СЕТИ С ОБРАТНЫМИ СВЯЗЯМИ

Рассмотренные ранее нейросетевые архитектуры относятся к классу сетей с направленным потоком распространения информации и не содержат обратных связей. После обучения на этапе функционирования сети каждый нейрон выполняет свою функцию – передачу выходного сигнала – ровно один раз. В общем случае может быть рассмотрена нейронная сеть, содержащая произвольные обратные связи, то есть пути, передающие сигналы от выходов к входам. Отклик таких сетей является динамическим, т. е. после подачи нового входа вычисляется выход и, передаваясь по обратной связи, модифицирует вход. Затем выход повторно вычисляется, и процесс повторяется снова и снова. Для устойчивой сети последовательные итерации приводят к все меньшим изменениям выхода, и в результате выход становится постоянным. Для многих сетей процесс никогда не заканчивается, такие сети называют неустойчивыми. Неустойчивые сети обладают интересными свойствами и могут рассматриваться в качестве примера хаотических систем, но для большинства практических приложений используются сети, которые дают постоянный выход.

Среди различных конфигураций искуственных нейронных сетей (НС) встречаются такие, при классификации которых по принципу обучения, строго говоря, не подходят ни обучение с учителем, ни обучение без учителя. В таких сетях весовые коэффициенты синапсов рассчитываются только однажды перед началом функционирования сети на основе информации об обрабатываемых данных, и все обучение сети сводится именно к этому расчету. С одной стороны, предъявление априорной информации можно расценивать, как помощь учителя, но с другой – сеть фактически просто запоминает образцы до того, как на ее вход поступают реальные данные, и не может изменять свое поведение, поэтому говорить о звене обратной связи с "миром" (учителем) не приходится. Из сетей с подобной логикой работы наиболее известны сеть Хопфилда и сеть Хэмминга, которые обычно используются для организации ассоциативной памяти. Далее речь пойдет именно о них.

Сеть Хопфилда

Проблема устойчивости ставила в тупик первых исследователей. Никто не был в состоянии предсказать, какие из сетей будут устойчивыми, а какие будут находиться в постоянном изменении. Более того, проблема представлялась столь трудной, что многие исследователи были настроены пессимистически относительно возможности ее решения.

К счастью, в работе [2 Cohen M. A., Grossberg S. G. 1983. Absolute stability of global pattern formation and parallel memory storage by compatitive neural networks. IEEE Transactions on Systems, Man and Cybernetics 13:815-26.] была получена теорема, описавшая подмножество сетей с обратными связями, выходы которых в конце концов достигают устойчивого состояния. Это замечательное достижение открыло дорогу дальнейшим исследованиям и сегодня многие ученые занимаются исследованием сложного поведения и возможностей этих систем.

Джон Хопфилд, физик из Калифорнийского технологического института. Он изучал свойства сходимости сетей на основе принципа минимизации энергии, а также разработал на основе этого принципа семейство нейросетевых архитектур. Сделал важный вклад как в теорию, так и в применение систем с обратными связями. Поэтому некоторые из конфигураций известны как сети Хопфилда.

Рассмотрим однослойную сеть с обратными связями, состоящую из n входов и n нейронов (рис. 1). Каждый вход связан со всеми нейронами. Так как выходы сети заново подаются на входы, то y_i – это значение i-го выхода, который на следующем этапе функционирования сети становится i-м входом.

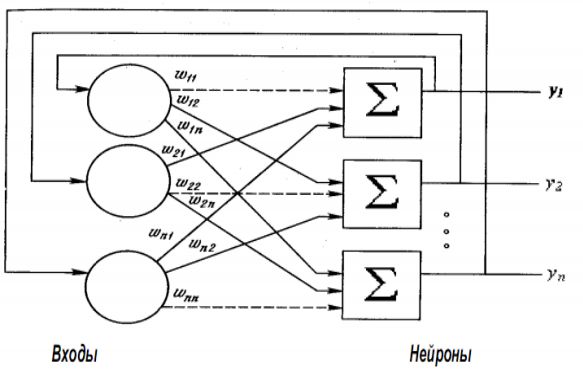

Рис. 1. Модель сети Хопфилда

Совокупность выходных значений всех нейронов y_i на некотором этапе N образует *вектор состояния сети* Y^N. Нейродинамика приводит к изменению вектора состояния на Y^{N+1}.

