

Глава 7

НЕЙРОННЫЕ СЕТИ ПРИ ИЗУЧЕНИИ КОГНИТИВНЫХ ПРОЦЕССОВ

Нейронные сети представляют собой область сетевого анализа, отличную от классической сетевой теории. Сходство заключается в использовании принципа дискретности моделируемого объекта и связности модели. Отличие в том, что здесь крайне важна последовательность соединений вершин графа. Можно сказать, что нейронные сети пред-ставимы ориентированными графами.

В отличие от классического сетевого анализа, вершина здесь имеет другой содержательный смысл. Это не актор, не единичный действующий субъект или страта, связанные с другими акторами или стратами. Вершина представляет собой скорее кибернетический черный ящик с одним или несколькими входами, одним или несколькими выходами. Этот ящик является мельчайшей единицей преобразования информации. Очевидно, что у одной такой единицы просто нет ресурсов для восприятия, анализа и представления информации о сложном объекте. И здесь особую роль приобретает организация множества единичных кибернетических ящиков в сложную сеть с множеством входов на одном полюсе (восприятие информации об объекте) и множеством выходов на другом полюсе (реакции системы на этот объект).

Организация (или сетевизация) единичных кибернетических ящиков диктуется последовательностью обработки информации (и сама диктует такую последовательность): после анализа на одном уровне информация поступает на следующий и т. д., пока системой не будет сделан окончательный результирующий вывод. Обычно система выстраивается слоями и информация не поступает в следующий слой, пока не прошла предварительную обработку в предыдущем.

Следующее принципиальное отличие от классического сетевого анализа: в одной системе объединены принципиально разные вершины, причем отдельно друг от друга они существовать не могут. Один тип вершин — воспринимающие единицы, задача которых состоит в том, чтобы определить: присутствует ли данная черта (за понимание которой ответственна именно эта единица) у объекта восприятия, или нет. Если черта присутствует, она активизируется и передает информацию в другой тип вершины — оценочные единицы. Если черта у объекта восприятия отсутствует, то воспринимающая единица остается неактивной. Итак, информация от воспринимающей единицы поступила в оценочную. Именно в этой вершине происходит оценивание поступившей информации, отнесение ее к определенной категории, если она согласуется с ранее известными образами, или выделение

ее в особый тип, если ничего похожего раньше не встречалось. Как уже говорилось, нейронные сети организованы послойно. Слои воспринимающих единиц перемежаются слоями оценивающих единиц. Для моделирования когнитивных процессов можно применять последовательно несколько однотипных слоев. Самой подходящей методологической аналогией при послойной организации сети является представление двух соседних взаимодействующих слоев в виде двудольного графа.

Интересным кажется соотнесение нейронных сетей с моделируемыми объектами. Нейронная сеть моделирует не изучаемый объект, а восприятие этого объекта, преобразование информации о нем и отнесение его к определенной категории. Несомненно, предварительно требуется «разбить» объект на составные части, продумать способ его представления (например, в пиксельном виде, как единую картину чередующихся пятен или в совокупности отдельных черт, которым к тому же приписана степень «важности» при чтении объекта). Способ представления будет диктовать организацию нейронной сети. В описанном выше случае объект отделен от сети, он предстает только в виде входного образа. Это характерно для задач классификации и типологизации объектов.

Для сложных нейронных сетей моделируются когнитивные процессы обучения оценочных единиц восприятию неполной информации. Это осуществляется, во-первых, за счет определенной организации оценочных клеток, во-вторых, за счет применения обучающих алгоритмов (или обучающих правил), предписывающих определенную последовательность обработки информации. В этом случае объектом изучения является процесс, который непосредственно моделируется сетью. Здесь также необходимо разбить объект на составные части, но каждая из них — не часть входной информации, а часть процесса, который моделируется соответствующей частью нейронной сети.

Обычно один тип сети моделирует один вид процессов или воспринимает один тип образов, однако это не является правилом. Чем разнообразнее объекты, тем более сложной является сеть.

Уместно еще раз подчеркнуть связь сетевых подходов с другими методами математического моделирования и статистическими методами. В нейронных сетях это, прежде всего, кибернетические методы. С учетом того, что нейронная сеть может воспринимать большое количество образов, связь со статистикой также представляется обоснованной.

Итак, нейронные сети — это адаптивные статистические модели, устроенные аналогично структуре головного мозга. Адаптивными они являются потому, что способны обучаться оценивать параметры, используя ограниченное число наблюдений во времени. Некоторые авторы считают, что принципы построения нейронных сетей схожи с дискриминант-

ным анализом, анализом главных компонент или логистической регрессией. Действительно, для анализа классических статистических моделей и нейронных сетей мы можем применять одинаковые математические инструменты. Нейронные сети могут использоваться в различных областях знаний, например, в психологии, инженерии, эконометрике.

