

В историческом развитии дан обзор основных моделей искусственных нейронных сетей, каждая из которых отражает определенный прогресс в приближении к принципам биологических вычислений. В обзор включены однослойные и многослойные перцептроны, карты признаков Кохонена, сети Хопфилда, двунаправленная ассоциативная память, машина Больцмана, адаптивная резонансная теория и сети Фукушимы.

1. Введение

Всем известны успехи, которых достигли цифровые компьютеры последовательного действия. Имеется ряд задач, с которыми они справляются гораздо лучше, чем человек. Процедура решения таких задач может быть точно сформулирована в виде последовательности инструкций, определяющих, что нужно делать с входными данными. Но с задачами типа распознавания образов эти компьютеры справляются весьма посредственно, так как такие задачи не могут быть точно сформулированы и относятся к категории плохо формализуемых. Решение их по сути сводится к быстрому выбору одного из множества уже готовых решений, которое является наилучшим при имеющихся входных данных, для чего в памяти должна храниться совокупность готовых решений. Поскольку мозг обладает уникальными способностями в решении плохо формализуемых задач, специалисты по вычислительной технике изучают механизмы его работы в надежде, что выявленные принципы можно будет использовать при разработке компьютеров.

Устройства, предназначенные для решения плохо формализуемых задач путем имитации нейронных структур мозга, называются нейрокомпьютерами. Основой нейронной вычислительной машины является нейронная сеть, представляющая совокупность процессорных элементов, объединенных множеством связей. Мощность такой машины определяется не быстродействием процессорного элемента, а степенью связанности нейронов. Обычно используется послойная компоновка нейронов и вычислительные возможности машины зависят также от числа слоев. Это выделяет нейронные сети в самостоятельный класс многопроцессорных вычислительных устройств.

Искусственные нейронные сети демонстрируют многие свойства, присущие мозгу, но даже самые сложные из них имеют лишь самое поверхностное

Поступила в редакцию 22.11.96.

сходство с биологическим мозгом. Однако они предлагают новый подход к машинным вычислениям, следуя которому можно будет создавать компьютеры новой архитектуры. Основные области применения нейронных сетей включают: обработку сигналов; неадаптивное и адаптивное распознавание образов и адаптивное управление; адаптивную ассоциативную память для экспертных систем, баз данных и знаний; решение математических задач большой размерности.

Главными элементами нейросетевой парадигмы являются:

- глобальная параллельность вычислений;
- замена программирования обучением;
- коллективный характер поиска решения за счет соединения процессорных элементов множеством связей;
- ассоциативный характер и распределенность памяти между связями, их гибкость (адаптивность);
- наличие конкуренции как на уровне отдельных нейронов, так и образов памяти;
- способность к абстрагированию.

В этом перечне оптик легко увидит близкие ему принципы, и в последние годы приходит осознание соответствия методов оптической обработки информации нейросетевой парадигме. Многие исследователи, в том числе и занимающиеся электроникой, считают, что, несмотря на первоначальные успехи электронных реализаций, будущее за оптическими и оптико-электронными нейронными сетями. Поэтому так важно, чтобы все большее число специалистов по оптике вовлекалось в это направление и представляло историю его становления. Этому служит данная статья, в которой на примере нескольких основных моделей искусственных нейронных сетей, ставших классическими, показано развитие данного направления. Не умаляя заслуг других исследователей, можно сказать, что основу направления заложил ряд блестящих ученых, которые, несмотря на сложность задачи, превратности

финансирования и научной моды, продолжали работу и создали модели искусственных сетей, каждая из которых отражает определенные достижения в соответствии с биологическим прототипом. В совокупности они отражают поступательное развитие направления и образуют некий его скелет в смысле топологического выражения принципов, лежащих в основе работы мозга. Топология положена в основу классификации, и единственным исключением являются сети обратного распространения сигнала ошибки, которые представляют одну из многих разновидностей многослойного персептрона, но рассматриваются отдельно ввиду чрезвычайной популярности. Вопросы реализации в данной статье не рассматриваются из-за ограниченности объема.

2. Персептрон

Основы нейросетевого подхода к вычислениям были заложены в 1943 году Уорреном Маккалоком и Уолтером Питтсом, изучавшими совокупность формальных нейронов, обменивающихся информацией посредством направленных связей [1]. Формальный нейрон Маккалока-Питтса состоит из набора межсоединений, сумматора и нелинейного оператора (рис. 1). Сумматор объединяет выходные сигналы предыдущего слоя нейронов или сигналы входов (сенсоров) x_i . T_i – веса межсоединений, хранящие результат обучения, θ – порог нейрона, NL – передаточная функция нейрона, которая обычно одинакова для всех нейронов одного слоя. Нейрон имеет два состояния – возбужденное и невозбужденное, он осуществляет пространственно-временное суммирование сигналов от других нейронов, подобно суммированию потенциалов на мембране клетки. Когда потенциал превышает пороговое значение, нейрон возбуждается. Эволюция состояния нейрона определяется изменением сигналов от остальных нейронов и весами межсоединений. Сигнал на выходе j -го нейрона определяется как

$$y_j = NL \left[\sum_i T_{ij} x_i - \theta_j \right], \quad (1)$$

где x – входные сигналы, T – веса межсоединений, θ – пороги нейронов, NL – передаточная (актива-

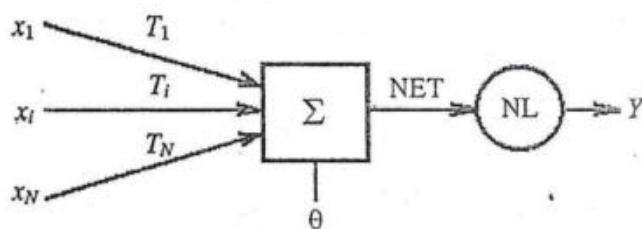


Рис. 1. Формальный нейрон Маккалока и Питтса.

ционная) функция нейрона, которая обычно одинакова для всех нейронов одного слоя. Видно, что это есть корреляция входа с весами межсоединений. В матричной форме выражение (1) может быть записано как

$$Y = NL(NET) = NL(TX), \quad (2)$$

где $NET = TX$ – операция внешнего произведения входного вектора (матрицы) X и матрицы межсоединений T , определяющая результат суммирования входов, умноженных на соответствующие веса связей. Этот алгоритм, называемый алгоритмом внешнего произведения, используется в большинстве предложенных моделей. Авторы показали, что объединение таких нейронов в связные сети ведет к увеличению вычислительной мощности, а отдельный нейрон при соответствующей настройке весов способен осуществлять логические функции Булевой алгебры (И, ИЛИ, И-НЕ). Эта простая схема с незначительными модификациями положена в основу практически всех моделей технических нейронов.

Следующим существенным вкладом в нейрониформатику была теория нейронных ансамблей психолога Дональда Хебба из университета McGill [2]. Он показал, что сила связей между нейронами может изменяться в процессе функционирования, повышаясь для одновременно возбужденных нейронов и при более частой их активизации. Это приводит к объединению нейронов в клеточные ансамбли и обособлению ансамблей друг от друга. При возбуждении достаточной части ансамбля он возбуждается целиком. Один и тот же нейрон может входить в разные ансамбли. Модель памяти, основанная на теории нейронных ансамблей, позволила объяснить некоторые свойства биологической памяти, такие как ассоциативность, распределенность, параллельность, устойчивость к помехам. Постулат Хебба об увеличении синаптической связи между двумя одновременно активными нейронами является основой многих методов обучения нейронных сетей.

Пятидесятые годы – это годы появления кибернетики и создания кибернетических моделей животных. Клод Шеннон создал электронную мышь, которая могла ориентироваться в лабиринте после его обследования. В это же время фон Нейман, считающийся создателем компьютеров последовательного действия, проявляет большой интерес к моделированию работы мозга [3]. В конце 50-х – начале 60-х годов Фрэнк Розенблатт, основываясь на идеях Маккалока и Питтса, предложил модель обучаемой машины, названную им "персептрон", позволяющую решать некоторые интеллектуальные задачи типа распознавания образов и вызвавшую большой интерес у инженеров [4, 5]. Это первое техническое устройство с изменяемой структурой.

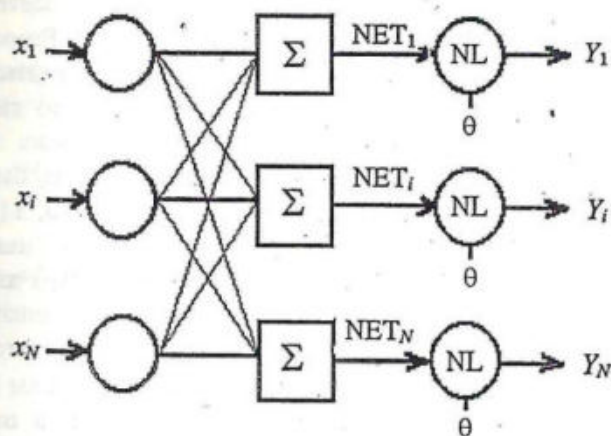


Рис. 2. Персептрон Розенблатта.

На рис. 2 представлен граф такой машины, приведенный к современным представлениям. Она характеризуется наличием только прямых связей и одного слоя межсоединений. Сенсорные нейроны обозначены кружками и не осуществляют никаких преобразований, а лишь передают входные сигналы на слой связей. Суммирующие нейроны обозначены прямоугольниками и образуют слой вычислительных нейронов. Между ними расположен слой межсоединений, которые подлежат настройке на образы памяти. Сенсорные ячейки Розенблатта состояли из фотоэлементов с двумя выходами с положительным и отрицательным сигналами, которые случайным образом соединялись с ассоциативными (суммирующими) ячейками. Переменные X могут быть дискретными и непрерывными. Если алгебраическая сумма сигналов на суммирующей ячейке превышала пороговую величину, то она срабатывала. Если срабатывала не та ячейка (не тот класс), то оператор менял параметры соединений, добываясь правильной реакции. После нескольких этапов обучения персептрон принимал решения самостоятельно.