Обозначим силу синапсической связи от i-го входа к j-му нейрону как w_{ij}. Каждый j-й нейрон сети реализует пороговую активационную функцию следующего вида:

$$y_j^{N+1} = f(s_j) = \begin{cases} -1, s_j < \Theta_j; \\ 1, s_j > \Theta_j; \\ y_j^N, s_j = \Theta_j. \end{cases}$$

Здесь $S_j = \sum_{\substack{i=1 \\ i \neq j}}^{n} y_i^N w_{ij}$; y_j^N – значение выхода j-го нейрона на предыдущем этапе функционирования сети, Θ_j – пороговое значение j-го нейрона.

Изменение состояний возбуждения всех нейронов может происходить одновременно, в этом случае говорят о *параллельной динамике*. Рассматривается также и *последовательная нейродинамика*, при которой в данный момент времени происходит изменение состояния только одного нейрона. Многочисленные исследования показали, что свойства памяти нейронной сети практически не зависят от типа динамики. При моделировании нейросети на обычном компьютере удобнее последовательная смена состояний нейронов. В аппаратных реализациях нейросетей Хопфилда применятся параллельная динамика.

Состояние сети – это просто множество текущих значений сигналов Y от всех нейронов. В первоначальной сети Хопфилда состояние каждого нейрона менялось в дискретные случайные моменты времени, в последующей работе состояния нейронов могли меняться одновременно. Так как выходом бинарного нейрона может быть только -1 или 1(или 0/1) (промежуточных уровней нет), то текущее состояние сети является двоичным числом, каждый бит которого является сигналом Y некоторого нейрона.

Функционирование сети легко визуализируется геометрически. На рис. 2а показан случай двух нейронов в выходном слое, причем каждой вершине квадрата соответствует одно из четырех состояний системы (00, 01, 10, 11). На рис. 2б показана трехнейронная система, представленная кубом (в трехмерном пространстве), имеющим восемь вершин, каждая из которых помечена трехбитовым бинарным числом. В общем случае система с n нейронами имеет 2n различных состояний и представляется n-мерным гиперкубом.

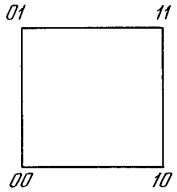

Рис. 2а. Два нейрона порождают систему с четырьмя состояниями

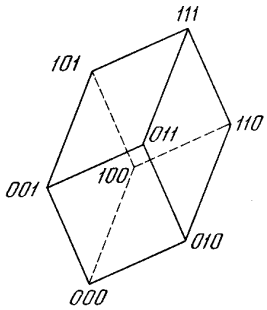

Рис. 2б. Три нейрона порождают систему с восемью состояниями

Когда подается новый входной вектор, сеть переходит из вершины в вершину, пока не стабилизируется. Устойчивая вершина определяется сетевыми весами, текущими входами и величиной порога. Если входной вектор частично неправилен или неполон, то сеть стабилизируется в вершине, ближайшей к желаемой.

Устойчивость

Как и в других сетях, веса между слоями в этой сети могут рассматриваться в виде матрицы W. В работе [2] показано, что сеть с обратными связями является устойчивой, если ее матрица симметрична и имеет нули на главной диагонали, т. е. если $w_{ij} = w_{ji}$ и $w_{ii} = 0$ для всех i.

Устойчивость такой сети может быть доказана с помощью элегантного математического метода.

Работа сети Хопфилда может быть объяснена терминами энергетического ландшафта. Существует ландшавт, который представляет собой гористую местность, на вершине которой находится шар. Потом шар катится вниз но наклонной, пока не остановится в некоторой ложбинке. Эти ложбинки отражают постоянные состояния сети, и каждая из них соответствует определенному заданному образу (образу обучения). В такой интерпретации шар, который имеет большую потенциальную энергию, катиться в ложбинку с меньшей потенциальной энергией, попадая в локальный минимум. Чтобы снова очутиться в начальном состоянии шар должен выполнить в физическом значении работу, т.е. потратить энергию. Т.о. работа сети Х-да. может быть охарактеризована энергетической функцией.

Если в процессе анализа персептрона выбор такой функции особенных сложностей не вызывал, поскольку известны реальные и требуемые значения выходов нейрона, в этом случае ситуация другая. Во-первых структура сети Х-да не позволяет заранее определить путь решения, т.е. требуемую или необходимую последовательность состояний нейронов. Во-вторых, наличие обратных связей приводит к тому что в любой момент времени выходы сети дают набор входов. Кроме того, необходимо учесть отсутствие у нейронов собственных обратных связей.