Нейронные сети построены из простых образований, мельчайших единиц (units), которые аналогичны нейронам. Эти единицы скреплены взвешенными связями, и обучение сети приводит к изменению их веса. Каждая единица кодируется или соответствует одной черте объекта, которую мы хотим изучить или предсказать. Единицы организованы в слои. Нейронные сети состоят из нескольких слоев. Первый слой является входом, последний — выходом. Срединные слои можно обозначить как скрытые. Информация о входном образе, проанализированная в первом слое, затем оценивается в следующих слоях. Каждая единица (нейрон) получает информацию от предыдущих, анализирует ее и передает заключение на выход — другому нейрону.

Целью сети является построение ассоциаций между информацией на входе и на выходе (входным и выходным образами). В статистических терминах это эквивалентно оценке связи между параметрами. Построение связей итеративно, при каждой итерации сеть «обучается» определять образ. Процесс обучения можно охарактеризовать как «алгоритм» правильного оценивания параметров входного образа.

7.1. ПРИНЦИПАЛЬНАЯ СХЕМА НЕЙРОННЫХ СЕТЕЙ, АРХИТЕКТУРА ВОСПРИЯТИЯ ИНФОРМАЦИИ И РАСПОЗНАВАНИЯ ОБРАЗОВ

Модель восприятия образов была сконструирована Ф. Розенблаттом в 1950-х годах (Rosenblatt, 1961). По сути, это была первая нейронная модель, способная обучаться. Модель могла узнавать, или классифицировать, образы — относить стимулирующий образ к той или иной категории.

В процессе восприятия объект сначала «регистрируется» на сетчатке глаза, а затем «узнается» клетками мозга. Клетки сетчатки являются входными — через них образ объекта попадает в сеть. Обычно они бывают либо активными (состояние 1), либо неактивными (состояние 0). Клетки мозга называют **ассоциативными**, их задача — отнести образ, воспринятый сетчаткой, к той или иной категории, известной ранее клеткам мозга. Они являются выходом для образа, воспринятого сетчаткой.

Выходные клетки, выдающие информацию, работают по принципу **базового оценочного модуля**, который был описан Маккаллохом и Питтсом (McCulloch & Pitts, 1943). Оценочный модуль является моделью нейрона, которая интерпретируется как простейшая логическая машина. Она способна находиться в двух состояниях: активном и неактивном. Степень активации определяется интегрированной информацией, по-

ступившей из нескольких источников. Оценочный модуль связан с каждой клеткой сетчатки, которая принимает только одно из двух значений: 1 или 0 (см. рис. 23); 0 соответствует неактивизированной входной клетке, 1 — активизированной.

