Лабораторная работа №4
Рекуррентные нейронные сети

Цель работы: изучение поведения рекуррентных нейронных сетей Хопфилда.

1. Теоретические сведения

Среди различных архитектур искусственных нейронных сетей встречаются такие, которые по принципу настройки нельзя классифицировать ни как обучение с учителем, ни как обучение без учителя. В таких сетях весовые коэффициенты связей между нейронами рассчитываются перед началом их функционирования на основе информации об обрабатываемых образах, 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 образы до того, как на ее вход поступают реальные данные, и не может изменять свое поведение. Из сетей с подобной логикой работы наиболее известна сеть Хопфилда, которая обычно используется для организации ассоциативной памяти.

Американский исследователь Хопфилд в 80-х годах предложил специальный тип нейросетей, названных в его честь. Они открыли новое направление в теории и практике нейросетей. Сети Хопфилда являются рекуррентными сетями или сетями с обратными связями (feedback networks) и были предназначены первоначально для решения следующей задачи. Имеется k образов (например, видеоизображений или фотоснимков), представленных, например, n-разрядными двоичными векторами. Задача нейросети состоит в запоминании и последующем распознавании этих k образов. Важно подчеркнуть, что распознаваемые образы при этом могут быть искажены или зашумлены.

Сети Хопфилда обладают рядом отличительных свойств:

1. симметрия дуг: сети содержат n нейронов, соединенных друг с другом. Каждая дуга (соединение) характеризуется весом w_{ij} , причем имеет место:

$$\forall i, j \in N : i \neq j \exists_1 w_{ij}$$

где $N = \{1, 2, ..., n\}$ – множество нейронов;

2. симметрия весов: вес соединения нейрона n_i с нейроном n_j равен весу обратного соединения

$$w_{ij} = w_{ji}; w_{ii} = 0;$$

бинарные входы: сеть Хопфилда обрабатывает бинарные входы {0,1} или {-1,1}. В литературе встречаются модели сетей как со значениями входов и выходов 0 и 1, так и –1, 1. Для структуры сети это безразлично. Однако формулы для распознавания образов (изображений) при использовании значений –1 и 1 для входов и выходов нейронов сети Хопфилда получаются нагляднее, поэтому эти значения и предполагаются ниже.

Определение. Бинарная сеть Хопфилда определяется симметричной матрицей W связей между нейронами с нулевыми диагональными элементами, вектором T порогов нейронов и знаковой функцией активации или выхода нейронов. Каждый выходной вектор O (O *om output*) с компонентами –1 или 1, удовлетворяющий уравнению $O = F(W_0 - T)$ называется образом для сети Хопфилда.

В начале сеть находится на высоком энергетическом уровне, из которого возможны переходы в различные состояния. Затем энергетический уровень сети уменьшается до тех пор, пока не достигается некоторое конечное состояние, соответствующее в общем случае некоторому локальному минимуму.

Класс сетей Хопфилда содержит только один слой нейронов, причем каждый нейрон соединен с остальными. Обратные связи с выхода нейрона на его же вход отсутствуют. На рисунке 5.1 приведен конкретный пример сети Хопфилда из четырех нейронов.

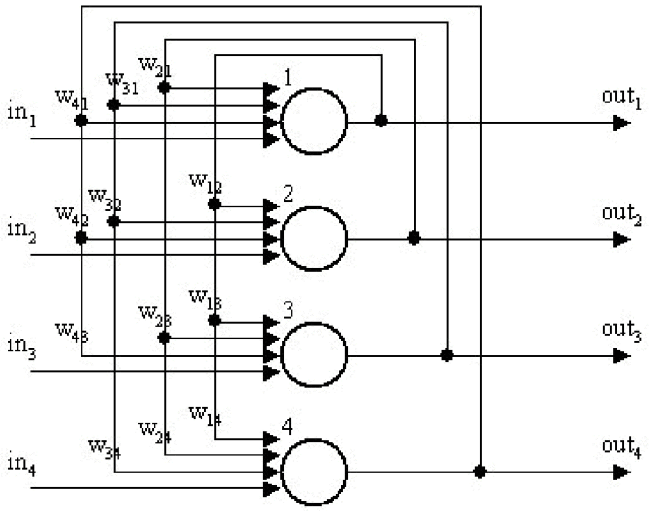

Рисунок 5.1 – Сеть Хопфилда из четырех нейронов.