Персептрон является простейшей однослойной сетью и использует обучение с учителем. Он открывает класс сетей без обратных связей и латерального торможения, называемых сетями прямого распространения (FEEDFORWARD). В соответствии с (1) персептрон разделяет пространство, $\{X\}$ на классы (не более $2M$, M — число классов) гиперплоскостями, определяемыми уравнениями

$$\sum T_i x_i - \theta = 0, \quad i = 1, \dots, M. \quad (3)$$

Розенблаттом был предложен простейший алгоритм обучения для случая $M = 1$:

$$T_i(t+1) = T_i(t) + \alpha [y_0(t) - y(t)] x_i(t), \quad (4)$$

где $0 < \alpha < 1$ — скорость обучения, $y(t)$ — состояние выхода при входе $x(t)$, $y_0(t)$ — состояние желаемого

выхода. Если входные векторы разделяются гиперплоскостью, то алгоритм сходится. Этот алгоритм иногда называют дельта-правилом, вводя обозначение ошибки $\delta = y_0 - y$, равное разности между требуемым и действительным выходами.

Один из ранних алгоритмов обучения персептронов — алгоритм Уидроу–Хоффа (правило наименьшего среднего), использованный в однослойных сетях ADALANE (Adaptive Linear Neuron) [6] и более поздней многослойной модификации MADALANE [7], разработанных в 60-х годах Бернардом Уидроу и его студентом Марцианом Хоффом. Они предназначались для адаптивной отстройки от радарных помех и подавления шумов в каналах связи и имели коммерческое применение более 20 лет. При вычислении ошибки использовался не выход нейрона после порогового оператора, а взвешенная сумма входных связей. В дальнейшем такой же подход был использован в методе обратного распространения сигнала ошибки.

Внутренние ограничения персептронов подробно проанализированы в работе Минского и Пейперта [8], которые показали, что однослойные персептроны Розенблатта неспособны решить многие простые задачи, например, реализовать логическую функцию *Исключающего ИЛИ*. Проблему *Исключающего ИЛИ* легко понять на примере функции двух аргументов X и Y , каждый из которых может быть нулем или единицей. Функция равна единице, когда один (но не оба!) из аргументов равен единице. Сеть описывается уравнением $T_1 X + T_2 Y = \theta$. По одну сторону прямой будут лежать все значения аргументов в плоскости XY , дающие превышение порога, по другую — меньше порога. Никакая комбинация весов не может обеспечить требуемого соотношения между входом и выходом. Существует обширный класс подобных функций, которые называют линейно неразделимыми. Еще хуже то, что при большом числе аргументов не существует простого способа определить, является ли функция линейно разделимой. Помимо проблемы разделимости в персептронах, как впрочем и в других типах нейронных сетей, существует проблема памяти. Минский привел примеры задач, в которых требуемая емкость памяти растет с размерностью задачи быстрее, чем по экспоненте. Если они не являются исключением, то персептроны ограничены только задачами малой размерности. Этот вопрос остается открытым для всех нейронных сетей.

Авторитет Минского, строгость его аргументации и возникший пессимизм относительно возможного прогресса в этой области привели к задержке исследований на два десятилетия. Однако ряд наиболее настойчивых ученых продолжали исследования. В работах Андерсона, Гроссберга, Кохонена,

Хопфилда, Фукушими разработаны более сложные и гибкие архитектуры и изучены их функциональные возможности. Особо следует отметить Стефена Гроссберга, директора Центра адаптивных систем при Бостонском университете, который с 1967 по 1988 год опубликовал с коллегами 146 математических и биологических работ. Постепенно складывался теоретический фундамент направления и было показано, что оценки Минского являлись излишне пессимистичными. В восьмидесятых годах рост активности в этом направлении носит взрывной характер. Появление многослойных перцептронов позволило расширить класс решаемых задач. Например, двухслойные перцептроны работают с выпуклыми множествами и легко решают функцию *Исключающего ИЛИ*. Она реализуется в двухслойной сети по схеме: $(X_1 \text{ И-НЕ } X_2)$ ИЛИ $(X_2 \text{ И-НЕ } X_1)$. Трехслойные перцептроны позволяют формировать структуру функционального пространства любой степени сложности, ограниченной лишь числом нейронов в сети. Обучение многослойных сетей оказывается чрезвычайно трудоемким. Однако такие процедуры были найдены эвристическим путем, одна из них – Back-Propagation.

Перцептрон был задуман как кибернетическая модель системы восприятия живого существа. Для зрительной системы датчики моделировали фоторецепторы сетчатки глаза (их использовалось несколько тысяч), суммирующие ячейки служили моделями ганглиозных клеток (также нескольких тысяч), а блок решения моделировал работу зрительного отдела коры головного мозга. Такая модель лишь отдаленно соответствует процессам, происходящим в зрительном канале. Система, обученная распознаванию некоторых образов, распознает только их. Модель тем не менее достаточно удобна и ее можно приспособить для моделирования нелинейных систем. Она допускает также обобщение для распознавания абстрактных понятий. Было предложено много физических реализаций перцептронов (электронных, микроэлектронных, оптических, электронно-оптических). Хотя перцептроны – в значительной степени уже история, теория перцептронов является основой для изучения других нейронных сетей и используются в них как составная часть.

3. Сети с обратным распространением сигнала ошибки

Сети этого типа используют обучение с учителем и представляют разновидность многослойного перцептрона, т.е. не имеют обратных и латеральных связей. Наименование сетей связано с поиском различных эвристик, позволяющих за ограниченное время найти приближенное решение задачи обуче-

ния. Наиболее известной из них является метод обратного распространения ошибки – Back-Propagation Error (BPE). Это один из лучших алгоритмов обучения многослойных перцептронов, хотя до сих пор нет гарантии, что сеть может быть обучена за конечное время и наилучшим образом. Метод был предложен в работе [9] и развит в работах [10, 11]. Алгоритм представляет обобщение метода наименьших квадратов или правила Уидроу–Хоффа для обучения простого перцептрона применительно к многослойным перцептронам. Минимизируется среднеквадратичная ошибка между фактическим и желаемым выходами. Начальные веса связей и пороги принимаются случайными. Затем предъявляется обучающий образ и вычисляется разность между желаемым и полученным выходом сети (ошибка). Эта разность подается на слой межсоединений (отсюда метод и получил свое название), начиная с последнего слоя, который подстраивается так, чтобы уменьшить ошибку. Процесс многократно повторяется, пока ошибка не стабилизируется на заданной величине. После обучения межсоединений выходного слоя аналогично настраиваются предыдущие (скрытые) слои.

Базовой идеей обучения является обобщенное дельта-правило

$$\Delta T_{ij}^s = \alpha (y_i^s - y_{i0}^s) x_j^s = \alpha \delta_i^s x_j^s, \quad (5)$$

где y_i и y_{i0} – фактический и желаемый выходы i -го нейрона, произведенные j -м компонентом s -го обучающего образа. Поведение сети при этом описывается с помощью функционала энергии среднего квадрата ошибки (функция цены)

$$E = 1/2 \sum_s \sum_i (Y_i^s - Y_{i0}^s)^2, \quad (6)$$

где суммирование осуществляется по нейронам выходного слоя и образам памяти. Для минимизации энергии необходимо изменить весовые коэффициенты по всем i . Авторы [11] дали три формулы для процедуры обучения: правило подстройки весов (5), формулу для различия желаемого и действительного выхода (сигнал ошибки)

$$\delta_i^s = (y_i^s - y_{i0}^s) f_i'(y_i^s), \quad (7)$$

где $f_i'(y_i^s)$ – производная активационной функции при уровне активации y_i^s , и формулу для сигнальных ошибок скрытого слоя

$$\delta_i^s = f_i'(y_i^s) \sum_m \delta_m^s T_{mi}. \quad (8)$$

Этот алгоритм требует непрерывной дифференцируемости активационной функции нейронов, так как требуется вычислять производную, и нулевого ее наклона при малых и сильных сигналах. Схема

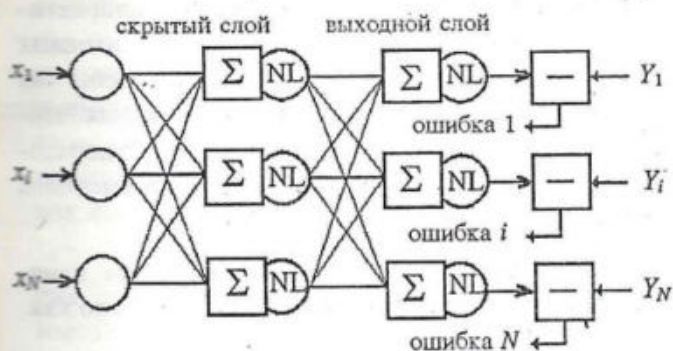


Рис. 3. Двухслойная сеть ВРЕ.

двухслойной нейронной сети с обратным распространением сигнала ошибки приведена на рис. 3, на котором Y обозначает желаемый выход сети.

Один из основных недостатков процедуры ВРЕ состоит в необходимости многократного (сотни раз) предъявления обучающих образов и процесс обучения может занимать несколько часов. Есть две опасности при обучении: возможность попадания в так называемый "паралич", когда сеть перестает обучаться и процесс случайно блуждает, и попадание в локальный минимум энергетического рельефа. Паралич сети происходит при замирании процесса обучения. Так как вычисляемая ошибка пропорциональна производной, обучение замирает, если большинство нейронов начинают функционировать в области малой производной активационной функции. Для выхода из паралича используются различные эвристики. Поскольку при обучении используется метод градиентного спуска, а поверхность ошибки имеет сложную форму, то сеть может попасть в локальный минимум и неправильно обучиться. Помочь этому могут статистические методы обучения (машини Коши и Больцмана), которые являются еще более медленными. Возможно сочетание статистических методов с градиентным спуском. Все это – вопросы современных исследований.

4. Карты признаков Кохонена

В 1982 году Тейво Кохонен из Технологического университета в Хельсинки предложил нейронную сеть с латеральными связями внутри слоя нейронов. Алгоритм Кохонена является простейшим среди самоорганизующихся, он использует конкуренцию нейронов за счет латеральных связей (механизм латерального торможения) и относится к моделям ассоциативной памяти, которые используют некоторое внутреннее представление сенсорной информации, происходящее в мозге. Один из вариантов нелинейного преобразования входной информации в двумерную "карту" признаков и предложил Кохонен [12, 13].

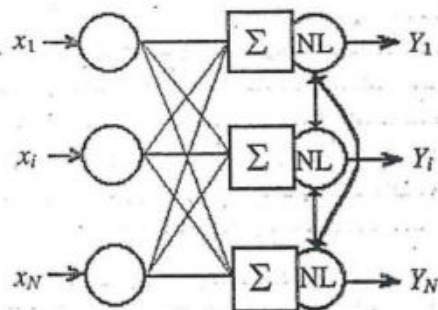


Рис. 4. Сеть Кохонена.