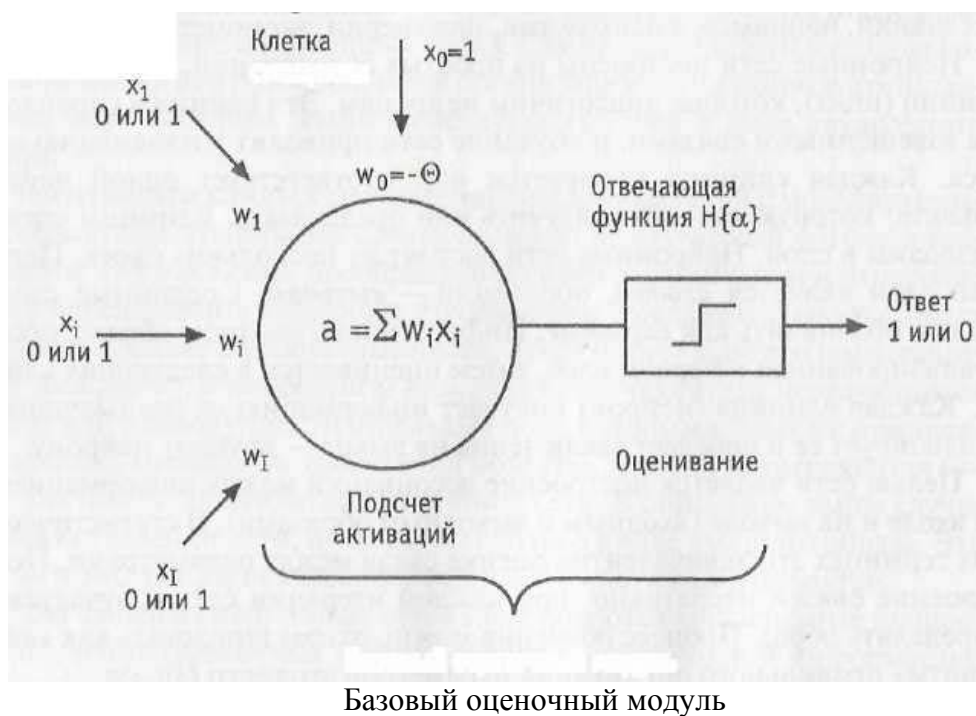


Рис. 23. Базовый оценочный модуль подсчитывает активации a как взвешенную сумму его входов. Если $a < 0$, то ответ модуля равен 0, если $a > 0$, то ответ модуля равен 1.

Выходные клетки связаны с нейронами, и эти связи подобны конъюнкциям хромосом в настоящих нейронах. Каждой связи присваивается вес, который увеличивает или уменьшает уровень активации данной входной клетки. Положительный вес может рассматриваться как возбудитель активации, а отрицательный — как ее замедлитель.

Базовый оценочный модуль подсчитывает активации как сумму позитивных и негативных весов, поступающих от активных входных клеток (ценность которых равна 1). Если суммарная активация больше некоторого порогового значения, тогда на выходе базового оценочного модуля мы получаем 1. Пороговое значение определяется при помощи отвечающей функции, которая может быть линейной или нелинейной (эти случаи мы рассмотрим позже). В нашем же простейшем случае она имеет ступенчатый вид. Если суммарная активация меньше некоторого порогового значения, то на выходе базового оценочного модуля мы получаем 0.

Формально вычислить уровень активации оценочного модуля можно следующим образом:

$$a = \sum_i^I w_i x_i = w_1 x_1 + w_2 x_2 + \dots + w_i x_i + \dots + w_I x_I, \quad (21)$$

где a — активация оценочного модуля, x_i — состояние i входной клетки, I — число входных клеток, w_i — вес, связывающий i -ю входную клетку с оценивающим модулем.

Оценочный модуль также может быть в неактивном состоянии (на выходе выдавать 0), если уровень его активации меньше, чем пороговое значение Θ , или в активном состоянии, если уровень его активации больше, чем Θ . Часто Θ просто равна 0. Формально ответ оценочного модуля можно представить как ступенчатую функцию:

$$\text{ответ} = H \{a\} = \begin{cases} 0 & \text{для } a \leq \Theta \\ 1 & \text{для } a > \Theta \end{cases}. \quad (22)$$

Ответ оценочного модуля может быть не только равным 1 или 0, но и располагаться в интервале $[0; 1]$ — принимать весовое значение. В этом случае одна входная клетка (нулевая входная клетка, так называемая клетка смещения, клетка систематической погрешности) всегда активна.

Ее вес w_0 всегда равен $-\Theta$. Мы можем переписать предыдущее выражение (20) для этого случая:

$$\text{ответ} = H \{a\} = \begin{cases} 0 & \text{для } a + w_0 \leq \Theta \\ 1 & \text{для } a + w_0 > \Theta \end{cases}, \quad (23)$$

где $w_0 = -\Theta$. Понятие «оценочное смещение» часто используется как синоним для порогового значения.

7.2. АРХИТЕКТУРА ВОСПРИЯТИЯ

Формально воспринимающая сеть состоит из I входных клеток, связанных с J выходными клетками — базовыми оценочными модулями. Каждая выходная клетка получает информацию из каждой входной, суммирует ее и трансформирует сумму в выход, используя взвешивающую функцию. Связи между входами и выходами удобно отобразить в виде матрицы W размерностью $I \times J$.

$$W = \begin{bmatrix} w_{1,1} & w_{1,2} & \dots & w_{1,J} \\ w_{2,1} & w_{2,2} & \dots & w_{2,J} \\ \dots & \dots & \dots & \dots \\ w_{I,1} & w_{I,2} & \dots & w_{I,J} \end{bmatrix} \quad (24)$$

В этой матрице каждый элемент w_{ij} представляет (показывает) силу связи между i -й входной клеткой и j -й выходной клеткой.

Всего может быть K входных образов (см. рис. 24). Каждый входной образ выражен I -мерным вектором, обозначаемым x_k . Каждому K соответствует выходной образ, который представлен J -мерным вектором, обозначаемым t_k . Обучение выбору связей между входными-выходными парами происходит через модификацию силы связей между входными и выходными клетками. Ответом на стимул x_k становится вектор $J \times 1$, обозначаемый t_k («шляпа» показывает, что это оценка целевого t_k), равный $t_k = H\{W^T x_k\}$, где T есть операция транспонирования (мы меняем местами строки и столбцы матрицы); а функция взвешивания W применяется элементарным умножением на x_k . Для достижения наиболее оптимальных весов связей используются специальные обучающие правила. Самыми известными являются правила Видроу-Хоффа (Widrow-Hoff, 1985), правило дельты и правило обучающего восприятия.

