Рекуррентные сети

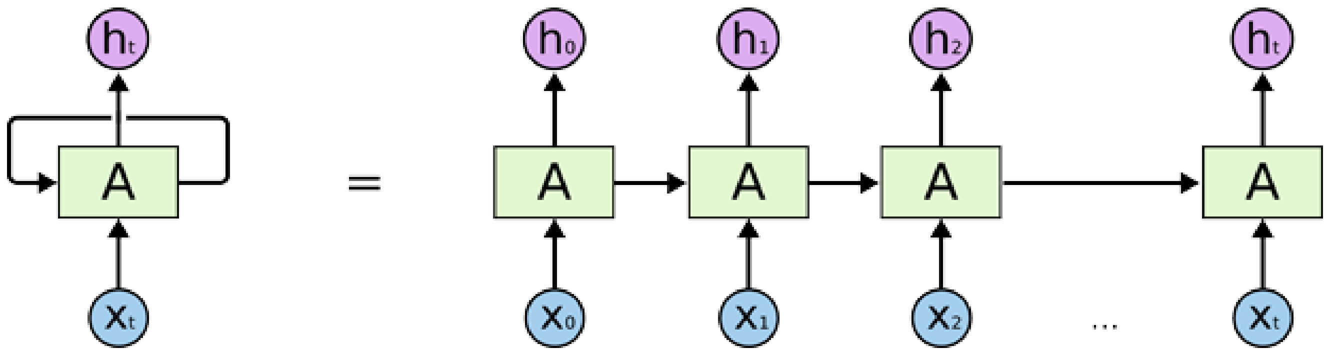
Что такое рекуррентные нейросети (РНС) и чем они отличаются от “обычных”, т.е. сетей прямого распространения?

У простого перцептрона, у сверточной нейросети есть общее ограничение. Такие сети с математической точки зрения ведут себя как обычные функции, хоть и очень сложно устроенные: у них есть заранее обозначенное число аргументов, а также обозначенный формат, в котором выдается ответ. Они не позволяют обрабатывать условно бесконечные последовательности, в которых важно не только содержание, но и порядок, в котором следует информация.

В рекуррентных нейросетях нейроны обмениваются информацией между собой: например, дополнительно к новому фрагменту входящих данных нейрон также получает некоторую информацию о предыдущем состоянии сети. Таким образом в сети реализуется «память», что принципиально меняет характер работы сети.

Рекуррентные сети

Схема однойслойной рекуррентной нейронной сети: на каждом цикле работы внутренний слой нейронов получает набор входных данных X и информацию о предыдущем состоянии внутреннего слоя A, на основании чего генерирует ответ h.



Если нейросети прямого распространения можно назвать «простой» функцией, то рекуррентные нейросети можно назвать программой. Память рекуррентных нейросетей (хотя и не полноценная) делает их Тьюринг-полными: при правильном задании весов нейросеть может успешно эмулировать работу компьютерных программ.

Рекуррентные сети

Схема однойслойной рекуррентной нейронной сети: на каждом цикле работы внутренний слой нейронов получает набор входных данных X и информацию о предыдущем состоянии внутреннего слоя A, на основании чего генерирует ответ h.

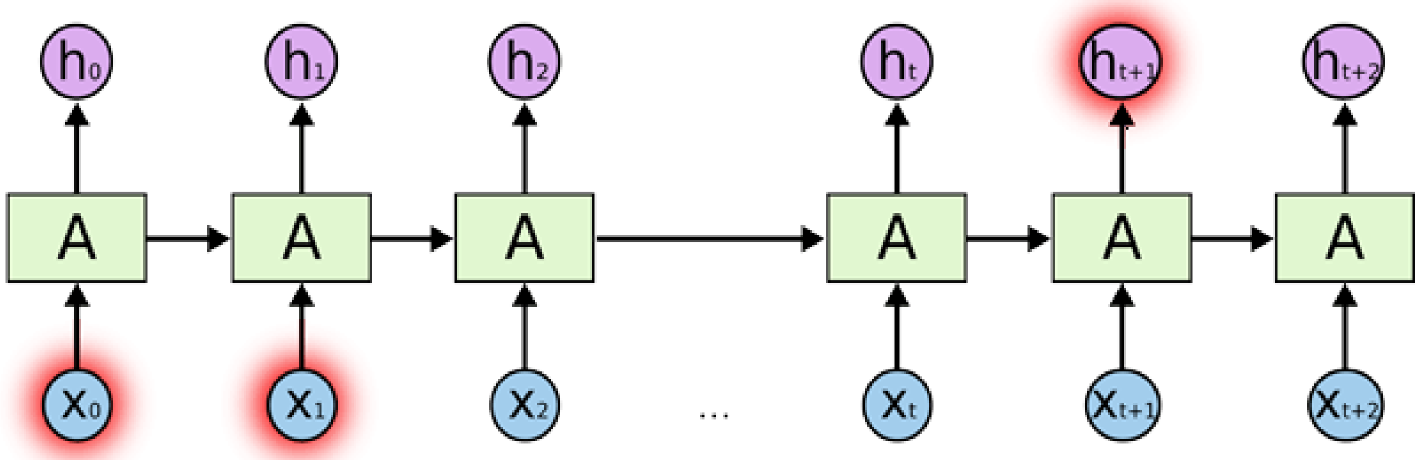
Как развивались рекуррентные сети?

Вероятно, первой РНС была сеть Хопфилда (впервые упомянута в 1974 году, окончательно оформилась в 1982-м), которая реализовывала на практике ячейку ассоциативной памяти. От современных РНС она отличается тем, что работает с последовательностями фиксированного размера.

Следующим шагом в эволюции РНС была «простая рекуррентная сеть» Джеффа Элмана, описанная в 1990 году. В ней автор затронул вопрос о том, как можно (и можно ли вообще) обучить нейросеть распознавать временные последовательности. Например, если есть входящие данные 1100 и 0110, можно ли их считать одним и тем же набором, сдвинутым во времени?

Рекуррентные сети

За простой РНС Элмана последовали все новые разработки, а в 1997 году Хохрейтер и Шмидхубер опубликовали статью «Long Short-term memory» («долгосрочная краткосрочная память»), заложившую основу для большинства современных РНС. В ней описывалась модификация РНС, решавшая проблему долгосрочной памяти простых сетей: их нейроны хорошо «помнят» недавно полученную информацию, но не имеют возможности надолго сохранить в памяти что-то, что обработали много циклов назад, какой бы важной та информация ни была.



Сети ассоциативной памяти

Ассоциативная память позволяет по неполной и даже частично недостоверной информации восстановить достаточно полное описание *знакомого* объекта. Слово знакомого является очень важным, поскольку *невозможно* вызвать ассоциации с незнакомыми объектами.

Решаемые ассоциативной памятью *задачи*:

- соотнести входную информацию со знакомыми объектами, и дополнить ее до точного описания объекта;
- отфильтровать из входной информации недостоверную, а на основании оставшейся решить первую задачу.

Сети ассоциативной памяти

Постановка задачи. Пусть задан набор из m эталонов - n - мерных векторов $\{x^i\}$. Требуется построить сеть, которая при предъявлении на вход произвольного образа - вектора x - давала бы на вы ходе «наиболее похожий» эталон.

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. С другой - сеть фактически просто запоминает образцы до того, как на ее вход поступают реальные данные, и не может изменять свое поведение.

Сети ассоциативной памяти

Сеть Хопфилда

Американским ученым Хопфилдом в 80-х годах предложен специальный тип нейронных сетей. В отличие от сетей с прямыми связями, сети Хопфилда являются рекуррентными или сетями с обратными связями.

Нейронная сеть Хопфилда - полносвязная НС симметричной матрицей связей. В процессе работы динамика таких сетей сходится (конвергирует) к одному из положений равновесия. Эти положения равновесия являются локальными минимумами функционала, называемого *энергией сети.*

Такая сеть может быть использована:

- как автоассоциативная память;
- как фильтр;
- для решения некоторых задач оптимизации.