Допустим, что найдена функция, которая всегда убывает при изменении состояния сети. В конце концов эта функция должна достичь минимума и прекратить изменение, гарантируя тем самым устойчивость сети. Такая функция, называемая функцией Ляпунова, для рассматриваемых сетей с обратными связями может быть введена следующим образом:

$$E = -\frac{1}{2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n} w_{ij} y_i y_j - \sum_{j=1}^{n} y_j x_j + \sum_{j=1}^{n} y_j \Theta_j \qquad (6.2)$$

где E – искусственная энергия сети; w_{ij} – вес от выхода нейрона i к входу нейрона j; y_j – выход нейрона j; x_j – внешний вход нейрона j; Θ_j – порог нейрона j.

Хопфилд доказал, что при активации сети функция (6,2) не возрастает и достигает локального минимума в некотором постоянном состоянии. А поскольку количество таких состояний ограничено, сеть достигает одного из них за конечное количество итераций. При этом низины енергет. Ландшафта соответствуют сохраненным образам. Чтобы сеть Х-да правильно функционировала эти низины не должны перекрываться.

Необходимо, выделить. Обычно в литературе подразумевается, что входящий сигнал x_i является кратковременным и в расчетах не учитывается, т.е. вместо (6.2), рассматриваемся функционал

$$E = -\frac{1}{2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n} w_{ij} y_i y_j + \sum_{j=1}^{n} y_j \Theta_j$$

Вычислим изменение функции энергии ΔE, вызванное изменением состояния j-нейрона Δy_j:

$$\Delta E = (-\sum_{i\neq j} w_{ij} y_i + \Theta_j)\Delta y_j = -(s_j - \Theta_j)\Delta y_j$$

(здесь мы воспользовались симметричностью связей и тем, что $w_{ii} = 0$).

Допустим, что величина $s_j > \Theta_j$. Тогда выражение в скобках (в правой части) будет положительным, а из вида активационной функции следует, что новый выход нейрона j должен быть 1, то есть измениться в положительную сторону (или остаться без изменения). Это значит, что $\Delta y_j \geq 0$, и тогда $\Delta E \leq 0$. Следовательно, энергия сети либо уменьшится, либо останется без изменения. Далее, допустим, что величина $s_j <$ порога. Тогда новое значение $y_j = -1$ и величина Δy_j может быть только отрицательной или нулем. Следовательно, опять энергия должна уменьшиться или остаться без изменения. Если величина $s_j =$ порогу Θ_j , Δy_j равна нулю и энергия остается без изменения.

Эти рассуждения показывают, что любое изменение состояния нейрона либо уменьшит функцию энергии, либо оставит ее без изменения. Так как функция энергии задана на конечном множестве ($\forall y_i \in \{-1,1\}$), то она ограничена снизу и вследствие непрерывного стремления к

уменьшению в конце концов должна достигнуть минимума и прекратить изменение. По определению такая сеть является устойчивой.

Поверхность функции энергии E в пространстве состояний имеет весьма сложную форму с большим количеством локальных минимумов. Стационарные состояния, отвечающие минимумам, могут интерпретироваться, как образы памяти нейронной сети. Сходимость к такому образу соответствует процессу извлечения из памяти. При произвольной матрице связей W образы также произвольны. Для записи в память сети какой-либо конкретной информации требуется определенное значение весов W, которое может получаться в процессе обучения.

Правило обучения Хебба

Работа [2 Hebb D. O. 1949. Organization of behavior. New York: Science Editions.] обеспечила основу для большинства алгоритмов обучения, которые были разработаны после ее выхода. В предшествующих этой работе трудах в общем виде определялось, что обучение в биологических системах происходит посредством некоторых физических изменений в нейронах, однако отсутствовали идеи о том, каким образом это в действительности может иметь место. Основываясь на физиологических и психологических исследованиях, Хэбб в [2] интуитивно выдвинул гипотезу о том, каким образом может обучаться набор биологических нейронов. Его теория предполагает только локальное взаимодействие между нейронами при отсутствии глобального учителя; следовательно, обучение является неуправляемым, Несмотря на то, что его работа не включает математического анализа, идеи, изложенные в ней, настолько ясны и непринужденны, что получили статус универсальных допущений. Его книга стала классической и широко изучается специалистами, имеющими серьезный интерес в этой области.

Правило обучения для сети Хопфилда опирается на исследования Дональда Хебба (D.Hebb, 1949), который предположил, что синаптическая связь, соединяющая два нейрона будет усиливаться, если в процессе обучения оба нейрона согласованно испытывают возбуждение либо торможение. Простой алгоритм, реализующий такой механизм обучения, получил название *правила Хебба*. Рассмотрим его подробно.