На рисунке 5.2 представлена структура сети Хопфилда общего вида, содержащая *n* нейронов.

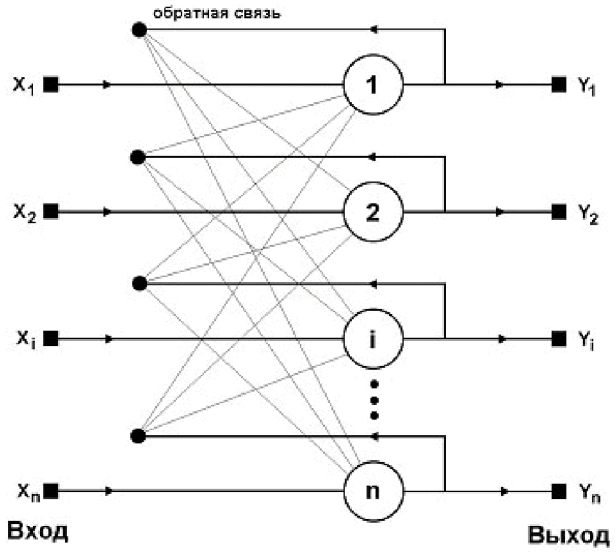

Рисунок 5.2. Структурная схема сети Хопфилда общего вида из n нейронов.

Алгоритм Хопфилда

Обозначим через *Xs* вектор значений признаков образа *s*-го класса, а через Xs_i – его *i*-ю составляющую. При этом алгоритм Хопфилда может быть описан следующим образом:

1. расчет весов связей между нейронами

$$w_{ij} = \begin{cases} \sum_{s=0}^{k-1} x_{si} S_{sj}, & i \neq j \\ 0, & i = j \end{cases}$$

где *k* – число запоминаемых классов образов;

2. инициализация сети путем ввода:

$$O_i(0) = x_i,\ 1 \le i \le n,$$

где $O_i(0)$ – выход *i*-го нейрона в начальный (нулевой) момент времени;

3. применение итерационного правила:

Repeat

$$O_j(t+1) = F\left(\sum_{i=0}^{n-1} w_{ij} O_i(t)\right)$$

$$Until \quad O_j(t+1) = O_j(t), \qquad j = 1,2...n$$

где $O_i(t)$ – выход i – го нейрона в момент времени t.

Соотношения (5.4), (5.5) представим в векторной форме:

$$O(t+1)=F(W_0(t)),\ t=1,2\ \dots\ n$$
$$O(0)=x(0),$$

где x – входной вектор, W – симметричная матрица сети Хопфилда, **F** – вектор-функция активации или выхода нейронов.

Взвешенная сумма входов j–го нейрона сети Хопфилда равна:

$$net_j = \sum_i w_{ij} O_i$$

Функция активации или выхода имеет вид:

$$f(net) = \begin{cases} 1 & для \quad \text{net} > 0 \\ -1 & для \quad \text{net} \le 0 \end{cases}$$

Таким образом, алгоритм Хопфилда может быть сформулирован следующим образом. Пусть имеется некоторый образ, который следует запомнить в сети Хопфилда, тогда веса искомой сети для распознавания этого образа могут быть рассчитаны, т.е. процесс обучения исключается. Если этот образ характеризуется n – мерным вектором $X = (x1, \dots, xn)$, то веса соединений определяются по формуле:

$$w_{ij} = \begin{cases} x_i x_j & для \quad i \neq j \\ 0 & для \quad i = j \end{cases}$$

то есть вес связи i-го нейрона с j-м равен произведению i-й и j-й составляющих вектора X, характеризующего данный образ. В этом случае сеть Хопфилда, характеризуемая матрицей W и порогами $T_i = 0$, $i = 1, 2, \dots, n$, запоминает предъявленный образ.