Сети ассоциативной памяти

Сеть Хопфилда

Алгоритм обучения сети Хопфилда существенно отличается от классических алгоритмов обучения персептронов. К этим алгоритмам относятся: *метод коррекции ошибки или метод обратного распространения ошибки*.

Отличие заключается в том, что вместо последовательного приближения к нужному состоянию с вычислением ошибок, все коэффициенты матрицы рассчитываются по одной формуле, за один цикл, после чего сеть сразу готова к работе.

Вычисление коэффициентов основано на следующем правиле: для всех запомненных образов Xi матрица связи должна удовлетворять уравнению $\mathbf{X}_i^T = \mathbf{W}\mathbf{X}_i$, поскольку при этом условии состояния сети Xi будут устойчивы - попав в такое состояние, сеть в нём и останется.

Отличительная особенность фильтров - матрица весовых коэффициентов настраивается детерминированным алгоритмом раз и навсегда, весовые коэффициенты затем не изменяются.

Сети ассоциативной памяти

Сеть Хопфилда

Алгоритм обучения сети Хопфилда существенно отличается от классических алгоритмов обучения персептронов. К этим алгоритмам относятся: *метод коррекции ошибки или метод обратного распространения ошибки*.

В сети Хопфилда есть обратные связи и поэтому нужно решать проблему устойчивости. Веса между нейронами в сети Хопфилда мо- гут рассматриваться в виде матрицы взаимодействий **W**. Сеть с обратными связями является устойчивой, если ее матрица симметрична и имеет нули на главной диагонали.

Условие симметричности является необходимым, но не достаточным. Только асинхронный режим работы сети гарантирует достижение устойчивого состояния сети, в синхронном случае возможно бесконечное переключение между двумя разными состояниями.

Сети ассоциативной памяти

Сеть Хопфилда

Запоминаемые векторы должны иметь бинарный вид. После этого происходит расчет весовых коэффициентов по формуле:

$$w_{ij} = \frac{1}{N} \sum_{d=1..m} X_{id} X_{jd}$$

где N - размерность векторов, *m* - число запоминаемых выходных векторов, *d* - номер запоминаемого выходного вектора, *Xij* - i-я компонента запоминаемого выходного j-го вектора.

Весовая матрица **W** может быть найдена вычислением внешнего произведения каждого запоминаемого вектора с самим собой и суммированием полученных матриц:

$\mathbf{W} = \frac{1}{N} \sum_i \mathbf{X}_i^T \mathbf{X}_i$, **X*i***-й запоминаемый вектор-строка.

Расчет этих весовых коэффициентов и называется обучением сети. Как только веса заданы, сеть может быть использована для получения запомненного выходного вектора по данному входному вектору, который может быть частично неправильным или неполным.

Сети ассоциативной памяти

Сеть Хопфилда

Сети Хопфилда обладают следующими свойствами:

- Симметрия дуг: сети содержат n нейронов, соединенных друг с другом. Каждая дуга (соединение) характеризуется весом w_{ij}, причем: $\forall i, j \in N : i \neq j : \exists_1 w_{ij}$,

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

- 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}. Для структуры сети это ,езразлично. Однако? формулы для распознавания образов при использовании значений -1 и 1 для входов и выходов нейронов сети Хопфилда получаются нагляднее, поэтому эти значения и предполагаются ниже.

Сети ассоциативной памяти

Сеть Хопфилда

Определение. Бинарная НС Хопфилда определяется:

- симметричной матрицей с нулевыми диагональными элементами;
- вектором **T** порогов нейронов;
- знаковой функцией активации или выхода нейронов.

Определение. Каждый вектор **O** с компонентами -1 или 1, удовлетворяющий уравнению

$$\mathbf{O} = \mathbf{f}(\mathbf{WO} - \mathbf{T}),$$

называется образом для сети Хопфилда.

Сети ассоциативной памяти

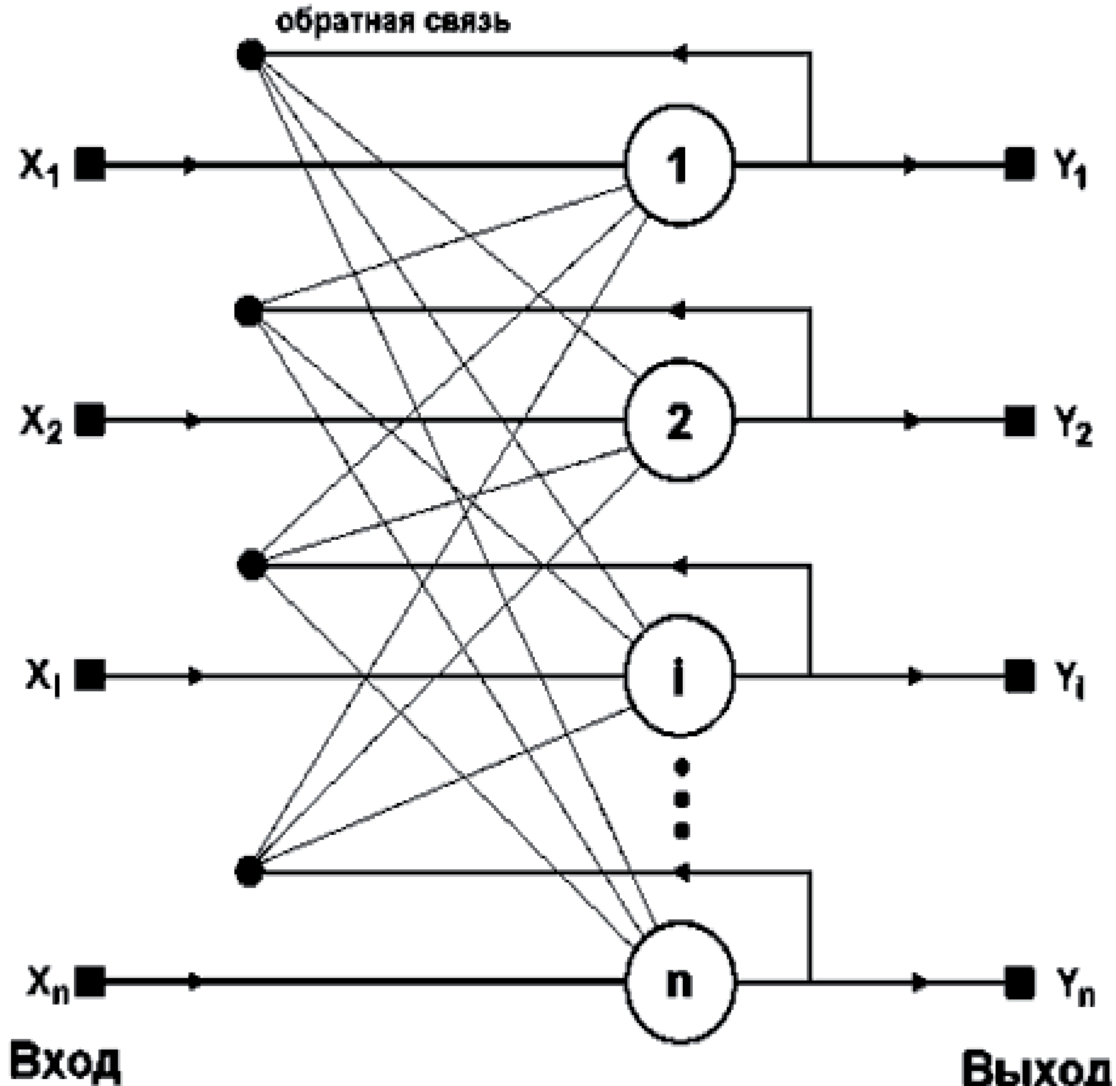
Сеть Хопфилда

Сети Хопфилда функционируют следующим образом. На входы сети подается некоторый образ. Он остается на входах до тех пор, пока не завершатся изменения состояний сети. В этом случае говорят о сходимости сети. Наглядно это можно представить следующим образом. В начале сеть находится на высоком энергетическом уровне, из которого возможны переходы в различные состояния. Затем энергетический уровень нейронной сети уменьшается до тех пор, пока не достигается некоторое конечное состояние.