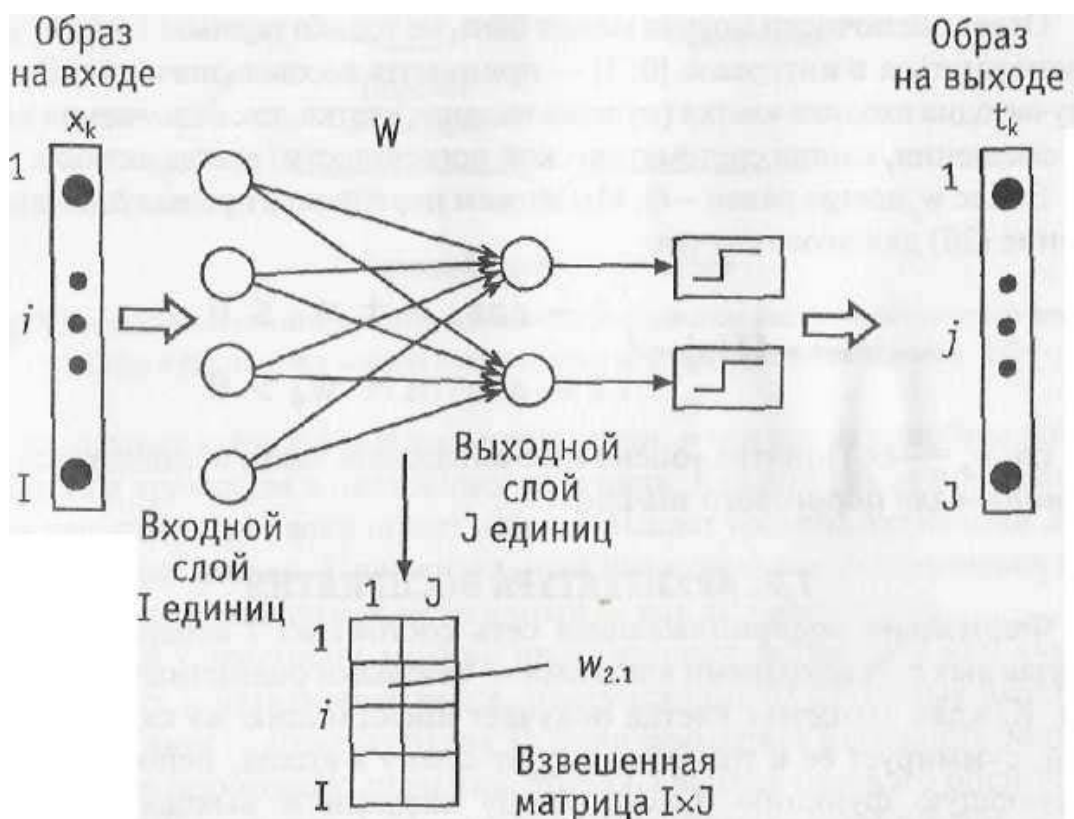


Рис. 24. Восприятие образа сетью из I входных и J выходных единиц

7.3. ЛИНЕЙНАЯ АВТОАССОЦИАТИВНАЯ ПАМЯТЬ

Выше была описана простейшая единица нейронной сети — элементарный блок; из таких блоков и складывается вся сеть. Принципиальный момент — разделение (различение) механизма, **воспринимающего информацию**, и механизма, **оценивающего** ее. Строение оценивающего

механизма, его работа по распознаванию образов и пополнению базы данных представлены ниже при описании автоассоциативной памяти. Автоассоциативная память — это внутренняя самообучающаяся часть нейронной сети. Обучение нейронной сети под влиянием внешних воздействий будет описано в разделе «Гетероассоциативная память».

Автоассоциативная память есть однослойная сеть, состоящая из полностью взаимосвязанных линейных модулей, которые управляются параллельно. Линейный модуль — это простейшая единица, чей выход прямо пропорционален уровню ее активации (выход есть линейная функция от активации). Целью автоассоциативной памяти является классификация и последующее опознание стимула, который поступает в сеть в виде входного образа. Каждый образ проходит идентификацию и добавляется в «банк данных» памяти. Также необходимо опознать стимул, у которого отсутствует какая-либо часть, или входящий образ, представляющий собой искаженную версию уже известного стимула.