Модель была разработана для топографического отображения как модель символического представления информации с сохранением топологической структуры. Под топографическим отображением в нервных структурах мозга подразумевается упорядоченная проекция предшествующей структуры слоев нейронов на последующую. В данном случае речь шла об отображении сетчатки глаза на зрительную кору мозга. Было уже известно, что это отображение гораздо сложнее, чем простое взаимно-однозначное отображение, так как каждая клетка коры получает сигналы от тысяч фоторецепторов. Карты Кохонена это один из простейших вариантов группировки клеток зрительной коры и формирования карт коры.

Сеть Кохонена является однослойной (рис. 4). Входные сигналы могут быть бинарными или аналоговыми. Нейроны выходного слоя имеют сигмоидальную характеристику и соединены между собой. Используются два вида связей – от сенсорных входов и латеральные, которые полагаются постоянными, – и все адаптивные процессы происходят во входных связях. Первоначально веса связей между слоями нейронов предполагаются случайными. Обучение осуществляется регулировкой связей при предъявлении на вход образа X , однако модель специально не описывает выходной результат. Те нейроны, веса которых более похожи на входной образ, возбуждаются сильнее. Если возникает первоначальная активность нейрона, то механизм латеральных связей приводит к увеличению контраста и возникновению зоны активности (процесс самоорганизации). Близко расположенные нейроны возбуждают друг друга дополнительно, с увеличением расстояния возбуждение сменяется торможением. Таким образом, только наиболее активные нейроны останутся возбужденными. Область активности первоначально может быть большой, но постепенно стягивается до размеров одного нейрона. Пространство памяти и выходное пространство эквивалентны в том смысле, что каждому нейрону на выходе соответствует свой образ памяти. В конце самообучения память подобна входному образу. Затем сети

предъявляется другой обучающий образ и процесс повторяется. Поскольку для различных обучающих образов зоны активности образуются в разных местах, то различные части сети настраиваются на разные образы. Близкие нейроны будут настроены на сходные входные образы. Формируется упорядоченная карта сигнального пространства.

Выходной вектор Кохонен определяет как нормализованную или ненормализованную корреляцию

$$y_i(t) = \sum_j T_{ij}(t)x_j(t) \quad (9)$$

или евклидово расстояние

$$y_i(t) = \left[\sum_j (x_j(t) - T_{ij}(t))^2 \right]^{1/2}, \quad (10)$$

где $x_j(t)$ — j -й компонент входного вектора. Оптимальные связи определяются как минимум расстояния в соответствующей метрике:

$$d_i(t) = \min_j \|X(t) - T_j(t)\|, \quad (11)$$

где C — точка в выходном пространстве, которой соответствуют наилучшие связи. Сначала устанавливаются наилучшие связи между входом и памятью, а затем определяется наилучшая выходная точка. Область связей в пространстве памяти первоначально может быть большой, но постепенно стягивается до размеров одной матрицы. В выходном пространстве первоначальная зона активности равна расстоянию между входным вектором и вектором памяти и постепенно стягивается до одного нейрона. Входные векторы могут быть бинарными или аналоговыми, тогда как образование векторов памяти идет как аналоговый процесс.

Закон адаптации весов входных связей имеет следующий вид:

$$dT_{ij}/dt = \alpha x_j y_i - \beta y_i T_{ij} \quad (12)$$

От правила Хебба это отличается вторым членом, который обуславливает активное забывание. Кохонен показал, что наличие этого члена ведет к нормализации длин векторов связей. Это позволяет определить центр зоны активности как нейрон с минимальной нормой векторной разности. Первоначально $\beta = 1$ и уменьшается в процессе обучения до нуля. Если латеральное взаимодействие принять бинарным и $\alpha = \beta$, то $dT_{ij}/dt = (x_i - T_{ij})$ внутри зон и $dT_{ij}/dt = 0$ вне зоны. Отсюда видно, что в процессе обучения веса стремятся следовать за входными сигналами. Поскольку для различных сигналов зоны активности образуются в разных местах, то различные части сети настраиваются на разные образы и

после длительного обучения происходит аппроксимация функции плотности вероятности входных образов. Близкие нейроны будут настроены на сходные входные образы. Формируется упорядоченная карта сигнального пространства. Эта упорядоченность была доказана только для одномерного отображения [13].

В разностной форме образование векторов памяти, соответствующее (12), может быть записано как

$$T_i(t) + \alpha(t)[X(t) - T_i(t)] - \text{при } i \text{ внутри } N_c(t), \\ T_i(t+1) = T_i(t) - \text{в остальных случаях.} \quad (13)$$

Здесь N_c — область победивших нейронов, а скорость обучения ($0 < \alpha < 1$) уменьшается со временем. Обычно принимают $\alpha(t) = \alpha(0) - \eta t$, где η — коэффициент обучения, t — временной индекс в дискретном пространстве итераций. Процедура повторяется для каждого вектора из обучающей выборки.

Расстояние d_i может быть записано как

$$d_i = \sum_j |x_j(t) - T_{ij}(t)|^2 = \\ = \sum_j x_j^2(t) + \sum_j T_{ij}^2(t) - 2 \sum_j T_{ij}(t)x_j(t), \quad (14)$$

где первый член постоянен. Если веса векторов нормализованы и их автокорреляции идентичны (т.е. сумма квадратов весов от всех входов к каждой выходной точке), то второй член также постоянен. Тогда минимум евклидова расстояния эквивалентен максимуму третьего члена, который отражает выход нейронной сети. Максимум на выходе ищется за счет механизма латерального торможения с использованием поперечных связей или так называемого MAXNET-алгоритма

$$y_i(t+1) = NL[y_i(t) - \varepsilon \sum_m y_m(t)], \quad (15)$$

где NL — нелинейный оператор, ε — константа и $m \neq i$. Для простоты NL можно считать пороговой функцией

$$NL(x) = \begin{cases} 1, & x > a, \\ x, & b < x < a, \\ -1, & x < a. \end{cases} \quad (16)$$

где a и b — произвольные постоянные. В итоге все точки выходной плоскости за исключением одной стремятся к нулю.

Сеть Кохонена открывает класс сетей с латеральными связями и конкуренцией нейронов, т.е. простейшим видом самоорганизации. Сеть также широко используется как составная часть более

сложных моделей. Предложено несколько вариантов применения таких сетей. В работе [14] было предложено использовать сеть для распознавания изображений и звуков (формирование двумерных карт – фонем). Сеть использовалась также для распознавания речи, музыкальных композиций и решения проблемы странствующего коммивояжера.

Существуют электронные и оптические реализации сети Кохонена. По сравнению с сетями ВРЕ и машиной Больцмана сеть Кохонена более проста в реализации, особенно при некоторых упрощениях, например при использовании прямоугольных окон, которые проще, чем евклидовы расстояния. Можно фиксировать величину подстройки весов. Однако такие упрощения могут влиять на сходимость. Большим преимуществом является нечувствительность к точности весов памяти, что важно для аналоговой техники.

5. Сети Хопфилда

Хотя еще Минский и Пейперт отмечали широкие алгоритмические возможности и возможность достижения устойчивости в сетях с обратными связями, отсутствие теоретической базы долгое время тормозило их развитие. Джону Хопфилду из Калифорнийского Технологического института удалось вовлечь в нейронистику мощный математический аппарат статистической физики и предложить физическую реализацию этого класса сетей. Именно с его работ началось практическое использование нейронных сетей с обратными связями. Создание модели связано с некоторыми аналогиями с физическими системами, в частности со спиновым стеклом и нелинейными динамическими системами с подходящей структурой фазового пространства, которые могут рассматриваться как запись информации в памяти.

Сети Хопфилда – это однослойные нейронные сети с обратными связями от выходных нейронов к

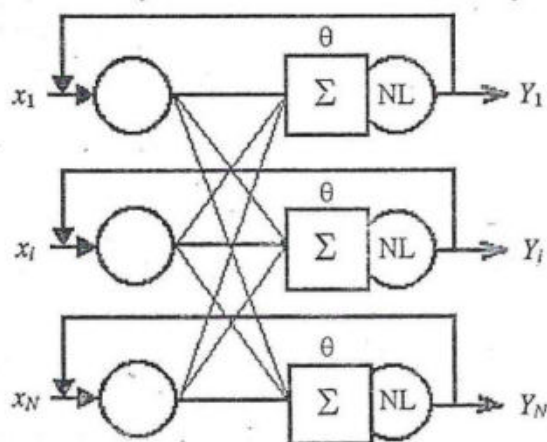


Рис. 5. Сеть Хопфилда.

сенсорным (рис. 5), использующие обучение с учителем. Цель алгоритма – хранение набора образов и вызов наиболее близко согласующегося с входными данными, т.е. сеть может рассматриваться как ассоциативная память. Результирующий выходной сигнал претерпевает пороговую операцию и используется как входные данные для следующей итерации. Процесс повторяется до сходимости.

Первая сеть Хопфилда [15] функционировала в асинхронном режиме. Нейроны – биполярные, связи – симметричные ($T_{ij} = T_{ji}$) и собственные связи отсутствуют ($T_{ii} = 0$). Каждый нейрон определял свое состояние в случайный момент времени с заданной средней частотой так, что в следующий момент времени случайно выбранный нейрон с номером i принимал значение 0 или 1 в зависимости от того, больше или меньше порога величина $\sum T_{ij} V_j$. Матрица памяти образуется взаимными внешними произведениями образов с последующим суммированием. Вызов осуществляется матрично-векторной мультипликацией матрицы памяти на входные данные. Если число запоминаемых образов не превышает емкости памяти M , то устойчивые состояния системы описываются M -мерным вектором, которому соответствует одна из вершин единичного гиперкуба в фазовом пространстве. Предполагается, что для всех нейронов порог $\theta = 0$ или $1/2 \sum T_{ij}$. Последний случай соответствует переходу к биполярным нейронам, описываемым спиновыми переменными $V = 2X - 1$.

Для задач ассоциативной памяти настройка весов может быть осуществлена по правилу Хебба

$$\Delta T_{ij}^s = 1/N (V_i^s V_j^s), \quad (17)$$

где $V^s (s = 1, \dots, M)$ – образы памяти. Тогда матрица памяти имеет "хеббовский" вид (разновидность первоначального варианта):

$$T_{ij} = \sum_s^M (2V_i^s - 1)(2V_j^s - 1), \quad i \neq j, \quad T_{ii} = 0. \quad (18)$$

Выражение (18) является матрицей кросс-корреляций, усредненной по всем векторам памяти. Это усреднение по мере роста M ведет к потере индивидуальности отдельных образов и переполнению сети.