Класс сетей Хопфилда содержит только один слой нейронов, причем каждый нейрон соединен с остальными. Обратные связи с вы- хода нейрона на его же вход отсутствуют.

Сети ассоциативной памяти

Сеть Хопфилда



Сети ассоциативной памяти

Сеть Хопфилда

«Энергетический» подход при описании процесса обучения сети.

При построении функции энергии исходят из следующих положений:

1. *Каждый эталон должен быть точкой минимума.*
2. *В точке минимума все координаты образа должны иметь значения* ± 1.

Функция $H = -\frac{1}{2}\sum_{i=1}^{m}\left(\mathbf{x},\mathbf{x}^i\right)^2 + \alpha\sum_{j=1}^{n}\left(x_j^2 - 1\right)^2$ не удовлетворяет этим требованиям строго, но можно предполагать, что первое слагаемое обеспечит притяжение к эталонам, а второе слагаемое - приблизит к единице абсолютные величины всех координат точки минимума. Величина α характеризует соотношение между этими двумя требованиями и может меняться со временем.

Сети ассоциативной памяти

Сеть Хопфилда

Используя выражение для энергии, можно записать систему уравнений, описывающих функционирование сети Хопфилда:

$$\dot{x}_j = -\partial H / \partial x_j = \sum_{i=1}^{m} \left(\mathbf{x}, \mathbf{x}^i\right) x_j^i - 4\alpha\left(x_j^2 - 1\right) x_j \qquad (1)$$

Сеть Хопфилда в виде (1) является сетью с непрерывным временем. Для реализации на компьютерах лучше воспользоваться сетями, функционирующими в дискретном времени.

Сеть Хопфилда должна осуществлять преобразование входного вектора **x** так, чтобы выходной вектор **x**’ был ближе к тому эталону, который является правильным ответом. Преобразование сети будем искать в следующем виде:

$$x' = \mathrm{Sign}\left(\sum_{i=1}^{m} w_i x^i\right)$$

где w_i - вес i -го эталона, характеризующий его близость к вектору **x**, Sign - нелинейный оператор, переводящий вектор с координатами y_i в вектор с координатами sign y_i .

Сети ассоциативной памяти

Сеть Хопфилда

Сеть работает следующим образом:

1. *На вход НС подается образ* $\mathbf{x}$ *, а на выходе снимается образ* $\mathbf{x}'$.
2. *Если* $\mathbf{x}' \neq \mathbf{x}$, *то полагаем* $\mathbf{x} = \mathbf{x}'$ *и возвращаемся к шагу* 1.
3. *Полученный вектор* $\mathbf{x}'$ *является ответом.*
4. Таким образом, ответ всегда является неподвижной точкой преобразования сети (2) и именно это условие и является условием остановки.

Пусть i* - номер эталона, ближайшего к образу $\mathbf{x}$. Тогда, если выбрать веса пропорционально близости эталонов к исходному образу $\mathbf{x}$, то следует ожидать, что $\mathbf{x}'$ будет ближе к эталону $\mathbf{x}^{i^*}$, чем $\mathbf{x}$, а после нескольких итераций он станет совпадать с $\mathbf{x}^{i^*}$.

Наиболее простой сетью вида (1) является дискретный вариант сети Хопфилда с весами равными скалярному произведению эталонов на предъявляемый образ: $\mathbf{x}' = \mathrm{Sign}\left(\sum_{i=1}^{m}(\mathbf{x}, \mathbf{x}^i)\mathbf{x}^i\right)$

Сети ассоциативной памяти

Сеть Хопфилда

Рассмотри задачу сохранения образца $x = [1, -1, 1, 1]$. На первом этапе необходимо определить весовые значения сети Хопфилда, а далее убедиться, что сеть сможет восстановить сохраненный образец при подаче на вход искаженного вектора $x_{test} = [-1, -1, 1, 1]$.

Весовая матрица определяется по формуле $W = x^T x$.

$$W = \begin{bmatrix} 1 \\ -1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 1 & 1 \\ -1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 1 \\ 1 & -1 & 1 & 1 \end{bmatrix}$$

После обнуления диагональных элементов матрицы:

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

На вход подадим вектор x_{test} . Нейроны поведут себя следующим образом.

Сети ассоциативной памяти

Сеть Хопфилда

Комбинированный ввод первого нейрона будет равен:

$$net_1 = (-1) * 0 + (-1) * (-1) + 1 * 1 + 1 * 1 = 3$$

Поскольку $net_1 > 0$, то нейрон 1 будет находиться в состоянии +1.

Для второго нейрона: $net_2 = (-1) * (-1) + (-1) * 0 + 1 * (-1) + 1 * (-1) = -1$

Нейрон 2 окажется в состоянии -1.

Нейрон 3 – +1: $net_3 = (-1) * 1 + (-1) * (-1) + 1 * 0 + 1 * 1 = 1$

Нейрон 4 – +1: $net_4 = (-1) * 1 + (-1) * (-1) + 1 * 1 + 1 * 0 = 1$

Таким образом, мы получили такой вектор $[\mathbf{1}, -\mathbf{1}, \mathbf{1}, \mathbf{1}]$, что соответствует тому образцу, который мы запоминали в первой части примера. Если теперь еще раз произвести расчет состояний нейронов в соответствии с теми состояниями, которые они приняли после прохождения вектора x_{test} по взвешенным связям, то мы увидим, что состояния не изменится, то есть сеть оказалась в устойчивом положении.

Сети ассоциативной памяти

Сеть Хопфилда

О сетях Хопфилда известно, что они способны запомнить и точно воспроизвести «порядка 0.14n слабо скоррелированных образов».

В этом высказывании содержится два ограничения:

- число эталонов не превосходит 0.14п .
- эталоны слабо коррелированны.

Наиболее существенным является второе ограничение, поскольку образы, которые сеть должна обрабатывать, часто очень похожи.

***Синхронный режим работы сети*.** Если работа сети моделируется на одном процессоре, то при синхронном режиме последовательно просматриваются нейроны, однако их состояния запоминаются отдельно и не меняются до тех пор, пока не будут пройдены все нейроны сети. Когда все нейроны просмотрены, их состояния синхронно меняются на новые.

Сети ассоциативной памяти

Сеть Хопфилда

При реально параллельном моделировании, этот режим фактически означает, что время передачи τ_{ij} для каждой связи между элементами u_i и u_j одинаково для каждой связи, что приводит к параллельной работе всех связей, они одновременно меняют свои состояния, основываясь только на предыдущем моменте времени.

При синхронном режиме возможно бесконечное чередование двух состояний с разной энергией, т.е. динамический аттрактор.

Асинхронный режим работы сети. В асинхронном режиме работы состояния нейронов в следующий момент времени меняются последовательно вычисляется локальное поле для первого нейрона в момент t, определяется его реакция, и нейрон устанавливается в новое состояние (которое соответствует его выходу в момент $t+1$), потом вычисляется локальное поле для второго нейрона с учётом нового состояния первого, меняется состояние второго нейрона, и так далее.

Сети ассоциативной памяти

Сеть Хопфилда

Таким образом, состояние каждого следующего нейрона вычисляется с учетом всех изменений состояний рассмотренных ранее нейронов.

В асинхронном режиме невозможен динамический аттрактор: вне зависимости от количества запомненных образов и начального состояния сеть непременно придёт к устойчивому состоянию (статическому аттрактору).