Представление индивидуального стимула не локализовано в памяти, но распределено по всей сети. Автоассоциативная память является «контентно-адресной»: каждая часть или черта образа хранится в отдельной клетке. В отличие от компьютерной памяти, которая использует точное местоположение или адрес в памяти для поиска информации, автоассоциативная память способна найти целые образы, представленные только частью информации. Например, Кохонен (Kohonen, 1977) показал, что автоассоциативная память может использоваться для восстановления изображений лиц. Сначала автоассоциативная память была заполнена изображениями 100 лиц, которые хранились в пиксельном формате. Затем на вход поступали тестовые задания — неполные изображения лиц. Результаты продемонстрировали, что система реконструировала изображения, убедительно похожие на оригинал.

Автоассоциативная память широко применяется для моделирования обучения человека восприятию и запоминанию. Она увеличивает число тестируемых (формализуемых) поведенческих наклонностей, а также возможности дать правильный ответ, даже если он требуется очень быстро или входная информация является неполной (поскольку мы используем контекст, «чувствуем» пробел). Например, у вас не будет проблем с ответом на вопрос: «Сколько животных каждого вида взял Нах на ковчег, чтобы спасти их от потопа?» Вы можете даже не заметить, что вопрос неправильный (Ной, а не Нах, взял животных на ковчег),

Линейная автоассоциативная память строится из полностью взаимосвязанных линейных модулей. Линейный модуль очень похож на базовый оценочный модуль (см. рис. 25). Линейный модуль подсчитывает активационный уровень, складывая все взвешенные внешние значения, а затем преобразует эти активации. Данное преобразование,

называемое **линейной трансферной функцией**, делает выход пропорциональным активации.

Если модуль имеет / входов, обозначаемых $x_1, \dots, x_i, \dots, x_I$, каждому из которых соответствует вес, выраженный w_i то активация клетки может быть вычислена так:

$$a = \sum_i^I w_i x_i \quad (25)$$

Ответ клетки будет равным γa , где γ — константа. Через векторные обозначения, если x — входной вектор и w — вектор весов, активация клетки может быть подсчитана так:

$$a = \sum_i^I w_i x_i = w^T x, \quad (26)$$

где $w^T x$ — скалярное произведение векторов w и x . Ответ клетки будет пропорционален ее активации, следовательно, он должен рассчитываться так:

$$\text{ответ} = \gamma a = \gamma w^T x \quad (27)$$

Обучение осуществляется путем изменения оценок нейронных весов.

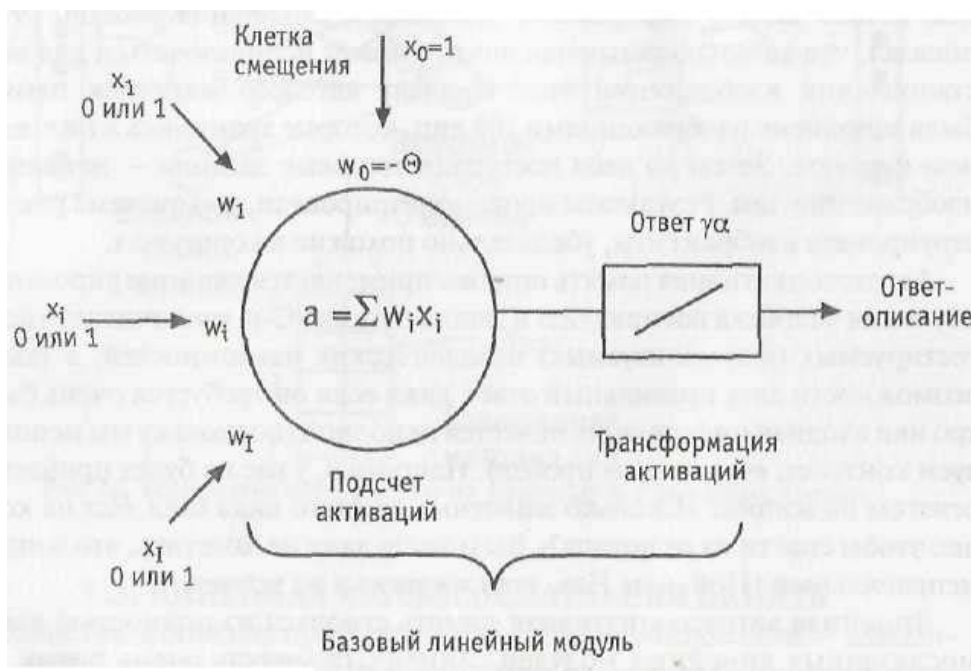


Рис. 25. Базовый линейный модуль подсчитывает активацию как сумму взвешенных активаций входов. Ответ-описание пропорционален активации модуля

Автоассоциативная память — это сеть I линейных модулей, которые полностью взаимосвязаны (каждая клетка соединена с любой другой и также сама с собой) изменяющимися связями (см. рис. 26).