Сила, действующая на i -й нейрон (аналог мембранного потенциала),

$$f_i = \sum_j^M T_{ij} V_j^k = \sum_s^M (2V_i^s - 1) \sum_j^M V_j^k (2V_j^s - 1) \sim \sim N(2V_i^k - 1) / 2, \quad (19)$$

так как для случайно выбранных векторов V среднее значение члена в скобках равно нулю кроме $s = k$.

Если $V_i^k = 1$, f – положительна, если $V_i^k = 0$, f – отрицательна. Поэтому при $M < N$ состояние образа V^k устойчиво (N – размерность входного вектора). При этом, однако, могут появиться ложные образы. Такое правило обучения обеспечивает формирование симметричной матрицы связей и допускает существование тормозящих модифицируемых связей и переход возбуждающих связей в тормозящие, что дает возможность сети автоматически регулировать уровень активности и работать с нулевыми порогами. Однако при этом значительно снижается емкость памяти (до 0,14 от числа нейронов) и возникают ложные минимумы. Емкость памяти сети Хопфилда незначительна, при $N = 100$ сеть надежно запоминает не более 5 векторов (расстояние Хемминга не более 30) [15]. Повышение емкости памяти может быть достигнуто введением дополнительной нелинейности. Квадратичная и кубическая нелинейности увеличивают емкость памяти соответственно в 30 и 300 раз.

Если ввести функцию энергии (Ляпунова) $E = -1/2 \sum_i \sum_j T_{ij} V_i V_j$, то ее изменение при изменении состояния i -го нейрона на δV_i

$$\delta E = -\delta V_i \sum_j T_{ij} V_j < 0 \quad (20)$$

и эволюция системы может закончиться только в устойчивой стационарной точке (аттракторе), соответствующей минимуму энергии. Эти минимумы могут соответствовать образам памяти, но могут быть и ложными.

В работе [16] Хопфилд предложил непрерывную модель и реализацию на аналоговых операционных усилителях в качестве суммирующих и нелинейных элементов и резисторов в качестве весов связей. Нейроны имели сигмоидальную характеристику и их состояния менялись одновременно. Поведение сети описывается системой дифференциальных уравнений

$$\tau dx_i(t)/dt = -x_i(t) + \sum_j^N T_{ij} y_j(t) - \theta, \quad (21)$$

$$y_i(t) = f(x_i(t)),$$

где $\tau = RC$ – постоянная времени связей. При моделировании на цифровом компьютере это может быть записано как $(x_i(t) = y_i(t))$

$$y_i(t) = (1 - \Delta t/\tau) y_i(t-1) + \Delta t/\tau \left[\sum_j^N T_{ij} y_j(t-1) - \theta \right]. \quad (22)$$

Это рекурсивное уравнение типично для сетей с обратными связями. Была доказана сходимость к локальным энергетическим минимумам и для такой сети, однако в таком режиме сеть Хопфилда в качестве аттракторов может иметь также предельные циклы.

Сети разрабатывались для проблем ассоциативной памяти. Хопфилд предложил также использовать свои сети для решения оптимизационных задач, например, проблемы странствующего коммивояжера [17]. Поиск минимума энергии занимает всего несколько постоянных времени нейрона и сеть быстро достигает стабильного состояния. В отличие от задачи ассоциативной памяти здесь требуется нахождение глобального минимума, поэтому бинарные нейроны и бинарный характер их взаимодействия здесь не годятся и используется сигмоидальная модель и аналоговое взаимодействие.

Адаптация межсоединений требует дополнительных операционных усилителей для замены переменных резисторов. Тогда даже минимальная оценка требует квадратичного числа усилителей относительно числа нейронов в слое. Такая электронная реализация возможна только при создании специализированных БИС. Наряду с электронными реализациями, алгоритм был реализован на оптическом вектор-матричном усилителе Фархата [18], представляющем первую оптическую нейронную сеть как вектор-матричный процессор. Эта реализация, наряду с реализациями модели двунаправленной ассоциативной памяти, стала началом интенсивного проникновения в нейронную оптику. Несмотря на свои недостатки, модель находит применение для случаев небольшого количества простых образов.

6. Двунаправленная ассоциативная память

Ассоциативная память – одно из наиболее изученных и развитых применений нейронных сетей. Если входной X и выходной Y образы принадлежат единому полю, то память – автоассоциативная, если разным – память гетероассоциативная. При гетероассоциативной памяти входной и выходной образы могут не иметь общих элементов и быть разной размерности. Для того чтобы память работала как гетероассоциативная, необходимо, чтобы выходной образ существовал на другом слое нейронов, т.е. сеть должна быть, по крайней мере, двухслойной. В этом ее основное отличие от сети Хопфилда, в которой выходной образ возникает на том же наборе нейронов, что и входной. И она развивает далее моделирование биологического прототипа, поскольку наша память перемещается по цепочке умственных ассоциаций от одного предмета к другому пока мы не восстановим требуемый образ, который, может быть, и никак не связан с первоначальным образом, кроме как этой цепочкой.

Линейная ассоциативная память описывается выражением $Y = XT$. В матрице T содержатся пары ассоциированных образов. В простейшем случае каждой паре сопоставляется соответствующая мат-

рица, которые затем суммируются $T = X^T Y$, где X^T — транспонированная матрица. Это операция внешнего векторного произведения. Такой способ формирования линейной ассоциативной памяти предложен в 1972 году Тенво Кохопеном [13, 19] и Андерсоном [20]. Процедура вызова имеет вид

$$Y = X_i T = X_i X_i^T Y_i + \sum (X_i X_i^T) Y_j. \quad (23)$$

Это суммирование внутренних произведений. Если входные образы ортонормированы, то $Y = Y_i$ и память совершенна. В противном случае появляются перекрестные искажения, описываемые вторым членом. При большом числе нейронов векторы X почти ортогональны и перекрестные искажения малы. Линейная ассоциативная память Кохопена-Андерсона минимизирует перекрестные искажения в статистическом смысле и в этом смысле является оптимальной.

В дальнейшем модель была развита для восстановления искаженных образов, был введен итеративный процесс и нелинейность. При этом межсоединения стали использоваться в двух направлениях, но с разными весами. Каждый итерационный шаг все более и более приближает искаженный входной образ к истинному, который запомнен в памяти. После нескольких шагов сеть приходит в стабильное состояние и сигналы X и Y не меняются, т.е. система стабилизируется на некоторой ассоциированной паре. На этом свойстве основана возможность системы опознавать образ по его части или зашумленный образ. Схема сети приведена на рис. 6.

В развитие модели внесли вклад многие ученые, в том числе и Стефен Гроссберг. Коско развил идею двунаправленности. Модель Коско [21] относится к сетям с синхронной динамикой и представляет интерес для оптических реализаций. Такие сети получили название двунаправленной ассоциативной памяти (ДАП). Нейроны разделены на слои X и Y , каждый связан с другим своими межсвязями T и T^T .

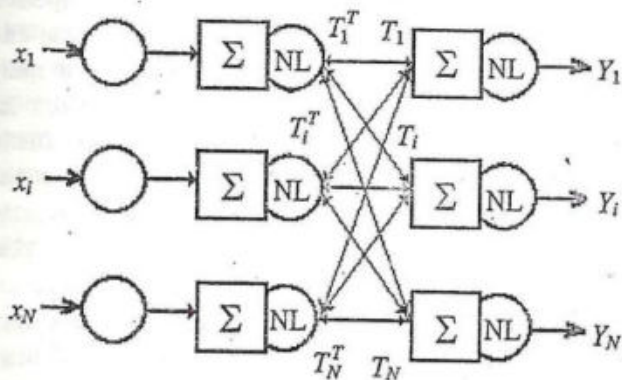


Рис. 6. Сеть двунаправленной ассоциативной памяти.

Эти межсвязи не независимы, а связаны операцией транспонирования. Процесс нелинейной двунаправленной ассоциативной памяти можно представить следующей схемой:

$$X-T-NL-Y-T^T-NL-X-T-NL-Y \text{ и т.д.,}$$

где NL — нелинейное преобразование.

Теорема Коско [22] утверждает, что двунаправленная сеть бинарных нейронов стабильна при произвольной матрице связей, даже если эта матрица не является матрицей памяти. Эта теорема обобщает теорему Хопфилда для случая автоассоциативной памяти [15]. Это важно не только для памяти, но и для задач оптимизации.

Энергия сети для прямого прохождения возбуждения $E(X, Y) = -1/2 XTY^T$ и для обратного прохождения $E(Y, X) = -1/2 YTX^T$ [21]. Система стремится к минимуму энергии при произвольной матрице связей. Если T — матрица ассоциативной памяти, то минимум соответствует ассоциированной паре. Если состояние системы попадает в бассейн притяжения устойчивого состояния, то после нескольких итераций оно оказывается на дне бассейна. Это возможно, если бассейн достаточно велик, т.е. число образов в памяти не слишком велико. Анализ показывает, что число ассоциированных пар не должно превышать минимального из значений размера матрицы образа (столбца или строки).

Возможны различные подходы к обучению сети: с учителем и адаптивное. Правило предписывающего обучения $\delta T_{ij} = \alpha x_i y_j$, где α — константа. Слоям предъявляются входные и выходные образы. По завершении обучения матрица памяти сохраняется неизменной. Адаптивное обучение использует метод проб и ошибок. Сети, уже содержащей случайную матрицу памяти, предъявляется образ, реакция сравнивается с требуемой и вырабатывается ошибка рассогласования, которая используется для корректировки матрицы связей.

Такие сети были созданы в виде электронных реализаций, но обладали ограниченным объемом памяти. Для того чтобы уплотнить память, были созданы оптические реализации, использующие когерентный свет. Долговременная память реализуется в области пространственных частот в виде объемных голограмм, каждая из которых представляет ассоциативную пару образов. Чтобы все голограммы работали одновременно, они регистрируются в области пространственных частот и используется архитектура коррелятора Ван дер Люгта. Итерационный процесс реализуется помещением всей системы в линейный резонатор из двух зеркал обращения волнового фронта. В таком виде ДАП получила наибольшее распространение и получила название

нелинейной голографической ассоциативной памяти (НГАП) [23, 24]. В качестве второго образа ассоциированной пары Y обычно используются точечные источники когерентного света, и устройство работает как автоассоциативная память. Используется ДАП и в сетях более высокого уровня. С этой модели началось использование в нейронинформатике голографии как метода регистрации ассоциированных образов.