Веса w_{ij} можно умножить на некоторый положительный коэффициент (например, $\frac{1}{n}$). Использование такого коэффициента особенно целесообразно в тех случаях, когда запоминаемые образы характеризуются большим числом признаков (например, видеоизображение отображается большим числом пикселов).

Пример 5.1. Определение матрицы синаптических весов

Пусть имеется изображение из четырех пикселов (рисунок 5.3):

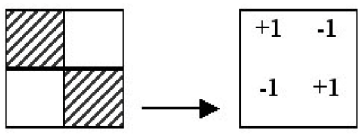

Рисунок 5.3 – Закодированное изображение из 4-х пикселов

Соответствующий вектор (образ) имеет вид: *x =(1,-1,-1,1)*. Стационарная сеть Хопфилда содержит при этом 4 элемента (нейрона). Ее весовая матрица рассчитывается на основе (5.10) и принимает вид:

$$W = \begin{pmatrix} 0 & -1 & -1 & 1 \\ -1 & 0 & 1 & -1 \\ -1 & 1 & 0 & -1 \\ 1 & -1 & -1 & 0 \end{pmatrix}$$

При этом легко можно убедиться в справедливости равенства $X = F(W x)$.

Подадим на входы сети искаженный образ (рисунок 5.4):

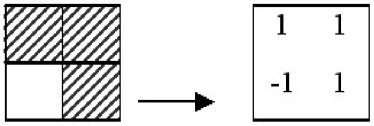

Рисунок 5.4 – Искаженное изображение из 4-х пикселов

Сети потребуется лишь одна итерация, чтобы выдать корректное изображение.

5.2. Распознавание образов сетями Хопфилда.

Обычно сети Хопфилда разрабатываются для запоминания и последующего распознавания большого количества образов (изображений). Обозначим через $x_1, ..., x_k$ эти образы, причем

$$x_j = (x_{j1}, ..., x_{jn})T, j = 1, 2, ..., k,$$

где x_{ji} [j = 1, 2, ..., k, i = 1, 2, ..., n] – *i*-я составляющая *j*-го образа.

Подобно (5.10) введем матрицу *Ws* для *s* – го образа с весами:

$$w_{ij}^{s} = \begin{cases} x_{si}x_{sj} & ,s = 1,2,...,k \quad ,i \neq j \\ 0 \quad , & s = 1,2,...,k \quad ,i = j \end{cases}$$

Из этих матриц можно образовать результирующую матрицу сети для всех k запоминаемых образов $W = (W1 + ... + Wk) / n$,

где *n* – размерность векторов (например, число пикселов в представлении видеоизображения). Если имеется не много изображений (т.е. *k* мало), то сеть с матрицей (5.13) запоминает *k* образов, полагая, естественно, что изображения не сильно коррелированы. С ростом *k* уменьшается вероятность воспроизведения отдельного образа. Приведем утверждение относительно числа *k* запоминаемых образов при использовании сети Хопфилда из *n* нейронов.

Пусть *k* – число запоминаемых образов, а *n* – число признаков (пикселов), и образы, подлежащие запоминанию, не коррелированы, т.е. для двух изображений *j* и *s* сумма $\left|\sum_i x_{ji}x_{si}\right|$ мала. В этом случае при подаче на вход сети с весовой матрицей (5.13) одного образа каждый пиксел корректно воспроизводится с вероятностью при выполнении условия:

$$k \leq 0.15n \quad (n \to \infty)$$

Пример 5.2. Образ характеризуется тысячью признаками (n = 1000). В этом случае бинарная сеть Хопфилда в состоянии запомнить и корректно воспроизвести до 150 образов (например, видеоизображений).

В работе сети Хопфилда можно выделить следующие три стадии:

Инициализация или фаза настройки сети: на этой стадии рассчитываются все веса сети для некоторого множества образов. Подчеркнем еще раз: эти веса не определяются на основе рекуррентной процедуры, используемой во многих алгоритмах обучения с поощрением. По окончании этой фазы сеть в состоянии корректно распознавать все запомненные образы.