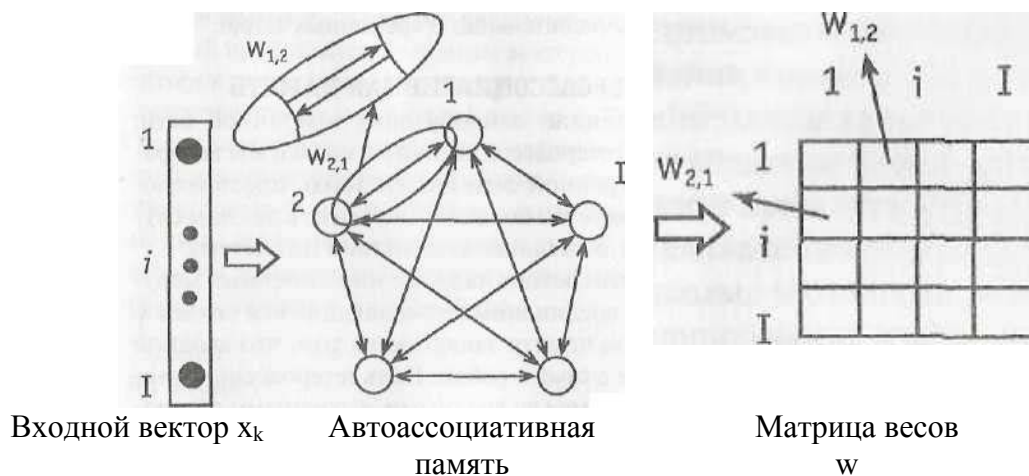


Рис. 26. Автоассоциативная память, состоящая из I клеток

Связи представлены в квадратной матрице W :

$$W = \begin{bmatrix} w_{1,1} & w_{1,2} & \dots & w_{1,I} \\ w_{2,1} & w_{2,2} & \dots & w_{2,I} \\ \dots & \dots & \dots & \dots \\ w_{I,1} & w_{I,2} & \dots & w_{I,I} \end{bmatrix} \quad (28)$$

В этой матрице элемент w_{ii} выражает силу связи между клетками i и i' . Данная связь двунаправлена и в общем случае симметрична ($w_{ii} = w_{i'i}$). Каждый из K объектов занесен в память в виде I -мерного вектора, обозначаемого x_k , в котором любой элемент соответствует определенной черте описываемого образа. Каждый компонент в этом векторе используется как вход в клетку автоассоциативной памяти. K входных образов образуют матрицу X размерностью $I \times K$, где k -й столбец есть x_k и каждый элемент x_{ik} представляет оценку i черты k образа.

Результатом модификации оценки связей через обучение будут выходные образы. При этом в качестве обучающего инструмента могут использоваться два разных правила обучения. Первое обучающее правило, Хеб-биана (Hebbian, 1949), требует минимальных вычислений. Его можно

описать как «непересматривающее», потому что обратная связь не заканчивается на корректировке ответа, вес корректируется в течение всего процесса обучения. Второе обучающее правило, Видроу-Хоффа (Widrow-Hoff, 1985), требует больших вычислительных и временных затрат.

7.4. ЛИНЕЙНАЯ ГЕТЕРОАССОЦИАТИВНАЯ ПАМЯТЬ

Выше был рассмотрен принцип самообучения нейронной сети. При исследовании линейной гетероассоциативной памяти мы возвращаемся к представлению нейронной сети как системы, преобразующей множество входов во множество выходов. Обучение в данном случае осуществляется в результате внешних воздействий на систему.

Гетероассоциативная память состоит из двух слоев линейных модулей — входного и выходного, соединенных изменяющимися связями. Отличие от автоассоциативной памяти заключается том, что входной образ связан с выходным, а не с самим собой. Цель гетероассоциативной памяти — установить связи между входными-выходными образами так, чтобы при получении определенного входного образа память могла выдать соответствующий ему выходной образ.

Гетероассоциативная память, в общем, решает задачи распознавания образов и классификации. Отличие этого типа памяти в том, что ее базовый модуль линейен, а вход и выход непрерывны. С позиции статистики, внутренняя работа гетероассоциативной памяти подобна технике многомерного регрессионного анализа и может быть исследована с помощью математического аппарата автоассоциативной памяти. С когнитивной точки зрения, эта память копирует ассоциативные правила человеческой памяти.