7. Машина Больцмана

В модели нейронной сети, известной под названием машины Больцмана, получил наибольшее свое развитие принцип случайности, который впервые был использован Метрополисом для вывода сети из локальных минимумов энергетического рельефа, означающих ложное решение. В дальнейшем стохастический фактор был использован для процесса обучения. В машине Больцмана и обучение и вычисления используют случайность.

Теория машины Больцмана и других сетей стохастического типа, включая стохастическое обучение, была развита в 1983–87 гг. рядом исследователей [25–28]. Машина состоит из случайно соединенных нейронов (рис. 7), при этом любая совокупность нейронов может быть использована для ввода информации и определена как входная, тогда как любая другая совокупность нейронов может использоваться для вывода информации и определена как реакция на вход. В машине используется обучение с учителем, и ее можно рассматривать как стохастический вариант сети Хопфилда, но с двунаправленными связями, пропускающими информацию в обоих направлениях.

Машина Больцмана используется для задач, в которых прямые математические вычисления невозможны или слишком длительны, например оптимизационных, если их можно сформулировать как топологию нейронной сети. Обычные методы в таких случаях не работают удовлетворительно, тогда

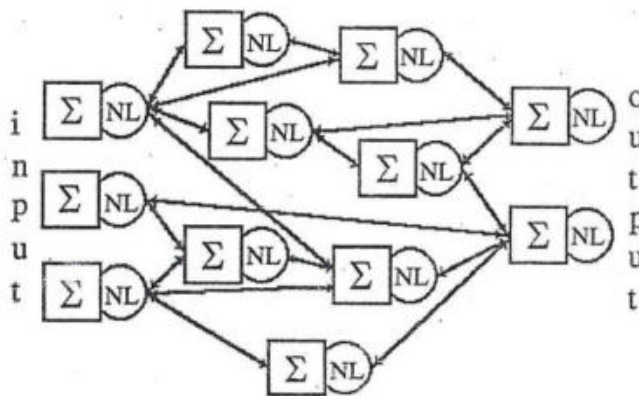


Рис. 7. Машина Больцмана.

вместо точных данных ищется оценка в условиях дефицита времени и информации. Поиск глобальной оценки и есть сфера машины Больцмана. Оценка ищется по минимуму энергии, которая может быть определена по-разному в соответствии с априорной информацией или целями. Однако, каким бы образом ни была сформулирована энергия, машина стремится ее минимизировать путем случайного изменения оценки добавкой, которая принимается или отвергается. Поиск осуществляется по сравнению вероятностей распределения Больцмана. Например, энергию сети можно определить как

$$E = - \sum_{i < j} T_{ij} y_i y_j + \sum_i \theta_i y_i, \quad (24)$$

где y_i определяет состояние или выход бинарного нейрона. Если на нейроны не действует никакая внешняя сила, то вся система движется к минимуму энергии. Чтобы сеть не попала в локальный минимум, нежелательный для оптимизационных задач, применяется вероятностное правило срабатывания элементов

$$P_k = 1 / [1 + \exp(-\Delta E_k / kT)], \quad (25)$$

где P_k – вероятность нахождения k -го элемента в единичном состоянии, T – регулируемый фактор масштаба, называемый "температурой" системы, а k – постоянная, выбираемая в зависимости от задачи. Изменение энергии сети при возбуждении k -го нейрона определяется как

$$\Delta E_k = E_{k \text{ выкл}} - E_{k \text{ вкл}} = \sum_i T_{ik} y_i - \theta_k. \quad (26)$$

Если оно положительно, k -й нейрон переключается на уменьшение энергии. При этом, как и во всех стохастических методах, резко падает вероятность попадания сети в локальный минимум и получения ошибочного решения, что является недостатком многих сетей, использующих поиск минимума энергии методом градиентного спуска.

Процесс достижения равновесия может быть длительным и неуверенным при низкой температуре и быстрым при высокой. Лучше начинать процесс при высоких температурах и получать быструю сходимость, а затем постепенно уменьшать ее и получать лучшее решение. Этот алгоритм получил наименование моделированного отжига [26] по аналогии с процессом отжига кристаллических материалов (медленное охлаждение металла для получения низкоэнергетической кристаллической решетки). В процессе отжига распределение энергетических уровней описывается распределением Больцмана. При высоких температурах вероятность приближается к единице и нахождение системы в высокоэнергетическом или в низкоэнергетическом состоя-

нии практически равновероятны. При уменьшении температуры вероятность высокоэнергетических состояний уменьшается по сравнению с низкоэнергетическими и, таким образом, при приближении температуры к нулю высокоэнергетическое состояние маловероятно. Если возникает тепловое равновесие, то относительные вероятности нахождения системы в двух глобальных состояниях подчиняются распределению Больцмана.

В работе [27] было показано, что скорость уменьшения температуры для попадания в глобальный минимум должна быть обратно пропорциональна логарифму времени:

$$T(t) = T_0 / \log(1 + t), \quad (27)$$

где T_0 — начальная температура, т.е. охлаждение должно быть очень медленным. В [28] распределение Больцмана было заменено на распределение Коши, имеющее более длинные хвосты, что увеличивает вероятность больших шагов,

$$P(x) = T(t) / [T^2(t) + x^2]. \quad (28)$$

Это резко сократило время обучения и искусственная температура как функция времени стала иметь вид

$$T(t) = T_0 / (1 + t). \quad (29)$$

Распределение (28) может быть решено относительно шага x , тогда можно получить аналитическое выражение величины изменения синаптической связи x_c :

$$x_c = \alpha T(t) \operatorname{tg}[P(x)], \quad (30)$$

где α — скорость обучения. Однако и в случае использования распределения Коши время обучения часто оказывается слишком большим (на два порядка больше, чем обучение по методу ВРЕ). Может быть использован метод искусственной теплоемкости, определяемой как средняя скорость изменения температуры с целевой функцией. В критических точках (минимумах), когда небольшое изменение температуры приводит к большим изменениям целевой функции, происходит скачкообразное уменьшение искусственной теплоемкости и скорость уменьшения температуры должна замедляться, чтобы гарантировать сходимость. При всех других температурах можно без риска уменьшать ее быстрее, чем и достигается экономия времени.

Алгоритм может быть выражен в терминах внешних и внутренних произведений. Вычисление внешнего произведения в машине позволяет утверждать, что она является нейронной сетью. Если, например, матрица памяти будет содержать желаемые

векторы в виде сумм внутренних произведений, то машина превращается в нейронную сеть ассоциативной памяти. При этом машина запускается при нулевой температуре и могут быть использованы оптические архитектуры. Поскольку машина ищет решение за счет накопления результатов вычислений, может сказаться падение точности, характерной для оптических систем. Кроме того, некоторые детали алгоритма моделированного отжига слишком тонки для понимания оптической системой, которая также имеет существенно меньший динамический диапазон по сравнению с цифровой.

Алгоритм обучения машины имеет типичные недостатки, присущие методам градиентного спуска при неточном вычислении градиента. Это происходит из-за неполного достижения равновесия и ограниченного времени сбора статистик. Из-за стохастичности обучение занимает гораздо больше времени, чем даже для сетей ВРЕ, поскольку многие шаги выполняются в неверном направлении. Бесконечная дисперсия распределения Коши ведет к появлению неограниченных изменений весов, сильному насыщению нейронов и риску паралича.

Для практической реализации машина имеет несколько особенностей. Во-первых, обучающий алгоритм локален и каждый синапс должен иметь несколько цепей для подбора статистики. Из этого следует низкая точность подбора весов. Однако для ряда простых проблем достаточно 3–7-битовой точности. Во-вторых, алгоритм требует случайных событий, т.е. каждый нейрон нуждается в независимом их источнике. Еще одной проблемой является реализация двунаправленных связей, которые не так просты как однонаправленные цепи. Все это ведет к тому, что всего лишь несколько реализаций машины было создано. Однако созданные статистические методы носят гораздо более общий характер, так как могут быть использованы при решении задач нелинейной оптимизации. В частности, при большом числе переменных и неизвестной их функциональной связи с целевой функцией машина может быть использована для построения модели системы, если могут быть собраны данные для обучающих множеств.

8. Адаптивная резонансная теория

Выше были рассмотрены относительно простые модели нейронных сетей. Существуют гораздо более сложные модели, которые ближе к биологическому прототипу и используют предыдущие как составные части. Наиболее известны из них две: адаптивная резонансная теория — Adaptive Resonance Theory (ART) и сети Фукушимы — Когнитрон и Неокогнитрон.

Сети ART явились результатом попытки решения дилеммы стабильности – пластичности памяти, т.е. попытки ответить на вопрос: как память может воспринимать новую информацию (быть пластичной), не забывая и не искажая при этом уже существующую (оставаясь стабильной). Все рассмотренные выше модели не обладают этим свойством и при предъявлении нового обучающего образа искажают уже запомненные. В некоторых моделях новый образ требует даже полного переобучения сети. Такие модели как ВРЕ требуют многократного предъявления каждого образа, что в реальных ситуациях может быть недопустимым. Все это может быть причиной отсутствия сходимости процесса обучения.

Адаптивность в широком смысле – это изменение выходных сигналов, отражающее изменение входных воздействий. В зависимости от типа обучения она проявляется по-разному. Адаптивная система, использующая обучение с учителем, адаптируется к конкретной ситуации каждый раз одинаково с точностью до шумов. Самообучающаяся система может адаптироваться каждый раз по-иному. В нейронном интеллекте под адаптивностью на уровне долговременной памяти подразумевают способность гибко видоизменять и наращивать память, приспосабливаясь к изменениям внешнего мира и накоплению опыта.

Стефен Гроссберг и его сотрудник Гейл Карпенгер предложили модель нейронной сети с адаптивным резонансом [29–32]. Цикл работ Гроссберга и Карпенгера объединяется авторами как "теория адаптивного резонанса". Концепция была разработана в начале 70-х годов и детализирована в дальнейшем. Важнейшей особенностью нейронной сети с адаптивным резонансом является способность к самообучению. Сеть относит входной образ к одному из классов, на который он более похож. Если входной образ соответствует образу памяти, то он классифицируется, если найденный образ памяти соответствует входному с определенным допуском, то он видоизменяется, чтобы быть более похожим на входной. Если образ не найден, создается новый класс. При этом происходит опробование классов до тех пор, пока не обнаружится соответствующий входному образу или не установится новый. Обучение осуществляется после того, как сеть стабилизировалась на конкретном классе (вошла в резонанс), отсюда метод и получил свое наименование. Обучение в резонансных состояниях либо уточняет ранее сформированный образ памяти на основе новой информации, либо инициирует установление нового образа. Существуют разновидности ART для работы с бинарными и непрерывными образами (ART-1, ART-2).