Пусть задана обучающая выборка образов X^k, k=1, ….,K. Требуется построить такую матрицу связей W, что соответствующая нейронная сеть будет иметь в качестве стационарных состояний образы обучающей выборки (значения порогов нейронов Θ_j положим равными нулю). В случае одного обучающего образа $X=(x_1, x_2, ..., x_n)$, $x \in \{-1,1\}$, правило Хебба приводит к матрице: $w_{ij}= x_i x_j$, $i \neq j$, $w_{ii}=0$.

Покажем, что состояние Y=X является стационарным для сети Хопфилда с данной матрицей W. Действительно, значение функции энергии в состоянии X является для нее глобальным минимумом:

$$E(X) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}x_i x_j = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j x_i x_j = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} x_i^2 x_j^2 = -\frac{1}{2}n^2,$$

то есть сеть прекратит изменения, достигнув состояния X.

Для запоминания K образов применяется итерационный процесс:

$$w_{ij}^k = w_{ij}^{k-1} + x_i^k x_j^k, k = \overline{1,K} \quad (\text{считаем, что } w_{ij}^0 = 0).$$

Этот процесс приводит к полной матрице связей:

$$w_{ij} = \sum_{k=1}^{K} x_i^k x_j^k$$

Сеть Хопфилда нашла широкое применение в системах *ассоциативной памяти,* позволяющих восстанавливать идеальный образ по имеющейся неполной или зашумленной его версии.

Пример. В качестве примера рассмотрим сеть, состоящую из 70 нейронов, упорядоченных в матрицу 10 × 7.

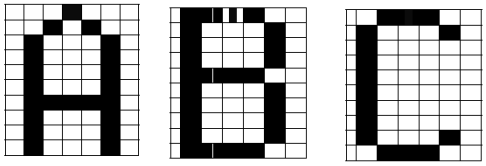

Рис. 3. Идеальные образы обучающей выборки

Сеть обучалась по правилу Хебба на трех идеальных образах – шрифтовых начертаниях латинских букв A, B и C (рис. 3). Темные ячейки соответствуют нейронам в состоянии +1, светлые – 1. После обучения нейросети в качестве начальных состояний нейронов предъявлялись различные

искаженные версии образов, которые в процессе функционирования сети сходились к стационарным состояниям. Для каждой пары изображений на рисунках 4,5,6 левый образ является начальным состоянием, а правый – результатом работы сети – достигнутым стационарным состоянием.

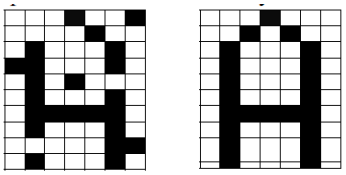

Рис. 4. Сеть Хопфилда распознает образ с информационным шумом

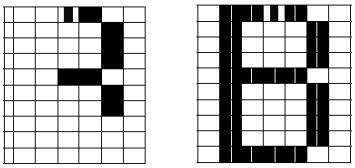

Рис. 5. Сеть Хопфилда распознает образ по его небольшому фрагменту

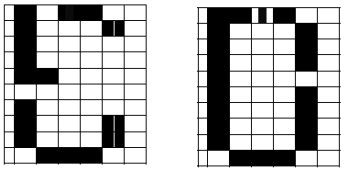

Рис. 6. Сеть Хопфилда генерирует ложный образ

Опыт практического применения сетей Хопфилда показывает, что эти нейросетевые системы способны распознавать практически полностью зашумленные образы и могут ассоциативно узнавать образ по его небольшому фрагменту. Однако особенностью работы данной сети является возможная генерация ложных образов. Ложный образ является устойчивым локальным минимумом функции энергии, но не соответствует никакому идеальному образу. На рис. 2 показано, что сеть не смогла различить, какому из идеальных образов (B или C) соответствует поданное на вход зашумленное изображение, и выдала в качестве результата нечто собирательное.

Ложные образы являются «неверными» решениями, и поэтому для исключения их из памяти сети на этапе ее тестирования применяется механизм «разобучения». Суть их заключается в следующем. Если обученная сеть на этапе тестирования сошлась к ложному образу $Z = (z_1,...,z_n)$ т.о. ее весовые коэффициенты пересчитываются по формуле: $w_{ij}' = w_{ij} - \varepsilon z_i z_j$ где − ε малое число (0 <ε<. 0,1), что гарантирует незначительное ухудшение полезной памяти. После нескольких процедур разобучения свойства сети улучшаются. Это объясняется тем, что состояниям ложной памяти соответствуют гораздо более «мелкие» энергетические минимумы, чем состояниям, соответствующим запоминаемым образом.