Определить, устойчиво или нет состояние нейрона, можно на основании искусственной энергии нейрона в данном поле $E_i = -s_i h_i$. Если знак выхода (+1 или -1) нейрона совпадает с направлением локального поля ($E_i < 0$), то его положение энергетически устойчиво и в следующий момент времени состояние нейрона остаётся неизменным. В противном случае ($E_i > 0$) положение нейрона неустойчиво и он меняет свой знак, переходя в состояние $s_i(t+1) = -s_i(t)$ с энергией $E_i(t+1) \le E_i(t)$.

Сети ассоциативной памяти

Сеть Хопфилда

Устойчивость при асинхроном способе достигается потому, что выполняется условие на общую энергию $E_i(t+1) \le E_i(t)$. В синхронном случае условие несколько изменяется, а именно: $E(t+1) \le E(t-1)$. В ситуации, когда происходят бесконечные циклические переходы, энергия двух разных состояний соответственно равна E(t) и E(t+1). При этом состояния t+1 и t-1, а также t и t+2 - совпадают. Если образуется такое состояние, то его называется динамическим аттрактором.

Если же совпадают состояния t и t + 1, аттрактор называют статическим.

В большинстве случаев динамические аттракторы являются нежелательными, так как не соответствуют какому-либо определённому ответу сети.

Сети ассоциативной памяти

Сеть Хопфилда

Пример. Зависимость работы сети Хопфилда от степени коррелированности образов можно продемонстрировать на примере.

Пусть даны три эталона x1 , x2 , x3 таких, что

$$\begin{aligned}&(x^1,x^2)+(x^1,x^3)>n,\\&(x^1,x^2)+(x^2,x^3)>n,\\&(x^1,x^3)+(x^2,x^3)>n,\\&(x^i,x^j)>0\ (\forall i,j).\end{aligned}\qquad(1)$$

Для любой координаты существует одна из четырех возможностей:

$$\begin{aligned}&1)\ x_l^i = x_l^j = 1\ (\forall i,j),\\&2)\ x_l^i = -x_l^j = x_l^k = 1\,,\\&3)\ x_l^i = -x_l^j = x_l^k = -1\,,\\&4)\ x_l^i = x_l^j = -1\ \ (\forall i,j)\end{aligned}\qquad(2)$$

Сети ассоциативной памяти

Сеть Хопфилда

В первом случае при предъявлении сети q -го эталона в силу формулы (1) получаем $x_l' = \mathrm{Sign}\left(\sum_{i=1}^{3}\left(x^q, x^i\right)\times 1\right) = 1$ так как все скалярные произведения положительны по условию (2). Аналогично получаем в четвертом случае $x_j' = -1$.

Во втором случае рассмотрим отдельно три варианта

$$x = x^i,\ x_j' = \mathrm{Sign}\left(-\left(x^i, x^i\right) + \left(x^i, x^j\right) + \left(x^i, x^k\right)\right) = 1$$

$$x = x^j,\ x_l' = \mathrm{Sign}\left(-\left(x^j, x^i\right) + \left(x^j, x^j\right) + \left(x^j, x^k\right)\right) = 1$$

$$x = x^k,\ x_l' = \mathrm{Sign}\left(-\left(x^k, x^i\right) + \left(x^k, x^j\right) + \left(x^k, x^k\right)\right) = 1$$

так как скалярный квадрат любого образа равен n, а сумма двух любых скалярных произведений эталонов больше n , по условию (2). Таким образом, независимо от предъявленного эталона получаем $x_j' = 1$. Аналогично в третьем случае получаем $x_j' = -1$.

Окончательный вывод таков: если эталоны удовлетворяют условиям (2), то при предъявлении любого эталона на выходе всегда будет один образ.

Сети ассоциативной памяти

Сеть Хопфилда

Сеть Хопфилда можно отнести к автоассоциативной памяти, т.е. такой, которая может завершить или исправить образ, но не может ассоциировать полученный образ с другим образом. Чтобы организовать устойчивую автоассоциативную память с помощью НС с обратными связями, веса нужно выбирать так, чтобы возникали энергетические минимумы в нужных вершинах единичного гиперкуба.

Рассмотрим подробнее функционирование сети в режиме формирования ассоциативной памяти.

Требование, детализирующие понятие «слабо коррелированных образов»: lля правильного распознавания всех эталонов достаточно (но не необходимо) потребовать, чтобы выполнялось

.

$$\sum_{\substack{i=1 \\ i\neq j}}^{m} \left(x^i, x^j\right) < n, \forall j$$

Более сильное условие можно записать в виде $\left(x^i, x^j\right) < \frac{n}{m}, \forall i \neq j$

Сети ассоциативной памяти

Сеть Хопфилда

Задача, решаемая сетью Хопфилда в качестве ассоциативной памяти, как правило, формулируется следующим образом.

Известен некоторый набор двоичных сигналов, 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.

Динамические НС Хопфилда получили два основных применения:

1. Сеть используется как ассоциативная память, так как она способна запоминать, а затем восстанавливать даже при неполной входной информации различные векторы $\mathbf{r} \in \mathbf{R}^n$. Эта способность обеспечивается устойчивостью достигнутого состояния после настройки сети.

Сети ассоциативной памяти

Сеть Хопфилда

Вектор выхода обученной сети $\mathbf{q} \in \mathbf{R}^{N_к}$ рассматривается как некоторый хранимый эквивалентный образ. Любой входной вектор $\tilde{\mathbf{r}} \in \mathbf{R}^{N}$ в классе $\mathbf{r} \in \mathbf{R}^{N}$ может быть использован как указатель для восстановления вектора $\mathbf{q} \in \mathbf{R}^{N_к}$.

Известно, что сеть с n бинарными базовыми элементами может хранить макси- мальное число случайных n-битных векторов, равное 0.15n , а сеть должна иметь 2n связей между базовыми элементами.

2. Поскольку сеть Хопфилда при обучении минимизирует некоторый функционал, то она может быть использована для решения комбинаторных задач оптимизации. Основная проблема при таком использовании сети состоит в формулировании исходной оптимизационной задачи в терминах обучения сети.

Сети ассоциативной памяти

Сеть Хопфилда

Ограничения сети Хопфилда. Нейронная сеть Хопфилда имеет ряд недостатков.

1. Относительно небольшой объем памяти, величину которого можно оценить выражением: $M = N\left(2\log_2 N\right)^{-1}$.

Попытка записи большего числа образов приводит к тому, что нейронная сеть перестаёт их распознавать.

2. Достижение устойчивого состояния не гарантирует правиль- ный ответ НС. Это происходит из-за того, что сеть может сой- тись к так называемым ложным аттракторам, которые иногда на- зывают «химерами», которые обычно склеены из фрагментов различных образов.