Ввод нового образа: нейроны сети устанавливаются в соответствующее начальное состояние по алгоритму (5.3) – (5.7),

Затухающий колебательный процесс: путем использования итеративной процедуры рассчитывается последовательность состояний сети до тех пор, пока не будет достигнуто стабильное состояние, т.е.

$$O_j(t+1) = O_j(t),\ j = 1,\ 2,\ \ldots,\ n,$$

или в качестве выходов сети используются значения (5.16), при которых сеть находится в динамическом равновесии:

$$O = S(W_0),$$

где O – выходной вектор сети.

Динамическое изменение состояний сети может быть выполнено по крайней мере двумя способами: синхронно и асинхронно. В первом случае все элементы модифицируются одновременно на каждом временном шаге, во втором - в каждый момент времени выбирается и подвергается обработке один элемент. Этот элемент может выбираться случайно.

Выход бинарной сети Хопфилда, содержащей n нейронов, может быть отображен бинарным вектором O состояния сети. Общее число таких состояний – $2n$ (вершины n – мерного гиперкуба). При вводе нового входного вектора состояние сети изменяется от одной вершины к другой до достижения сетью устойчивого состояния.

Из теории систем с обратными связями известно: для обеспечения устойчивости системы ее изменения с течением времени должны уменьшаться. В противном случае возникают незатухающие колебания. Для таких сетей Коэном и Гроссбергом доказана теорема, формулирующая достаточные условия устойчивости сетей с обратными связями: «*Рекуррентные сети устойчивы, если весовая матрица* $W = (w_{ij})$ *симметрична, а на ее главной диагонали – нули:*

$$w_{ij} = w_{ji} \text{ для всех } i \neq j;$$
$$w_{ii} = 0 \text{ для всех } i. \text{ »}$$

Обратим внимание, что условия данной теоремы достаточны, но не необходимы. Для рекуррентных сетей отсюда следует, что возможны устойчивые сети, не удовлетворяющие приведенному критерию.

Для доказательства приведенной теоремы Коэна и Гроссберга используется энергетическая функция E, принимающая лишь положительные значения. При достижении сетью одного из своих устойчивых состояний эта функция принимает соответствующее минимальное значение (локальный минимум) – для бинарных образов в результате конечного числа итераций. Эта функция определяется следующим образом:

$$E = -\frac{1}{2}\sum_i \sum_{j \neq i} w_{ij} x_i x_j + \sum_i x_i T_i$$

где x_i – вход i-*го* нейрона, T_i – порог i-го нейрона.

Для каждого образа, вводимого в сеть, можно определить энергию E. Определив значение функции E для всех образов, можно получить поверхность энергии с максимумами (вершинами) и минимумами (низинами), причем минимумы соответствуют образам, запомненным сетью. Таким образом, для сетей Хопфилда справедливо утверждение: *минимумы энергетической функции*

соответствуют образам, запомненным сетью. В результате итераций сеть Хопфилда в соответствии с (5.3) – (5.7) сходится к запомненному образу.

Для запоминания каждого следующего образа в поверхность энергии, описываемой энергетической функцией E, необходимо ввести новую «низину». Однако при таком введении уже существующие «низины» не должны быть искажены.

Для некоторого образа $x = (x_1, x_2, ..., x_n)$ следует минимизировать обе составляющие функции E (5.19). Для того, чтобы второе слагаемое $\sum_i x_i T_i$ было отрицательным, необходимо обеспечить различие знаков входов x_i и порогов T_i. При фиксированных порогах это невыполнимо, поэтому выберем пороги равными нулю. При этом второе слагаемое в (5.19) исключается, а остается лишь первое:

$$E = -\frac{1}{2}\sum_i \sum_{j \neq i} w_{ij} x_i x_j$$

Из общего числа k образов выделим некоторый образ с номером s. При этом энергетическую функцию E (5.20) можно представить так:

$$E = -\frac{1}{2}\sum_i \sum_{j \neq i} w'_{ij}\, x_i x_j - \frac{1}{2}\sum_i \sum_{j \neq i} w^s{}_{ij} x_{si} x_{sj}$$

где w^s_{ij} – составляющая весового коэффициента w_{ij}, вызванная s-м образом, w'_{ij} – составляющая весового коэффициента w_{ij} , вызванная остальными образами, запомненными сетью, x^s_i – значение i-го входа для s-го образа.