Машине было предложено изучить снимки человеческого лица в фас и на три четверти для того, чтобы затем смоделировать вид сбоку. После анализа снимка в фас машина дала более точное изображение вида сбоку, чем после изучения снимка на три четверти.

Строительный блок линейной гетероассоциативной памяти — базовый линейный модуль (см. рис. 27). Особенностью линейной гетероассоциативной памяти является композиция сети, которая состоит из I входных клеток, связанных с J выходными клетками. Каждая выходная клетка получает веса от всех входных клеток, суммирует их и преобразует сумму в ответ через линейную трансферную функцию. Веса связей между входными и выходными клетками представлены в матрице W размером $I \times J$:

$$W = \begin{bmatrix} w_{1,1} & w_{1,2} & \dots & w_{1,J} \\ w_{2,1} & w_{2,2} & \dots & w_{2,J} \\ \dots & \dots & \dots & \dots \\ w_{I,1} & w_{I,2} & \dots & w_{I,J} \end{bmatrix} \quad (29)$$

В этой матрице элемент w_{ij} отражает силу связей между i -й входной и j -й выходной клетками.

Каждый из K входных образов представлен I -мерным вектором, обозначаемым x_k . Каждому входному соответствует выходной образ, который представлен J -мерным вектором, обозначаемым t_k . K входных векторов располагаются в матрице X размерностью $I \times K$, где k -му столбцу соответствует k -й входной вектор x_k . Точно так же местоположение K выходных образов отражено в матрице T размерностью $J \times K$, где k -му столбцу соответствует k -й выходной вектор t_k . Обучение сопоставлению входных-выходных образов осуществляется через модифицирование силы связи между входными и выходными клетками. Эти модификации также могут быть осуществлены при использовании обучающих правил Хеббiana и Видроу-Хоффа.