Сети ассоциативной памяти

Сеть Хэмминга

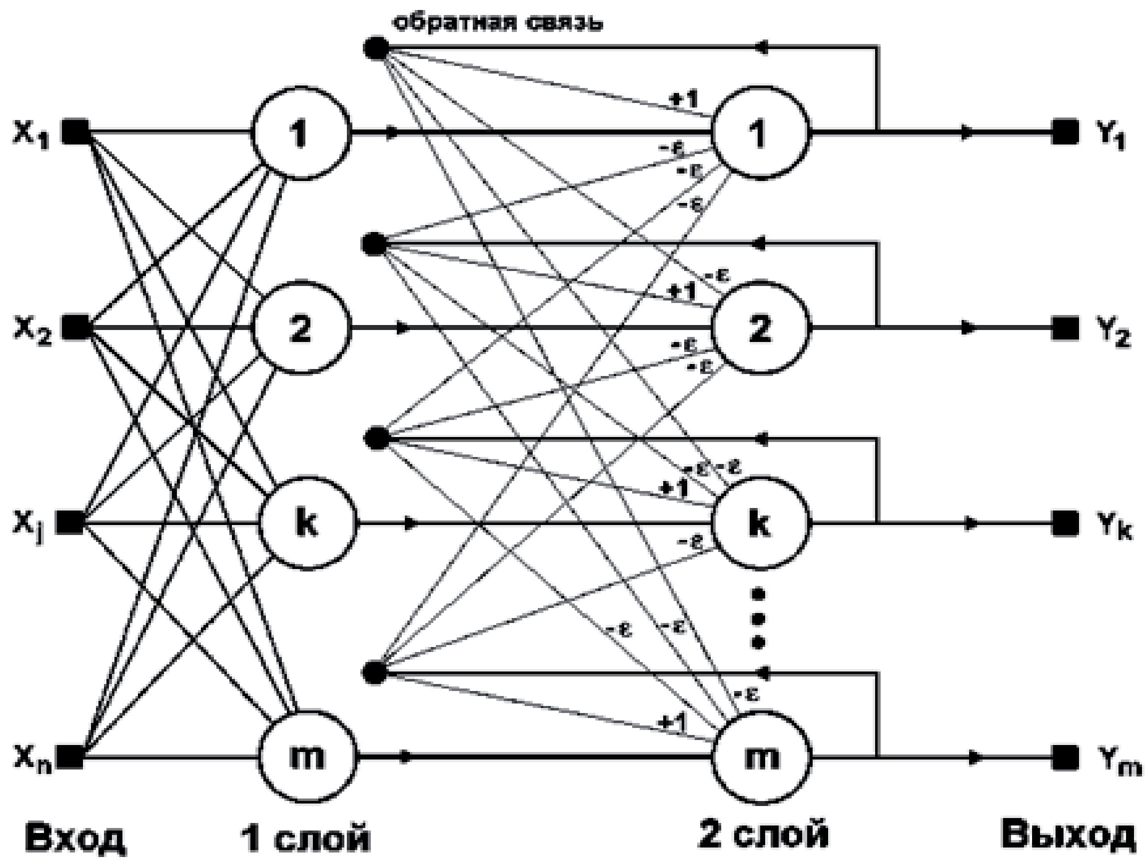
Сеть Хемминга - расширение сети Хопфилда. Эта сеть была разработана Ричардом Липпманом в середине 80-х годов. Сеть Хемминга реализует классификатор, базирующийся на наименьшей погрешности для векторов двоичных входов, где погрешность определяется расстоянием Хемминга.

В условиях, когда нет необходимости в явном виде выдавать образец, а достаточно получать номер образца, ассоциативную память успешно реализует сеть Хэмминга. Она характеризуется, по сравнению с сетью Хопфилда, меньшими затратами на память и объемом вычислений.

Сети ассоциативной памяти

Сеть Хэмминга

Структурная схема сети Хэмминга



Сети ассоциативной памяти

Сеть Хэмминга

Сеть состоит из двух слоев. Первый и второй слои имеют по m нейронов, где m - число образцов. Нейроны первого слоя имеют по n синапсов, соединенных со входами сети (образующими фиктивный нулевой слой). Нейроны второго слоя связаны между собой ингибиторными (отрицательными обратными) синаптическими связями. Единственный синапс с положительной обратной связью для каждого нейрона соединен с его же аксоном.

Сеть обладает простой архитектурой прямого распространения с входным уровнем, полностью подсоединенным к слою категорий. Каждый элемент обработки в слое категорий является обратно подсоединенным к каждому нейрону в том же самом слое и прямо подсоединенным к выходному нейрону. Выход из слоя категорий к выходному слою формируется через конкуренцию.

Сети ассоциативной памяти

Сеть Хэмминга

Обучение сети Хемминга похоже на методологию Хопфилда. На входной слой поступает желаемый обучающий образ, а на выход выходного слоя поступает значение желаемого класса, к которому принадлежит вектор. Выход содержит лишь значение класса к которому принадлежит входной вектор. Рекурсивный характер слоя Хопфилда обеспечивает средства коррекции всех весов соединений.

Идея работы сети состоит в нахождении расстояния Хэмминга от тестируемого образа до всех образцов. Расстоянием Хэмминга называется число отличающихся битов в двух бинарных векторах.

Сеть Хэмминга должна выбрать образец с минимальным расстоянием Хэмминга до неизвестного входного сигнала, в результате чего будет активизирован только один выход сети, соответствующий этому образцу.

Сети ассоциативной памяти

Сеть Хэмминга

Алгоритм функционирования сети Хэмминга:

1. На стадии инициализации весовым коэффициентам первого слоя и порогу активационной функции присваиваются следующие значения: $w_{ik} = \frac{x_i^k}{2},\ i = \overline{0, n-1},\ k = \overline{0, m-1}$, $\quad T_k = \frac{n}{2},\ k = \overline{0, m-1}.$

   x_i^k - i-ый элемент k-ого образца.

Весовые коэффициенты тормозящих синапсов во втором слое берут равными $0 < \varepsilon < \frac{1}{m}$. Синапс нейрона, связанный с его же аксоном, имеет вес +1.

2. На входы нейронной сети подается неизвестный вектор $\mathbf{X} = \{x_i\},\ i = \overline{0, n-1}$, исходя из которого рассчитываются состояния нейронов первого слоя:

$$y_j^{(1)} = s_j^{(1)} = \sum_{i=0}^{n-1} w_{ij} x_i + T_j,\ \ j = \overline{0, m-1}$$

После этого полученными значениями инициализируются значения аксонов второго слоя: $y_j^{(2)} = y_j^{(1)},$

Сети ассоциативной памяти

Сеть Хэмминга

3. Вычислить новые состояния нейронов второго слоя:

$$s_j^{(2)}(p+1) = y_j(p) - \varepsilon \sum_{k=0}^{m-1} y_k^{(2)}(p), k \neq j, \quad j = \overline{0, m-1}$$

и значения их аксонов:

$$y_j^{(2)}(p+1) = f\left[s_j^{(2)}(p+1)\right], \quad j = \overline{0, m-1}$$

Активационная функция f имеет вид порога, причем величина F должна быть достаточно большой, чтобы любые возможные значения аргумента не приводили к насыщению.

4. Проверить, изменились ли выходы нейронов второго слоя за последнюю итерацию. Если да - перейди к шагу 2. Иначе - конец.

Из оценки алгоритма видно, что роль первого слоя весьма условна: воспользовавшись один раз на шаге 1 значениями его весовых коэффициентов, сеть больше не обращается к нему, поэтому первый слой может быть заменен на матрицу весовых коэффициентов.

Сети ассоциативной памяти

Сеть Хэмминга

Сеть Хемминга имеет ряд преимуществ над сетью Хопфилда:

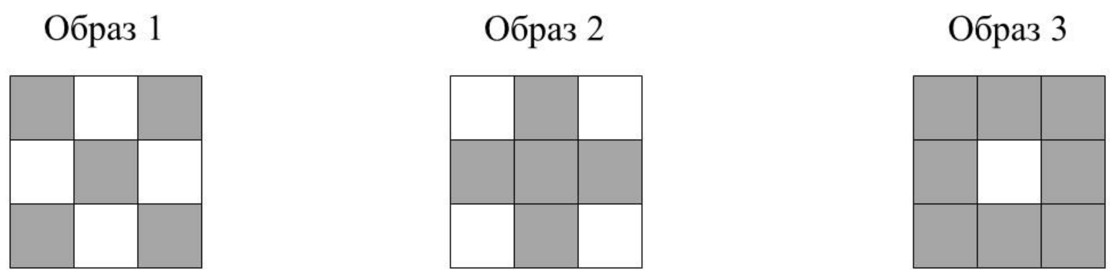
- она способна найти минимальную погрешность, если погрешности входных бит являются случайными и независимыми;
- для функционирования сети Хемминга нужно меньшее количество нейронов, поскольку средний слой требует лишь один нейрон на класс, вместо нейрона на каждый входной узел;
- nакже сеть Хемминга не страдает от неправильных классификаций, которые могут случиться в сети Хопфилда.