Рассмотрим основы архитектуры нейронной сети с адаптивным резонансом следуя работе [31] (рис. 8). Прямоугольники представляют поля, в которых содержатся образы кратковременной памяти – Short Term Memory (STM). Черные полуокружности представляют движение сигналов через адаптивные фильтры, а стрелки представляют пути сигналов, которые не являются адаптивными. Черные круги представляют устройства, которые суммируют входные сигналы. Система содержит две подсистемы: внимания (ATTENTIONAL SUBSYSTEM) и ориентации (ORIENTING SUBSYSTEM). Подсистема внимания состоит из двух нейронных полей: поля образов F_1 (слой сравнения) и поля категорий F_2 (слой распознавания), образующих кратковременную память и соединенных матрицей межсоединений (LTM – Long Term Memory). Совокупность межсоединений образует фильтр для сигналов, направленных снизу–вверх, или долговременную память и фильтр сверху–вниз, который обеспечивает пороговый характер самостабилизации системы. Адаптивные сигналы, идущие сверху–вниз играют роль ожидаемых образов, полученных в результате обучения системы и позволяют системе реализовать механизмы ожидания, параллельного поиска и сравнения ожидаемого сигнала с произвольными входными образами.

Вспомогательная ориентирующая подсистема становится активной, если направленный снизу–вверх входной сигнал в поле F_1 не соответствует ожиданиям, считываемым из поля категорий F_2 . В этом случае ориентирующая подсистема активизируется и приводит к быстрой перестройке поля активных категорий F_2 . Такая перестройка автоматически заставляет подсистему внимания продолжать параллельный поиск. При этом происходит

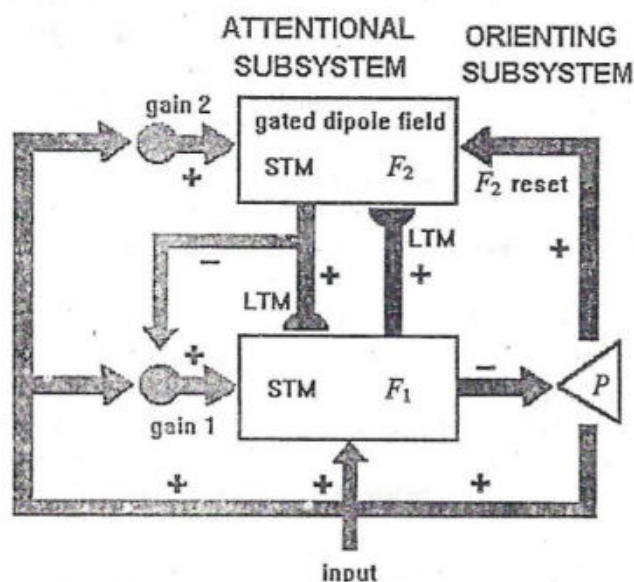


Рис. 8. Типичная архитектура ART [31].

опробование альтернативных категорий до тех пор, пока не обнаружится адекватная входному образу или не установится новая. Стратегия адаптивно об-новляется посредством процесса самообучения. По-иск осуществляется быстро по сравнению со скоро-стью процесса обучения. По терминологии авторов самообучение системы носит характер "конкурент-ного" в сочетании с "ассоциативным" обучением, характерным для многих других моделей нейронных сетей. Обучение в резонансных состояниях либо уточняет ранее полученный код опознавания на ос-нове новой информации, которую могут содержать входные образы, либо инициирует установление но-вого кода в рамках еще незанятых межсоединений.

Рассмотрим работу системы более подробно. Входной образ I после предобработки поступает на поле F_1 , в подсистему ориентации и блок управле-ния усилением. Его наличие вызывает сигнал управления усилением G , подвозбуждающий все нейроны поля F_1 . Для них действует правило 2/3, согласно которому нейрон становится активным только тогда, когда на него действуют сигналы двух источников (из трех). В результате в слое F_1 появ-ится вектор X , совпадающий с I . При прохожде-нии по связям осуществляется трансформация X во входной вектор S поля F_2 :

$$S_j = \sum_j^M T_{ij}(t)x_i, \quad 1 \leq j \leq M. \quad (31)$$

В результате взаимодействия нейронов этого слоя с помощью механизма латерального торможения в нем образуется новый вектор Y с усиленным контрастом. В простейшем случае это может быть один возбужденный нейрон на месте максимального компонента вектора S . Это соответствует формиро-ванию зоны активности в картах признаков Кохо-пена. Возможность активизации нейронных групп в слое F_2 и их функции рассмотрены в работе [32]. Вектор Y представляет гипотезу системы относи-тельно того, на какой из закодированных в LTM классов похож входной образ I . Проходя по свя-зям сверху-вниз он считывает из долговременной памяти эталонный образ V класса, к которому сис-тема отнесла I . Наличие активности в слое F_2 бло-кирует сигнал управления усилением G и согласно правилу 2/3 в F_1 останутся активными только ней-роны, получающие возбуждение от I и V . Если та-ких нейронов мало, активность F_1 значительно тор-мозится, и, если она становится меньше порога внимания подсистемы ориентации, на длительное время тормозятся активные нейроны слоя F_2 и сни-мается эталон V . В слое F_1 восстанавливается пер-воначальный вектор X , но из-за долговременного торможения он уже не сможет активизировать в F_2 прежний вектор V и создаст новый вектор-эталон.

Последовательный поиск в долговременной памяти продолжается до тех пор, пока не наступит одно из двух событий: либо вектор Y в F_2 возбуждет эталон, который достаточно близок к I , и входной вектор будет отнесен к кластеру, соответствующему V , ли-бо в F_2 возбуждается совершенно новый вектор (нейрон), производится модификация весов с целью сформировать в памяти новый образ-эталон I . Под-система ориентации перестает генерировать сигнал сброса, и состояние системы стабилизируется. В этом состоянии происходит самообучение. Если поиск привел к активации имеющегося эталона, то может быть подкорректирован критерий допуска к нему или он сам уточнен, чтобы стать более похо-жим на входной образ. Если активируется новый вектор, то самообучение создает новую категорию. Таким образом, сеть обладает свойством обучения в потоке данных и производит классификацию обра-зов, изменяя при необходимости структуру класте-ров.

Итак, в системе осуществляется характерный цикл обработки сигналов. Этот цикл содержит адаптивную фильтрацию снизу-вверх входных обра-зов, формирование кода (гипотезы), формирование с помощью фильтрации сверху-вниз ожидаемого об-раза, полученного в результате обучения, сравнение с предъявляемыми образами и перестройку кода. Тот факт, что самообучение системы происходит только в состоянии резонанса, позволяет реализо-вать компромисс между ее пластичностью и ста-бильностью. Под пластичностью понимается потен-циальная возможность системы быть готовой к изу-чению новых событий до тех пор, пока не истощит-ся полная емкость памяти.

Долговременная память подчиняется правилу обучения

$$dT_{ij}/dt = ky_j(x_i - E_{ij}T_{ij}), \quad (32)$$

которое включает забывание, но в более сложной форме, чем у Кохонена. Связи между активными в F_2 и неактивными в F_1 нейронами исчезают, что и ведет к их уточнению входными образами. Перед началом обучения связям T_{ij} снизу-вверх присваи-ваются небольшие случайные значения, а связям T_{ji} сверху-вниз – близкие к максимальным. Для связей сверху-вниз $E_{ij} = 1$, а для связей снизу-вверх $E_{ij} = x_i + L \sum_k x_k$ ($L = \text{const}$, $k \neq j$), что обеспечивает вы-полнение правила Вебера. Оно требует, чтобы при обучении связи для вектора с меньшим числом ак-тивных нейронов были больше, чем для вектора с бóльшим числом. Это обеспечивается конкуренцией за синаптические ресурсы, которую описывает E .

Сеть выдвигает встречные ожидания и по сути начинает функционировать, когда в ней уже сфор-мировалось предварительное мнение о классифика-

ции образа взаимодействием методов фильтрации и латерального торможения. За счет нелинейного усиления уменьшается уровень шумов и искажений в совпадающих частях образа и уменьшается порог бдительности к несовпадающим частям (адаптивная регулировка порога узнавания). Значительное несовпадение входного образа и встречного ожидания может быть устранено применением принципа переключения внимания. Наиболее активный нейрон в распознающем слое, направивший процесс по неправильному пути, временно тормозится, и сеть начинает искать вариант более удачной классификации [31].

Архитектура адаптивного резонанса позволяет реализовать самоорганизацию нейронной сети, в результате которой происходит стабильная классификация произвольной последовательности аналоговых или бинарных входных образов. Следует отметить, что в ART адаптивное распознавание образов является лишь частным случаем ее работы. В общем случае система реализует когнитивный процесс, содержащий обнаружение, тестирование, поиск, обучение и опознавание. В настоящее время разработана математическая теория сетей ART, доказывающая их способность к надежной самостабилизации, проведено их машинное моделирование. Показано отсутствие в сетях ART ложных образов, которые могут быть характерны для других классов нейронных сетей. Результаты большого цикла работ указывают на перспективность ART-систем для обработки больших и сложных баз данных. В силу своей сложности метод резонанса пока имеет мало реализаций.

9. Сети Фукушимы (Когнитрон и Неокогнитрон)

В 1975 году Кунихико Фукушима из NHK лаборатории в Токио предложил специализированную сеть для распознавания образов, названную Когнитрон и обладающую инвариантностью к смещению и вращению [33]. Далее эта модель была развита и получила название "Неокогнитрон" [34-39]. Сети Фукушимы позволяют реализовать селективное восприятие при опознавании образов, т.е. расчленение в процессе восприятия сложного объекта на его составные части (сегменты) и отдельное опознавание сегментов.