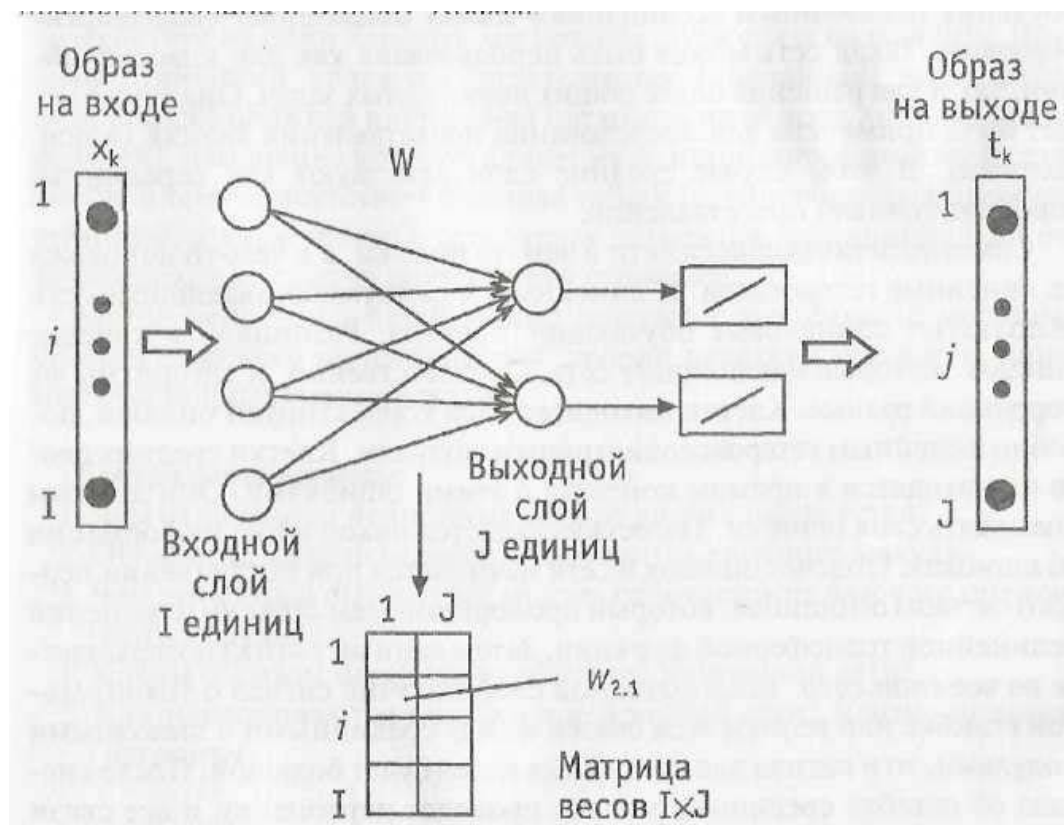


Рис. 27. Архитектура гетероассоциативной памяти

7.5. САМОПРОДУЦИРОВАНИЕ ОШИБОК

Основной проблемой ранних сетевых моделей была их ограниченность линейными функциями. Маккалох и Питтс (McCulloch & Pitts, 1943) предложили между входными и выходными слоями клеток добавлять промежуточные слои. Однако упускалось из виду, что правила обучения должны учитывать и изменять веса промежуточных слоев. Таким образом, промежуточные слои выпадали из зоны контроля,

ошибки накапливались; в результате усложненная нейронная сеть работала некорректно. Это, вероятно, и было одной из причин падения интереса к нейронным сетям в конце 1960-х годов. Повышение интереса к ним в 1980-х годах было обусловлено открытием алгоритма, который позволял адаптировать веса любой связи сети при обучении ассоциирования пар образов.

Самоувеличивающаяся сеть— это многослойная сеть, состоящая из нелинейных модулей. Подобно линейному, нелинейный модуль подсчитывает уровень активации путем суммирования всех взвешенных активаций, которые он получает. В отличие от линейных модулей, нелинейный преобразует эти активации в ответ через нелинейную трансферную функцию. Целью самоувеличивающейся сети является обучение нелинейным ассоциациям между входящими-выходящими образами. Такая сеть может быть использована как для классификации, так и для решения более общих нелинейных задач. Она также может быть применена для исследования представления знаний (в психологии). В этом случае средние слои действуют как срединные (промежуточные) представления.

Самоувеличивающиеся сети в чем-то похожи, а в чем-то непохожи на линейные гетеросвязи. И линейные, и самоувеличивающиеся сети используют одинаковые обучающие правила. Разница— в природе ошибок, которые накапливает сеть. Соответственно, и алгоритмы их коррекции разные. Клетки выходного слоя корректируют ошибки, подобно линейным гетероассоциативным модулям. Клетки средних слоев не находятся в прямом контакте с этими ошибками. Они должны оценивать свои ошибки. Так осуществляется накопление информации об ошибках. Подсчет ошибок в сети начинается при поступлении первого сигнала об ошибке, который пропорционален степени изменений нелинейной трансферной функции. Затем данный сигнал возвращается во все слои сети. Если выходной слой получил сигнал о значительной ошибке или велики веса связей между срединными и выходными модулями, то и сигнал для срединных слоев будет большой. После сигнала об ошибке срединные модули проводят переоценку, и все связи пропорционально перевзвешиваются. Кратко алгоритм коррекции ошибок выглядит так:

1. определение составных слоев нелинейных модулей;
2. подсчет сигналов об ошибке с учетом степени трансформации нелинейной функции;
3. рассредоточение ошибки между всеми модулями сети пропорционально их связям;
4. оценивание сигнала об ошибке срединными модулями и коррекция весов их связей.

Потенциал применения нейронных сетей в социологии крайне велик. Нейронные сети помогают осмыслить и последовательно представить мыслительные и когнитивные процессы; возникновение и изменение терминов, понятий; воздействие массового сознания на изменение терминов; последовательность представления информации и ее воздействия и т. д.

С одной стороны, нейронные сети являются одной из областей дискретной математики и, соответственно, сетевых подходов. С другой стороны, наблюдаются существенные отличия, которые и позволяют отделить это направление от классического сетевого анализа. Например, вершины графа в нейронных сетях не являются закрытыми и самодостаточными, как в классической сетевой теории. Их нельзя вырвать из контекста, они работают как необходимый элемент всей сети. Воздействуя на один элемент, мы воздействуем сразу на всю сеть. Другими отличиями являются определенная (послойная) архитектура сети, а также сложная внутренняя организация ее вершин (клеток или модулей), изменение которой влияет на функционирование всей сети. Вместе с тем, присутствует большая гибкость, способность к преобразованиям, к включению в сеть новых элементов, принципиально отличающихся по своему внутреннему строению.

Метод нейронных сетей можно сочетать с различными методами кибернетического моделирования, теории нечетких множеств, лингвистической переменной и т. д.

Контрольные вопросы:

1. Чем отличаются нейронные сети от других типов сетей?
2. Объясните механизм работы базового оценочного модуля.
3. Чем линейный оценочный модуль отличается от базового оценочного модуля?
4. Каким образом происходит обучение нейронной сети?
5. Как накапливаются ошибки в нейронной сети? Кратко опишите алгоритм.