В целом, сеть Хемминга быстрее и точнее, чем сеть Хопфилда.

Сети ассоциативной памяти

Сеть Хэмминга

Пример использования искусственной нейронной сети Хэмминга для классификации по трем эталонным графическим образам, представленным бинарными элементами в квадратной матрице (рис.). Таким образом, входной вектор содержит 9 элементов, выходной – 3 элемента.

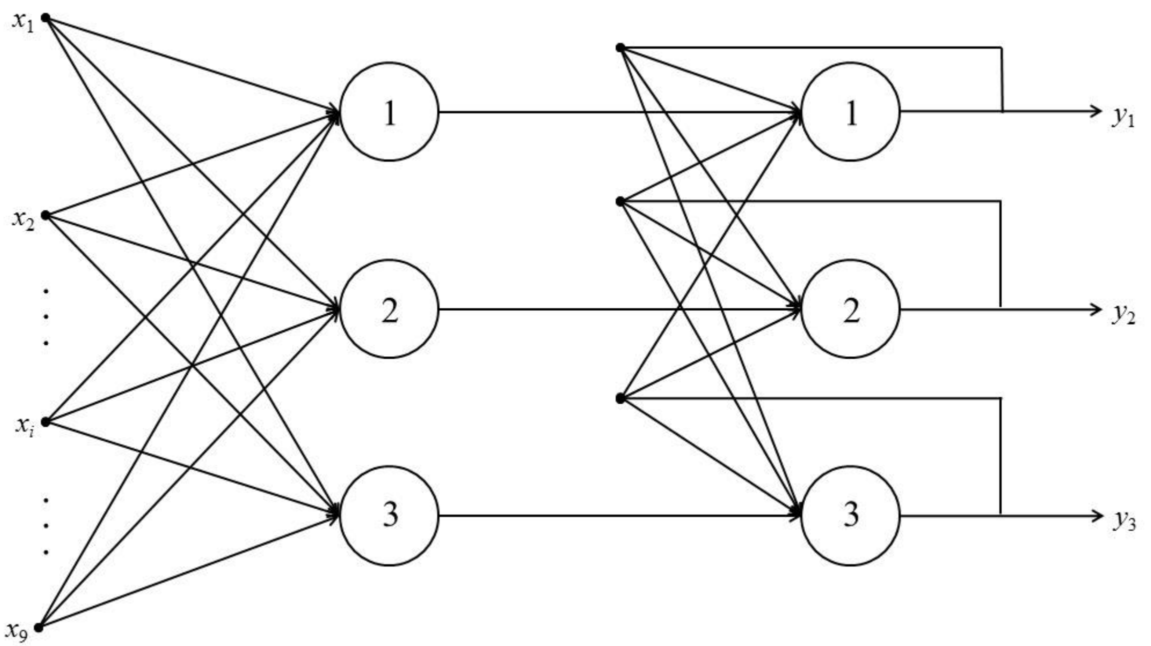


Наличие рисунка в любом из 9 элементов кодируется значением 1, отсутствие – значением –1.

Сети ассоциативной памяти

Сеть Хэмминга

Создаваемая нейронная сеть будет включать 9 входных переменных и по 3 нейрона в первом и втором (выходном) слоях



Сети ассоциативной памяти

Сеть Хэмминга

Сформируем матрицу эталонных образов:

№ образа	№ входной переменной								
	1	2	3	4	5	6	7	8	9
1	1	–1	1	–1	1	–1	1	–1	1
2	–1	1	–1	1	1	1	–1	1	–1
3	1	1	1	1	–1	1	1	1	1

Рассчитаем весовые коэффициенты нейронов первого слоя:

№ нейрона первого слоя	№ входной переменной								
	1	2	3	4	5	6	7	8	9
1	0,5	–0,5	0,5	–0,5	0,5	–0,5	0,5	–0,5	0,5
2	–0,5	0,5	–0,5	0,5	0,5	0,5	–0,5	0,5	–0,5
3	0,5	0,5	0,5	0,5	–0,5	0,5	0,5	0,5	0,5

Сети ассоциативной памяти

Сеть Хэмминга

По формуле T=n/2 определяем порог активационной функции T = 4,5. С учетом ограничения $\varepsilon \in \left(0, \frac{1}{3}\right]$ абсолютное значение веса каждого ингибиторного синапса примем ε = 0,3. Зададимся предельной величиной критерия стабилизации выходного вектора Emax = 0,1.

Составим матрицу весовых коэффициентов обратных синапсов на основе выражения для определения значений синапсов обратных связей нейронов второго слоя

$$\varepsilon_{jp} = \begin{cases} 1, j = p; \\ -\varepsilon, j \neq p, \end{cases}$$

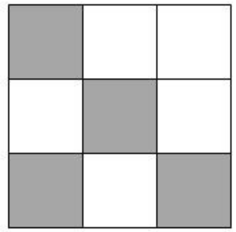
№ выходного нейрона	**№ нейрона первого слоя**		
	1	**2**	**3**
1	**1,0**	**–0,3**	**–0,3**
2	**–0,3**	**1,0**	**–0,3**
3	**–0,3**	**–0,3**	**1,0**

Сети ассоциативной памяти

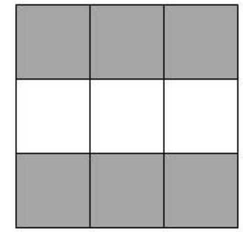
Сеть Хэмминга

Для тестирования настроенной сети используем два зашумленных графических образа:

Образ 4



Образ 5



Подаем на входы сети бинарный вектор, соответствующий образу 4: $\vec{x}^{\mathrm{T}} = [1, -1, -1, -1, 1, -1, 1, -1, 1]$. Применение расчетного соотношения дает вектор-столбец состояний нейронов первого слоя, а активационной функции (линейная активационная функция f(s))

$$f(s) = \begin{cases} 0, s \le 0; \\ s, 0 < s \le T; \\ T, s \ge T; \end{cases}$$

к состояниям – вектор-столбец выходных значений нейронов первого слоя: $\vec{s}_1 = \begin{bmatrix} 8{,}00 \\ 2{,}00 \\ 3{,}00 \end{bmatrix}$ $\vec{y}_1 = \begin{bmatrix} 4{,}50 \\ 2{,}00 \\ 3{,}00 \end{bmatrix}$

Выходам нейронной сети присваиваются соответствующие выходные значения нейронов первого слоя.

Сети ассоциативной памяти

Сеть Хэмминга

Далее итерационно рассчитывается серия выходных векторов (расчеты выполняются аналогично) до выполнения условия стабилизации.

Сигналы нейронной сети Хэмминга, получаемые на протяжении полного цикла расчета при подаче тестового образа 4:

Номер итерации	Вектор состояний			Вектор выходов			
	$s_{21}^{(q)}$	$s_{22}^{(q)}$	$s_{23}^{(q)}$	$y_{21}^{(q)}$	$y_{22}^{(q)}$	$y_{23}^{(q)}$	
1	8,00	2,00	3,00	4,50	2,00	3,00	–
2	3,00	–0,25	1,05	3,00	0,00	1,05	10,05
3	2,69	–1,22	0,15	2,69	0,00	0,15	0,91
4	2,64	–0,85	–0,66	2,64	0,00	0,00	0,02

Критерий остановки цикла возврата сигнала по обратным связям выполнен после 4-й итерации. Положительное выходное значение 1-го нейрона указывает на то, что зашумленный входной образ следует отнести к 1-му классу.