Когда мы наблюдаем сложную фигуру, состоящую из многих образов, мы обычно обращаем внимание на один из образов, выделяем его среди остальных и опознаем его отдельно. После этого мы переключаем внимание на другой образ. Даже если образ, на который мы обращаем внимание, подвергнут действию шумов или содержит искажения, мы

можем восстановить полный образ, в котором дефекты и шум устранены. Даже если форма образа искажена или изменен его размер, мы можем опознать его и убрать дефекты посредством интерполяции. В процессе интерполяции мы используем эффективно даже малые остатки контуров образа, на которые мы обращаем селективное внимание. Приведенное описание зрительного восприятия моделируется в сетях Фукушимы. Простейшие модели ассоциативной памяти хотя и позволяют восстановить объект по его искаженному варианту, однако хорошо работают только, если предъявляемый образ соответствует по величине, ориентации и, возможно, положению объекта, на примере которого был проведено обучение. Модель Фукушимы позволяет реализовать автоассоциативное опознавание деформированных образов независимо от их положения. Кроме того, модель осуществляет сегментацию описываемых объектов.

Когнитрон

Сеть имеет послойную структуру со связями от предыдущего слоя к последующему и латеральным связями. Нейрон постсинаптического слоя не обладает полной связанностью, а связан только с частью нейронов предсинаптического слоя, называемых областью связей. Это соответствует послойной структуре зрительного канала с ограниченной длиной связей. С другой стороны, в постсинаптическом слое близко расположенные нейроны объединяются латеральными связями в области конкуренции, которых в процессе обучения происходит усиление синапсов только у наиболее возбужденного нейрона, а остальные остаются неизменными. Это соответствует процессу самоорганизации в картах Кохена. Таким образом, в Когнитроне используется самообучение и не требуется предъявления выходных сигналов. Области связей соседних нейронов слоя значительно перекрываются, что компенсируется процессом конкуренции.

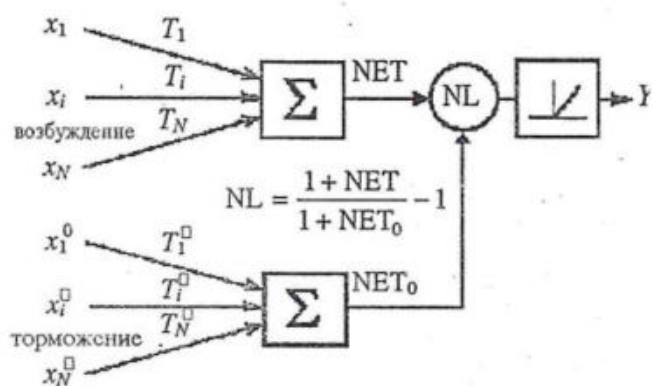


Рис. 9. Возбуждающий нейрон Фукушимы.

В каждом слое существует два типа нейронов — возбуждающие и тормозящие. Модель возбуждающего нейрона приведена на рис. 9. Веса связей имеют только положительные значения и могут возрастать без ограничений. Возбуждающие и тормозящие связи суммируются отдельно и используется специальный тип нелинейного преобразования:

$$Y = NL, \text{ NET} \geq 0, \\ NL = (1 + \text{NET}) / (1 + \text{NET}_0) - 1, \quad (33) \\ Y = 0, \text{ NET} < 0,$$

где NET и NET_0 — взвешенные суммы возбуждающих и тормозящих входов. При $\text{NET}_0 \ll 1$ выход примерно пропорционален разности сумм возбуждающих и тормозящих входов и нейрон работает как пороговый элемент с нулевым порогом. Если же $\text{NET} \gg 1$ и $\text{NET}_0 \gg 1$, выходная функция пропорциональна отношению возбуждающих и тормозящих входов и возрастает по закону Вебера-Фехнера, что и делает эту модель биологически более правдоподобной.

Для каждого постсинаптического нейрона в предсинаптическом слое имеется тормозящий нейрон с той же областью связей с возбуждающими нейронами. Синаптические веса тормозящих нейронов фиксированы, не меняются при обучении и их сумма равна единице ($\sum T_i^0 = 1$). Таким образом, выход тормозящего нейрона NET_0 является среднearифметическим значением выходов возбуждающих нейронов данной области связей.

Латеральное торможение в зоне конкуренции осуществляется тормозящим нейроном (рис. 10), вырабатывающим сигнал торможения на каждый возбужденный нейрон X_i путем нелинейного преобразования $NL = (1 + X_i) / (1 + \text{NET}_0) - 1$. Такой метод латерального торможения делает конкуренцию более правдоподобной, поскольку исключает ситуацию предопределенной победы изначально более возбужденного нейрона, как это имеет место при обычном, латеральном торможении. В процессе самообучения веса связей каждого нейрона постсинаптического слоя настраиваются на определенный образ, который часто появлялся в предсинаптическом слое.

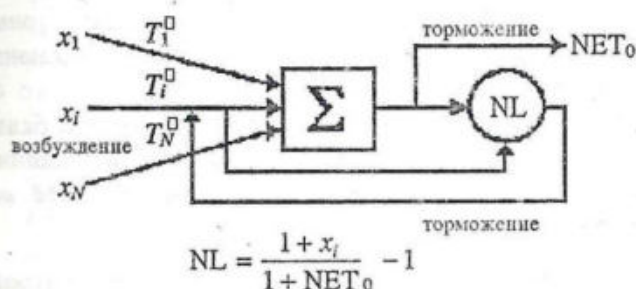


Рис. 10. Тормозящий нейрон Фукушимы.

Когнитрон состоит из нескольких слоев, в которых, подобно зрительной коре, реализуются различные уровни обобщения образа и отклик слоя становится все более сложным и абстрактным. Хотя постсинаптические нейроны напрямую связаны лишь с небольшой частью предсинаптических нейронов, степень косвенной связи с сенсорными нейронами возрастает по мере увеличения номера слоя и нейроны выходного слоя реагируют на все входное поле. При ограниченных размерах области связей может потребоваться большое число слоев. Оно может быть уменьшено увеличением размера области связей, но из-за большого их перекрытия происходит нивелирование реакции выходных нейронов. Частично это может быть скомпенсировано расширением области конкуренции. Таким образом, размеры областей связи и конкуренции, а также число слоев могут быть оптимизированы для данной задачи. Для уменьшения числа слоев может быть использовано вероятностное соединение нейронов с уменьшением вероятности появления более длинных связей.

Таким образом, в Когнитроне получило свое дальнейшее развитие моделирование топографического отображения зрительной коры по сравнению с моделью Кохонена.

Неокогнитрон

Модель Когнитрона в дальнейшем была существенно переработана в свете большего соответствия работе зрительной системы, в которой на низшем уровне были обнаружены нейроны, реагирующие на линии и углы определенной ориентации. На более высоком уровне нейроны реагируют на окружности, треугольники и прямоугольники. Степень абстракции возрастает, пока не определяются нейроны, реагирующие на сложные формы и устойчивые к искажениям. При этом в Неокогнитроне существенно были улучшены возможности модели к распознаванию искаженных образов.

Неокогнитрон также является иерархической многослойной сетью, содержащей каскады слоев простейшей сети. Через эти каскады возможно прямое и обратное распространение сигналов. Пря-

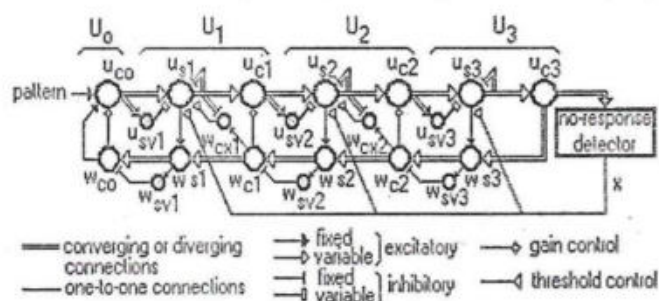


Рис. 11. Неокогнитрон [39].

мые сигналы от предъявляемого образа распространяются внутрь сети и выполняют функции опознавания, в то время как обратные сигналы осуществляют функции селективного восприятия и ассоциативного выбора. Прямые и обратные сигналы взаимодействуют друг с другом на всех стадиях их распространения в иерархической сети.

Структура Неокогнитрона сложна для полного графического отображения. Рис. 11 из работы [39] частично иллюстрирует структуру трехслойного Неокогнитрона. Начальный каскад сети характеризуется состоянием U_0 , l -й каскад – состоянием U_l и высший каскад – состоянием U_L . В частности, на рис. 11 $L = 3$. Используется также обозначение U_{sl} , чтобы определить слой ячеек u_s в l -м каскаде. Кружки представляет нейроны разного типа. Хотя нейроны разного типа присутствуют в большом количестве в каждом каскаде, для каждого каскада показан только один нейрон данного типа. Возбуждающие нейроны обозначены светлыми кружками, тормозящие – темными. В каждом слое разные типы нейронов, такие как u_s – простые нейроны и u_c – сложные нейроны и т. д., собраны в двумерные матрицы (плоскости). Таких плоскостей много в каждом слое. Между нейронами существуют соединения, обозначенные одиночными и двойными линиями. Одиночная линия показывает, что существует соединение "один к одному" между двумя группами нейронов, а двойная линия показывает, что существует сходящаяся либо расходящаяся система связей.

Нижним каскадом для прямого распространения является слой U_{c0} , содержащий рецепторные нейроны. Информация о стимулах, полученных во входном слое, распространяется в прямом направлении, обрабатывается в каждом каскаде и передается на каскады более высокого порядка. В процессе прямого распространения происходит опознавание образа. При этом высший каскад для прямого распространения есть опознающий слой U_{cl} , отклик которого показывает окончательный результат опознавания. В то же самое время отклик опознающего слоя передается в обратном направлении к низлежащим каскадам. Прямые и обратные сигналы взаимодействуют друг с другом на каждом каскаде. При этом обратные сигналы (сигналы селективного восприятия) облегчают обработку прямых сигналов. Верхним каскадом для обратного распространения является слой W_{c0} . В этом слое появляется сигнал, соответствующий ассоциативному восприятию и результату сегментации.

Простые нейроны u_s данной плоскости имеют область связей в каждой плоскости сложных нейронов предыдущего слоя и реагируют на один и тот же образ. Для искаженных версий данного образа

существуют другие плоскости простых нейронов. Области связей в плоскостях сложных нейронов предыдущего слоя значительно перекрываются. Простые нейроны других плоскостей реагируют на другие образы и их искаженные версии. Таким образом, простой нейрон данной плоскости реагирует на определенную версию данного образа, если она появится внутри одной из его областей связей.