Другим существенным недостатком сетей Хопфилда является небольшая емкость памяти. Многочисленные исследования показывают, что нейронная сеть, обученная по правилу Хебба, может в среднем, при размерах сети n , хранить не более чем 0.14 n различных образов. Для некоторого увеличения емкости памяти сети используется специальный алгоритм ортогонализации образов.

Несмотря на интересные качества, нейронная сеть в классической модели Хопфилда далека от совершенства. Она обладает относительно скромным об'емом памяти, пропорциональным числу нейронов сети N, в то время как системы адресной памяти могут хранить до 2N различных образов, используя N битов. Кроме того, нейронные сети Хопфилда не могут решить задачу распознавания, если изображение смещено или повернуто относительно его исходного запомненного состояния. Эти и другие недостатки сегодня определяют общее отношение к модели Хопфилда, скорее как к теоретическому построению, удобному для исследований, чем как повседневно используемому практическому средству.

Алгоритм

Задача, решаемая данной сетью в качестве ассоциативной памяти, как правило, формулируется следующим образом. Известен некоторый набор двоичных сигналов (изображений, звуковых оцифровок, прочих данных, описывающих некие объекты или характеристики процессов), которые считаются образцовыми. Сеть должна уметь из произвольного неидеального сигнала, поданного на ее вход, выделить ("вспомнить" по частичной информации) соответствующий образец (если такой есть) или "дать заключение" о том, что входные данные не соответствуют ни одному из образцов. В общем случае, любой сигнал может быть описан вектором **X** = { x_i } :i=1...n, n – число нейронов в сети и размерность входных и выходных векторов. Каждый элемент x_i равен либо +1, либо -1. Обозначим вектор, описывающий k-ый образец, через $\mathbf{X^k}$, а его компоненты, соответственно, – x_i^k, k=1..K, k – число образцов. Когда сеть распознаёт (или "вспомнит") какой-либо образец на основе предъявленных ей данных, ее выходы будут содержать именно его, то есть **Y** = $\mathbf{X^k}$, где **Y** – вектор выходных значений сети: **Y** = { y_i}: i=1,...n. В противном случае, выходной вектор не совпадет ни с одним образцовым.

Если, например, сигналы представляют собой некие изображения, то, отобразив в графическом виде данные с выхода сети, можно будет увидеть картинку, полностью совпадающую с одной из образцовых (в случае успеха) или же "вольную импровизацию" сети (в случае неудачи).

В работе сети выделяют 3 фазы:

1. инициализация
2. ввод входного образа
3. вычисление состояния нейронов

1. Инициализация

На стадии *инициализации* сети весовые коэффициенты синапсов устанавливаются следующим образом :

$$w_{ij} = \begin{cases} \sum_{k=1}^{K} x_i^k x_j^k, i \neq j \\ 0, \ i = j \end{cases} \text{ для } i, j = \overline{1,n} \quad (1)$$

Здесь i и j – индексы, соответственно, предсинаптического и постсинаптического нейронов; x_i^k, x_j^k – i-ый и j-ый элементы вектора k-ого образца.

2. ввод входного образа

Фактически осуществляется непосредственной установкой компонент выходных сигналов:

$y_i^0 = x_i$, $i = 1...n$, (2)

поэтому обозначение на схеме сети входных синапсов в явном виде носит чисто условный характер. Ноль справа от y_i означает нулевую итерацию в цикле работы сети.

3. Рассчитывается новое состояние нейронов

Здесь $S_j = \sum_{\substack{i=1 \\ i \neq j}}^{n} y_i^N w_{ij}$; y_j^N – значение выхода j-го нейрона на предыдущем N этапе функционирования сети, N обозначает итерацию, Θ_j – пороговое значение j-го нейрона=0.

$$y_j^{N+1} = f(S_j) = f\left(\sum_{i=1}^{n} w_{ij} y_i^N \right), \quad (3)$$

f – бинарная / биполярная функция активации;

$$f^{N+1}(S_j) = \begin{cases} -1, S_j < 0 \\ 1, S_j > 0 \\ y_j^N, S_j = 0 \end{cases} \quad (4)$$

Проверка, изменились ли выходные значения аксонов за последнюю итерацию. Если да – переход к пункту 3, иначе (если выходы застабилизировались) – конец. При этом выходной вектор представляет собой образец, наилучшим образом сочетающийся с входными данными.