory formation in the pair-association task // *Proceedings of the 2000 International Conference on Neural Information Processing, Taejeon*. 2000. Vol. 2. P. 915–919. URL: <http://volga.esys.tsukuba.ac.jp/~mor/paper/ICONIP2000.pdf>

Sutherland G., McNaughton B. Memory trace reactivation in hippocampal and neocortical neuronal ensembles. *Current Opinions in Neurobiology*. 2000. Vol. 10. P. 180–186.

Thorpe S.T., Imbert M. Biological constraints on connectionist modelling // *Connectionism in Perspective* / eds. R.Pfeifer, Z.Schreter, F.Fogelman-Soulie, and L.Steels. North-Holland: Elsevier, 1989.

Williams R.J., Zipser D. A learning algorithm for continually running fully recurrent neural networks. *Neural Comput.* 1989. Vol. 1. P. 270–280.

URL: <http://citeseer.ist.psu.edu/williams89learning.html>

Yoshizawa S., Morita M., Amari S. Capacity of associative memory using a nonmonotonic neuron model. *Neural Networks*. 1993. Vol. 6. P. 167–176.

URL: <http://volga.esys.tsukuba.ac.jp/~mor/paper/yoshi1993.html>

Дата публикации 19 апреля 2009 г.

СВЕДЕНИЯ ОБ АВТОРАХ

* Терехин Анатолий Тимофеевич. Доктор биологических наук, профессор кафедры общей экологии, биологический факультет, Московский государственный университет им. М.В.Ломоносова, Ленинские горы 1, стр. 12, 119991 Москва, Россия.
E-mail: terekhin_a@mail.ru

** Будилова Елена Вениаминовна. Кандидат технических наук, ст. научн. сотр. кафедры общей экологии, биологический факультет, Московский государственный университет им. М.В.Ломоносова, Ленинские горы 1, стр. 12, 119991 Москва, Россия.
E-mail: evbudilova@mail.ru

*** Качалова Лариса Андреевна. Кандидат биологических наук, директор Института когнитивной нейрологии, Современная гуманитарная академия, Нижегородская 32, 109029 Москва, Россия.
E-mail: lefti@muh.ru

**** Карпенко Михаил Петрович. Доктор технических наук, профессор, президент Современной гуманитарной академии: Современная гуманитарная академия, Нижегородская 32, 109029 Москва, Россия.
E-mail: rectorat@muh.ru