Сети ассоциативной памяти

Сеть Хэмминга

Результаты расчетов изменения сигналов нейронной сети для для тестового образа 5 с бинарным входным вектором вида:

$$\vec{x}^{\mathrm{T}} = [1, 1, 1, -1, -1, -1, 1, 1, 1]$$.

Номер итерации	**Вектор состояний**			**Вектор выходов**			
	$s_{21}^{(q)}$	$s_{22}^{(q)}$	$s_{23}^{(q)}$	$y_{21}^{(q)}$	$y_{22}^{(q)}$	$y_{23}^{(q)}$	
1	**6,00**	**2,00**	**7,00**	**4,50**	**2,00**	**4,50**	**–**
2	**2,55**	**–0,70**	**2,55**	**2,55**	**0,00**	**2,55**	**11,61**
3	**1,79**	**–1,53**	**1,79**	**1,79**	**0,00**	**1,79**	**1,17**
4	**1,25**	**–1,07**	**1,25**	**1,25**	**0,00**	**1,25**	**0,57**
5	**0,87**	**–0,75**	**0,87**	**0,87**	**0,00**	**0,87**	**0,28**
6	**0,61**	**–0,52**	**0,61**	**0,61**	**0,00**	**0,61**	**0,14**
7	**0,43**	**–0,37**	**0,43**	**0,43**	**0,00**	**0,43**	**0,07**

В этом случае критерий остановки был выполнен после 7-й итерации, однако уже на 2-й итерации стало понятно, что сеть не может отдать предпочтение 1-му и 3-му классам при отнесении входного зашумленного образа 5. В условиях малого количества входных характеристик следует сделать вывод, скорее, о том, что сеть вовсе не смогла классифицировать образ, чем о том, что она в равной степени отнесла его к двум классам.

Сети ассоциативной памяти

Двунаправленная ассоциативная память

Эта сетевая модель была разработана Бартом Козко и расширяет модель Хопфилда. Множество парных образов учится за образами, которые представлены как биполярные векторы. Подобно сети Хопфилда, когда представляется зашумленная версия одного образа, определяется ближайший образ, ассоциированный с ним.

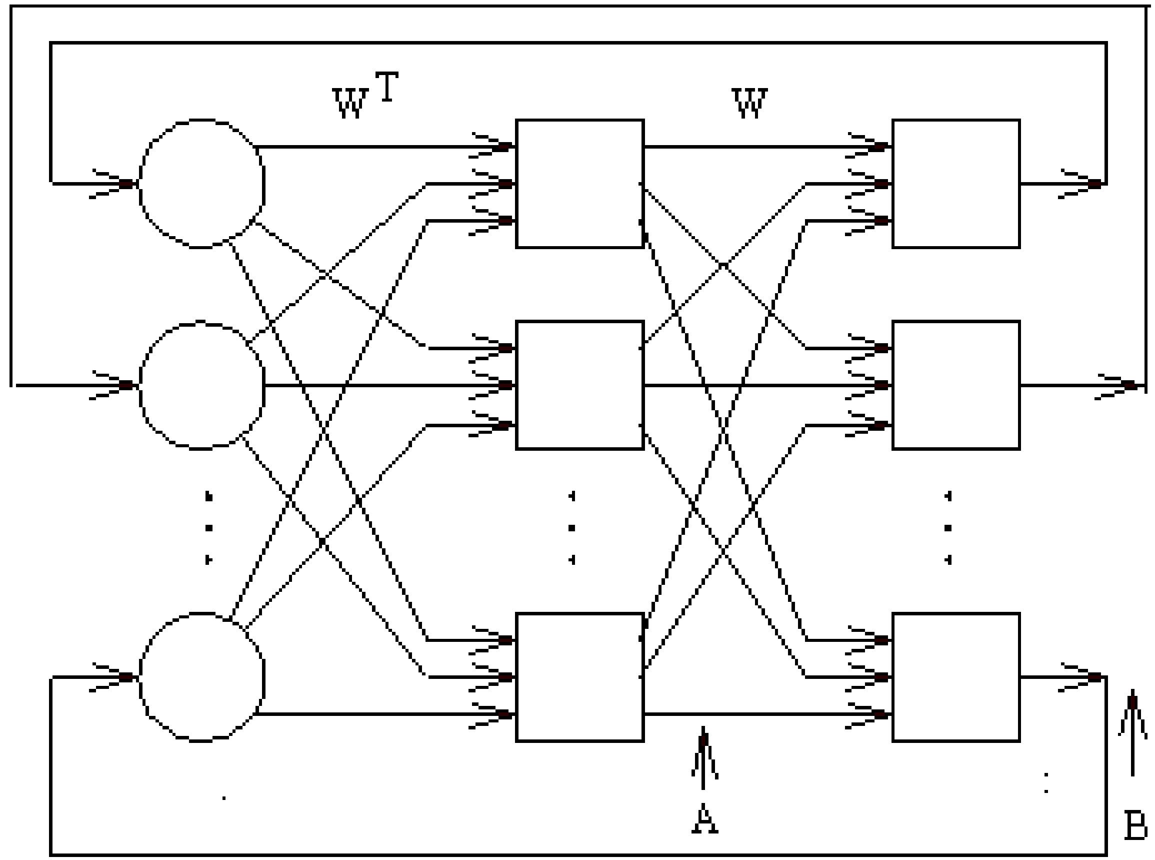
Она имеет столько входов, сколько есть выходов. Два скрытых слоя помещаются на двух отдельных ассоциативных элементах памяти и представляют удвоенный размер входных векторов. Средние слои полностью соединены между собой. Входной и выходной слои нужны для реализации средств ввода и воссоздания информации из сети.

Средние слои необходимы для сохранения ассоциированных пар векторов.

Сети ассоциативной памяти

Двунаправленная ассоциативная память

Структура двухнаправленной ассоциативной памяти



Сети ассоциативной памяти

Двунаправленная ассоциативная память

Если на вход НС поступает зашумленный вектор образа, средние слои начинают колебаться до достижения стабильного состояния, что отвечает ближайшей наученной ассоциации и на выходе генерируется образ из обучающего множества.

Подобно НС Хопфилда, двунаправленная ассоциативная память расположена к неправильному отыскиванию наученного образа, если поступает неизвестный входной вектор, который не был в составе обучающего множества.

Сеть Хопфилда автоассоциативная. Двунаправленная ассоциативная память относится к гетероассоциативной памяти. Входной вектор поступает на один набор нейронов, а соответствующий выходной вектор продуцируется на другом наборе нейронов. Входные образы ассоциируются с выходными.

Сети ассоциативной памяти

Двунаправленная ассоциативная память

Двунаправленная ассоциативная память, как и сеть Хопфилда, способна к обобщению, вырабатывая правильные выходные сигналы, несмотря на поврежденные входы.

Двунаправленная ассоциативная память работает по следующей схеме. Входной вектор A обрабатывается матрицей весов **W** сети, в результате чего продуцируется вектор выходных сигналов сети **B**. Вектор **B** обрабатывается матрицей $\mathbf{W}^T$ весов сети, которая продуцирует сигналы, представляющие новый входной вектор **A**. Этот процесс повторяется до тех пор, пока сеть не достигнет стабильного состояния, в котором ни вектор **A**, ни вектор **B** не изменяются.

Нейроны в слоях 1 и 2 функционируют, как и в других парадигмах, вычисляя сумму взвешенных входов и значения передаточной функции F.

Сети ассоциативной памяти

Двунаправленная ассоциативная память

Функция F вычисляется как $b_j = F\left(\sum_j a_j w_{ij}\right)$ или в векторной форме: B = F (AW), где B - вектор выходных сигналов нейронов слоя 2; **A** - вектор выходных сигналов нейронов слоя 1; **W** - матрица весов связей между слоями 1 и 2; F - передаточная функция.