Сложные нейроны u_c служат для уменьшения зависимости от положения образа и имеют перекрывающиеся области связей только в своей плоскости простых нейронов. Внутри своей плоскости они связаны областями конкуренции. По мере увеличения номера слоя сложный нейрон реагирует на данную версию образа при любом его положении в поле зрения. Таким образом, по мере роста номера слоя возрастает степень абстрагирования и уменьшается чувствительность к положению и искажениям образа.

Как и в Когнитроне простой нейрон состоит из возбуждающего u_s и тормозящего u_{sv} нейронов с одинаковыми областями связей. Выход тормозящего нейрона равен корню квадратному из взвешенной суммы квадратов входов по всем плоскостям сложных нейронов:

$$NET_0 = \left[\sum (T_i^0 x_i)^2 \right]^{0,5}, \quad (34)$$

причем веса T_i^0 фиксированы и $\sum T_i^0 = 1$. Обучаются только связи простого возбуждающего нейрона, в том числе и выходная связь от тормозящего нейрона. Самообучение и латеральное торможение аналогичны Когнитрону и описаны в работах [34, 35]. Увеличение синаптической связи простого нейрона происходит только при условии возбуждения предсинаптического сложного нейрона и более сильного возбуждения простого нейрона внутри области конкуренции. При этом, однако, происходит и некоторое увеличение синапсов остальных нейронов плоскости, чтобы они учились распознавать одинаковые образы и это распознавание не зависело от позиции образа в предсинаптической плоскости сложных нейронов и от возможности отказа наиболее возбужденного нейрона.

Приведенное рецептивное поле нейронов возрастает по мере увеличения номера слоя, но количество нейронов в слое уменьшается. В идеале только один сложный нейрон выходного слоя, отражающий определенную искаженную версию конкретного образа, должен возбуждаться. Практически это оказывается не лучшим вариантом при классификации и лучше использовать выходы группы наиболее возбужденных нейронов.

В работе [38] описана версия Неокогнитрона, использующая обучение с учителем, в котором реак-

ния каждого слоя заранее задается экспериментатором с учетом возрастания уровня абстрагирования.

Фукушимой проведено детальное теоретическое рассмотрение работы сети и выполнено ее машинное моделирование. Результаты показывают, что при предъявлении сети стимула, состоящего из нескольких образов, реализуется селективное восприятие и независимое опознавание каждого образа. Модель способна опознавать деформированные образы независимо от их пространственного положения и может быть применена при опознавании сложных зашумленных зрительных образов, а также для опознавания речи в присутствии посторонних звуковых шумов. Неокогнитрон требует громадных вычислительных ресурсов и сложен для реализации на современном уровне развития технологий.

Заключение

Человеческий мозг, блуждая в лабиринтах памяти, сформированной в результате многолетнего обучения, и строя цепочки ассоциаций, способен делать самые абстрактные заключения, внешне весьма далекие от той информации, которая воспринимается нашими органами чувств. Современная наука и тем более технология еще очень далеки даже от моделирования этого сложнейшего процесса. Биологические нейронные сети являются сложнейшими электрохимическими системами, далекими от существующих математических и физических моделей. Во многих случаях невозможно выразить происходящие процессы математически. Иногда методы и алгоритмы вообще имеют мало общего с биологической системой. Но и эти простые абстракции необходимы для понимания и физической реализации основных механизмов самоорганизации и обучения в системах массивированной связанности.

ЛИТЕРАТУРА

1. *McCulloch W.S., Pitts W.A.* // Bulletin of Mathematical Biophysics - 1943. - Vol. 5. - P. 115-133. (Перевод: в кн. "Автоматы" / Под ред. Шеннона К.Э., Маккарти Д. - М.: Мир, 1965. - 403 с.)
2. *Hebb D.O.* The Organization of Behaviour. - New York: Wiley, 1949. - 319 p.
3. *Von Neumann J.* The Computer and the Brain. - New Haven: Yale University Press, 1958. - P. 66-82.
4. *Rosenblatt F.* // Psych. Rev. - 1958. - Vol. 65. - P. 386.
5. *Rosenblatt F.* Principles of Neurodynamics. - New York: Spartan Books, 1962. (Перевод: Розенблатт Ф. Принципы нейродинамики. Перцептроны и теория механизмов мозга. - М.: Мир, 1965. - 403 с.)
6. *Widrow B., Hoff M.E.* Adaptive Switching Circuits. - IRE WESCON Conv. Record, 1960. - Part 4. - P. 96.

Описанные модели отличаются топологией. Наибольшей простотой и легкостью реализации обладают сети прямого распространения. Добавление латеральных или обратных связей полностью меняет поведение сети. Кроме того, сети отличаются типами нелинейных активационных функций нейронов, синаптическими операциями и алгоритмами обучения. Из-за этого трудности при реализации различны. Алгоритмы, подобные ВРЕ, требуют большой точности вычислений и непрерывной дифференцируемости активационной функции. Наоборот, в сетях Хемминга или сетях с радиальными базисными функциями (RBF - Radial Basis Function) при обучении достаточно 5-битовой точности. В системах с меньшей степенью параллелизма (сети Фукушимы) требуется большая сложность и гибкость.

Чисто оптические реализации нейронных сетей работают экстремально быстро, поскольку не имеют паразитных емкостных и индуктивных ограничений в цепях. Однако пока они еще очень громоздки, наиболее сложны и не завершены. Оптика имеет много общего с аналоговой электроникой, но технологические проблемы еще более велики, производство дорого, спектр возможных материалов велик. До реализации на уровне чипа дело еще не дошло даже на уровне простых сетей и конкурентноспособного продукта нет. Но в лабораториях очень много занимаются оптикой. Идет комбинирование с электроникой, где оптика осуществляет межсоединения, и такие реализации, возможно, более реальны в ближайшем будущем, чем чисто оптические. Зато оптические сети позволяют сочетать полную связанность с большим числом нейронов в слоях и достичь алгоритмического упрощения.

Работа выполнена при поддержке Российского фонда фундаментальных исследований (проект 95-02-05892-а).

7. *Winter R., Widrow B.* // IEEE Inter. Conf. on Neural Networks. - San-Diego: CA, 1988. - Vol. I. - P. 401-408.
8. *Minsky M.L., Papert S.* Perceptrons. - Cambridge, MA: MIT Press, 1969. (Перевод: Минский М., Пейперт С. Перцептроны. - М.: Мир, 1971. - 264 с.)
9. *Werbos P.J.* Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences // Ph.D. Dissertation in Statistics. - Harvard University, 1974.
10. *Rumelhart D.E., Hinton G.E., Williams R.J.* // Nature. - 1986. - № 323. - P. 533-536.
11. *Rumelhart D.E., McClelland J.L.* Parallel Distributed Processing: Exploration in the Microstructure of Cognition. Vols. 1 and 2. - Cambridge, MA: MIT Press, 1986.

12. *Kohonen T.* // *Biolog. Cybernetics.* - 1982. - Vol. 43, № 1. - P. 59-69.
13. *Kohonen T.* *Self-organisation and Associative Memory.* Berlin: Springer-Verlag, 1984.
14. *Kohonen T.* // *Computer.* - 1988. - № 3. - P. 11-22.
15. *Hopfield J.J.* // *Proc. Natl. Acad. Sci. USA.* - 1982. - Vol. 79, № 8. - P. 2554-2558.
16. *Hopfield J.J.* // *Proc. Natl. Acad. Sci. USA.* - 1984. - Vol. 81. - P. 3088-3092.
17. *Hopfield J.J., Tank D.W.* // *Biolog. Cybernetics.* - 1985. - Vol. 52. - P. 141-152.
18. *Farhat N.H., Psaltis D., Prata A., Paek E.* // *Appl. Opt.* - 1985. - Vol. 24, № 10. - P. 1469-1475.
19. *Kohonen T.* // *IEEE Trans. Comput.* - 1972. - Vol. C-21. - P. 353.
20. *Anderson J.A.* // *Math. Biosci.* - 1972. - Vol. 14. - P. 197.
21. *Kosko B.* // *Appl. Opt.* - 1987. - Vol. 26, № 23. - P. 4947-4960.
22. *Kosko B.* // *Inf. Sci.* - 1986. - Vol. 40. - P. 165.
23. *Psaltis D., Farhat N.* // *Opt. Lett.* - 1985. - Vol. 10. - P. 98.
24. *Owechko Y., Dupning G.J., Maron E., Soffer B.H.* // *Appl. Opt.* - 1987. - Vol. 26, № 10. - P. 1900.
25. *Kirkpatrick S., Gelatt C.D., Vecchi M.P.* // *Science.* - 1983. - Vol. 220. - P. 671.
26. *Hinton G.E., Sejnowski T.J.* // *Proc. IEEE. Computer Vision and Pattern Recognition.* - 1983. - P. 448-453.
27. *Geman S., Geman D.* // *IEEE Transactions on Pattern Analysis and Machine Intelligence.* - 1984. - № 6. - P. 721-741.
28. *Szu H., Hartley R.* // *Phys. Lett.* - 1987. - Vol. 1222 (3, 4). - P. 157-162.
29. *Carpenter G.A., Grossberg S.* // *Computing Vision, Graphics, Image Processing.* - 1987. - Vol. 37. - P. 54-115.
30. *Grossberg S.* // *Cognitive Science.* - 1987. - Vol. 11. - P. 23-63.
31. *Carpenter G.A., Grossberg S.* // *Appl. Opt.* - 1987. - Vol. 26, № 23. - P. 4919-4930.
32. *Grossberg S.* // *Neural Networks.* - 1988. - Vol. 1 № 1. - P. 17-61.
33. *Fukushima K.* // *Biolog. Cybernetics.* - 1975. - Vol. 20. - P. 121-136.
34. *Fukushima K.* // *Biolog. Cybernetics.* - 1980. - Vol. 36 (4). - P. 193-202.
35. *Fukushima K., Miyake S.* // *Pattern Recognition* - 1982. - Vol. 15(6). - P. 455-469.
36. *Fukushima K.* // *Biolog. Cybernetics.* - 1984. - Vol. 50. - P. 105-113.
37. *Fukushima K.* // *Biolog. Cybernetics.* - 1986. - Vol. 55 (1). - P. 5-15.
38. *Fukushima K., Miyake S., Takayuki I.* // *IEEE Trans on Systems, Man and Cybernetics.* - 1983. - Vol. SMC-13, № 5. - P. 826-834.
39. *Fukushima K.* // *Appl. Opt.* - 1987. - Vol. 26 № 23. - P. 4985-4992.