Выделенный образ с номером s определяет лишь второе слагаемое в (5.21):

$$E_s = -\frac{1}{2}\sum_i \sum_{j \neq i} w^s{}_{ij} x_{si} x_{sj}$$

Задача минимизации величины E_s эквивалентна максимизации выражения

$$\sum_i \sum_{j \neq i} w^s{}_{ij} x_{si} x_{sj}$$

Входы сети x_{si} принимают значения из множества {-1, 1}, поэтому $x^s{}_i$ всегда положительны. Следовательно, путем разумного выбора весовых коэффициентов $w^s{}_{ij}$ можно максимизировать соотношение (5.23):

$$\sum_i \sum_{j \neq i} w^s{}_{ij} x_{si} x_{sj} = \sum_i \sum_{j \neq i} (x_{si})^2 (xs_j)^2$$

где $w^s{}_{ij} = x^s{}_i\, x^s{}_j$. Таким образом, минимум энергетической функции E достигается при выборе следующего значения весового коэффициента

$$w^s{}_{ij} = x_{si} x_{sj}$$

Это справедливо для s-го образа, подлежащего запоминанию сетью Хопфилда. Для всех k образов, запоминаемых сетью, получаем

$$wi_j = \sum_s w^s{}_{ij} = \sum_s x_{si} x_{sj}$$

Пример 5.3.

Дана сеть из трех нейронов (рисунок 5.5):

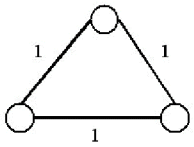

Рисунок 5.5 – Сеть Хопфилда с тремя нейронами и весами, равными 1.

Соответствующая весовая матрица имеет вид:

$$W = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

В качестве функции активации или выхода нейронов выберем знаковую функцию (*sign*) с нулевыми порогами. Пусть на вход такой сети подается вектор: $x = (1,-1,1)$.

Рассчитаем для него выходной вектор сети. При этом в соответствии с (5.3) – (5.7) получим:

$$O(1) = S(W x) = (-1,1,-1);$$
$$O(2) = S(Wo(1)) = (-1,-1,-1);$$
$$O(3) = S(Wo(2)) = (-1,-1,-1).$$

Так как $O(3) = O(2)$, то после 3-го шага выходы сети не изменятся, т.е. выходной вектор определяется сетью после 3-го шага.

Основная область применения сетей Хопфилда – распознавание образов. Например, каждое черно-белое изображение, представляемое пикселами, можно отобразить вектором $x = (x_1, \ldots, x_n)$, где x_i для i – го пиксела равен 1, если он черный, и $x_i = -1$, если – белый. При подаче на входы обученной сети Хопфилда искаженного изображения сеть после некоторого числа итераций выдает на выходы корректное изображение. На рисунке 5.6 приведены корректные образы, запомненные сетью, а на рисунке 5.7 – последовательность состояний сети Хопфилда при вводе искаженного образа. Важно подчеркнуть, что для искаженного изображения из четырех запомненных прототипов наиболее близким является второй прототип. После четвертой итерации сеть выдает корректный образ (второй прототип).

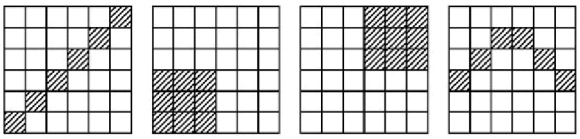

Рисунок 5.6 – Четыре видеоизображения для запоминания сетью Хопфилда

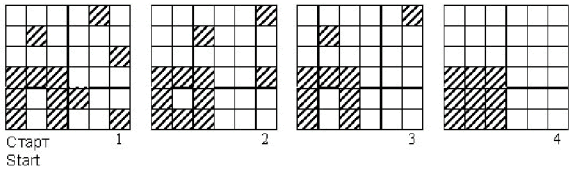

Рисунок 5.7 – Последовательность из четырех итераций по распознаванию искаженного изображения

Задание.

Обучите нейронную сеть Хопфилда распознанию первых 5 букв вашей фамилии. При распознании образа выедите шаги восстановления изображения аналогично изображениям 5.6 и 5.7