Аналогично $\mathbf{A} = F\,(\mathbf{BW}^T)$, где $\mathbf{W}^{\mathrm{T}}$ является транспозицией матрицы **W**.

В качестве передаточной функции используется экспонентная сигмоида. Слой 0 не делает вычислений и не имеет памяти. Он является лишь средством распределения выходных сигналов слоя 2 к элементам матрицы $\mathbf{W}^T$.Формула для вычисления значений весов имеет следующий вид: $\mathbf{W} = \sum_j \mathbf{A}_j^T \mathbf{B}_j$,

где $\mathbf{A}j$ и $\mathbf{B}j$ - входные и выходные сигналы обучающего множества.

Сети ассоциативной памяти

Двунаправленная ассоциативная память

Весовая матрица вычисляется как сумма произведений всех векторных паров обучающего множества.

Системы с обратной связью имеют тенденцию к колебаниям. Они могут переходить от состояния к состоянию, никогда не достигая стабильности. Доказано, что двунаправленная ассоциативная память безусловно стабильна при любых значениях весов сети.

Применение. Ассоциативная память, распознавание образов.

Недостатки. Емкость двунаправленной ассоциативной памяти жестко ограничена. Если n - количество нейронов во входном слое, то число векторов, которые могут быть запомнены в сети, не превышает $L=n/2\log_2 n$. Так, если n=1024, то сеть способна запомнить <= 25 образов.

Двунаправленная ассоциативная память имеет некоторую непредсказуемость в процессе функционирования, возможные ошибочные ответы.

Сети ассоциативной памяти

Двунаправленная ассоциативная память

Преимущества. 1. По сравнению с автоассоциативной памятью (например, сетью Хопфилда), двунаправленная ассоциативная память дает возможность строить ассоциации между векторами А и В, которые в общем случае имеют разные размерности. За счет таких возможностей гетероассоциативная память имеет более широкий класс применений, чем автоассоциативная память.

2. Двунаправленная ассоциативная память - простая сеть, кото- рая может быть реализована в виде отдельной СБИС или оптоэлектронным способом.

3. Процесс формирования синаптических весов простой и быстрый. Сеть быстро сходится в процессе функционирования.

4. Сигналы в сети могут быть как дискретными, так и аналого- выми. Для обоих случаев доказана стабильность сети.

Сети ассоциативной памяти

Двунаправленная ассоциативная память

Пример. Каждый нейрон *a.* в первом слое *A* имеет синапсы, соединяющие его с нейронами *Ь.* во втором слое *B*. Пусть нейроны имеют следующий «смысл»: а, — валюта, $я_2$ — дол лары, a_3 — марки, $а_4$ — рубли, b_t — США, b_2 — Россия, b_3 — Канада, b_4 — Германия.

Режим обучения бинарными образами

Подадим на нейросеть три бинарных связи:

$(x_{1,} y_1), (x_{2,} y_2), (x_{3,} y_3)$.

Пусть $x_1 = (1,1,0,0) \rightarrow y_1 = (1,1,1,0)$;

$x_2 = (1,0,1,0) \rightarrow y_2 = (0,1,0,1)$;

$x_3 = (1,0,0,1) \rightarrow y_3 = (0,1,0,0)$.

Смысл обучающих связей : если возбуждены нейроны a_1 и a_2 (в нашем распоряжении есть доллары), то по соответствующим синапсам возбудятся нейроны b_1, b_2, b_3 (мы можем ими воспользоваться в США, России и Канаде), и т. д.

Сети ассоциативной памяти

Двунаправленная ассоциативная память

От бинарных связей перейдем к биполярным (это сделано исключительно для простоты, чтобы не нужно было вводить ненулевой порог срабатывания нейронов):

$x_1 = (1,1\text{-}1,\text{-}1) \rightarrow y_1 = (1,1\ \text{-}1,\text{-}1)$;

$x_2 = (1,\text{-}11,\text{-}1 \rightarrow y_2 = (1,\text{-}11,\text{-}1)$;

$x_3 = (1,\text{-}1\text{-}1,1 \rightarrow y_3 = (1,\text{-}1\text{-}1,1)$;

Составим матрицу весов:

$$W = \sum_{l=1}^{3} x_l^T y_l = \begin{pmatrix} -1 & 3 & -1 & -1 \\ 3 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 \\ -1 & -1 & -1 & -1 \end{pmatrix}.$$

Режим распознавания Оценим эффективность запоминания обучающих связей. Убедимся, что матрица W хранит связи $(x_{1,} y_1)$, $(x_{2,} y_2)$, $(x_{3,} y_3)$.

Сети ассоциативной памяти

Двунаправленная ассоциативная память

Подадим на вход x_1 следующего вида $x_1 = (2, 2, 2, -2)$ — это означает, что в слое *B* возбудятся первые три нейрона (порог срабатывания принят равным нулю). Тогда в бинарной форме $y = (1, 1, 1,0)$, что является требуемой ассоциацией. Это означает, что подача на вход x_1, приводит к y_1, то есть ЭВМ действительно «запомнила» связь ($x_{1,}$ y_1). Аналогично проверяется запоминание остальных связей.

Сеть является двунаправленной: $y_1 W^{T} = (1, 5, -3, -3) \rightarrow (1, 1, 0, 0) \rightarrow x_1$, и т. д. Определим энергию связей в памяти: аналогично $E(x_2, y_2) =- 4$ и $E(x_3\ y_3) = -2$. Следует ожидать, что при ошибке в исходной информации связь ($x_{1,}$ y_1) будет притягивать к себе больше образов, так как это точка устойчивого равновесия *с* минимальным энергетическим уровнем.

Сети ассоциативной памяти

Двунаправленная ассоциативная память

Действительно, подадим на вход образ $x' = (1,1,0,1)$ — искаженный на один бит x_1 и x_3 тогда $x'W = (1,1,1,-3) \rightarrow (1,1,1,0) \rightarrow y_1$. Аналогично, если взять $x'' = (1, O, 1, 1)$ — вектор, расположенный «между» x_2 и x_3, то получим $(-3, 1, -3, 1) \rightarrow (0, 1, 0, 1) \rightarrow y_2$ — связь (x_2, y_2), притягивает к себе, так как ее энергия меньше энергии (x_3, y_3).

Работа с неопределенными данными. Рассмотрим случай, когда тип валюты неопределен $x'=(1,0, 0,0)$, тогда $x'W = (-1, 3, -1, -1) \rightarrow (0, 1, 0, 0) \rightarrow y_3$. Это означает, что она может быть использована только в той стране, где в ходу любая валюта.

Если валюта может быть любой, например, доллары и марки, то она может использоваться везде: $x' =(1, 1, 1,0) \rightarrow x'M = (1, 1,1,1) \rightarrow y'$.

Сети ассоциативной памяти

Двунаправленная ассоциативная память

Таким образом, построенная нейросеть способна запомнить необходимую информацию на этапе обучения, а в рабочем режиме позволяет решать задачи распознавания, то есть реализует функции ассоциативной памяти. Вся полученная при обучении информация сосредоточена в матрице W . За счет параллельной структуры сеть решает задачу «мгновенно» — за одно действие — умножение входного вектора на матрицу памяти. Так как информация как бы интегрирована в матрицу W, сеть способна достаточно эффективно решать задачу и при частичных искажениях в исходных данных.

Использованные материалы

В презентации использованы материалы, предоставленные:

Бутырским Евгением Юрьевичем, доктором физико-математических наук, профессором кафедры теории управления СПбГУ;

Гришкиным Валерием Михайловичем, кандидат технических наук, доцентом кафедры компьютерного моделирования и многопроцессорных систем.

А также следующие источники:

https://github.com/vdumoulin/conv_arithmetic

https://habr.com/ru/company/oleg-bunin/blog/340184/

http://bioinformaticsinstitute.ru/sites/default/files/vvedenie_v_deep_learning.pdf