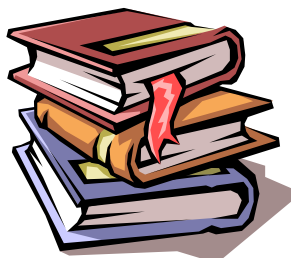


Luận văn tốt nghiệp

**Tìm hiểu về mạng nơron Kohonen
(hay mạng nơron tự tổ chức – SOM).**



MỤC LỤC

MỤC LỤC	1
CÁC HÌNH VẼ, BẢNG BIỂU TRONG LUẬN VĂN	4
THUẬT NGỮ TIẾNG ANH	4
Lời cảm ơn	5
Chương 1. Giới thiệu về mạng nơron nhân tạo	7
1.1 Cấu trúc và mô hình mạng nơron	9
1.1.1 Mô hình một nơron sinh học.....	9
1.1.2 Cấu trúc và mô hình của một nơron nhân tạo.....	10
1.2 Cấu tạo và phương thức làm việc của mạng nơron	13
1.2.1 Mạng nơron một lớp.....	15
1.2.2 Mạng nơron truyền thẳng nhiều lớp.....	16
1.2.3 Mạng nơron phản hồi.....	17
1.2.4 Mạng nơron hồi quy	17
1.2.5 Mạng Hopfield	17
1.2.6 Mạng BAM	19
1.3 Các luật học	20
1.3.1 Học có giám sát.....	21
1.3.2 Học củng cố.....	22
1.3.3 Học không có giám sát	22
1.4 Thuật toán lan truyền ngược.....	24
1.5 Kết luận	30
Chương 2. Tìm hiểu mạng nơron Kohonen.....	32
2.1 Giới thiệu.....	32
2.2 Mạng nơron Kohonen	33
2.2.1 Mạng nơron Kohonen nhận dạng.....	35
2.2.2 Cấu trúc của mạng nơron Kohonen.....	35
2.2.3 Chuẩn hóa dữ liệu đầu vào.....	36
2.2.4 Tính toán dữ liệu đầu ra của nơron	37

2.2.5 Ánh xạ lưỡng cực	37
2.2.6 Chọn nơron thắng	38
2.2.7 Quá trình học của mạng Kohonen	39
2.2.8 Tỷ lệ (tốc độ) học	40
2.2.9 Điều chỉnh các trọng số (cập nhật trọng số)	41
2.2.10 Tính toán sai số	41
2.3. Thực thi mạng nơron Kohonen	42
2.3.1 Thực thi mạng nơron truyền thẳng	42
2.3.2 Thực thi lan truyền ngược	47
2.3.3 Các tập huấn luyện	48
2.3.4 Báo cáo tiến trình	49
2.3.4.1 Lớp mạng cơ sở	49
2.3.4.2 Lớp KohonenNetwork	51
2.4 Kết luận	59
Chương 3. Nhận dạng ký tự quang sử dụng mạng nơron Kohonen	61
3.1 Giới thiệu chung	61
3.2 Huấn luyện mạng	62
3.3 Thử nghiệm sử dụng mạng nơron Kohonen để nhận dạng ký tự quang	63
3.4 Trường hợp ngoại lệ	66
3.5 Kết luận	66
KẾT LUẬN	68
Tài tham khảo	69

CÁC HÌNH VẼ, BẢNG BIỂU TRONG LUẬN VĂN

Hình 1.1 Một nơron sinh học.....	10
Hình 1.2. Sự liên kết các nơron.....	10
Hình 1.3 Mô hình một nơron nhân tạo.....	11
Hình 1.4 Đồ thị các dạng hàm kích hoạt.....	13
Hình 1.5 Mạng nơron ba lớp	14
Hình 1.6 Một số dạng mạng nơron	16
Hình 1.7 Cấu trúc của mạng Hopfield	18
Hình 1.8 Cấu trúc của BAM	19
Hình 1.9: Cấu trúc huấn luyện mạng nơron.....	20
Hình 1.10: Học có giám sát.....	22
Hình 1.12: Học không có giám sát.....	22
Hình 1.13: Sơ đồ cấu trúc chung của quá trình học.....	23
Hình 1.14: Mạng 3 lớp lan truyền ngược.....	25
Hình 2.1: Một dạng mạng nơron Kohonen	36
Hình 2.2: Sơ đồ khối biểu diễn huấn luyện mạng nơron Kohonen.....	40
Hình 3.1:Mô hình chung trong nhận dạng chữ viết.....	62
Hình 3.2: Sơ đồ huấn luyện mạng.....	62
Hình 3.4: Biểu diễn ký tự e theo ma trận 5×7 pixel.....	64
Hình 3.5: Biểu diễn ký tự e theo ma trận 5×7 bởi các giá trị.....	64
Hình 3.6: Biểu diễn ký tự e ở vector đầu vào	64
Hình 3.7 Vẽ và gán ký tự	65
Hình 3.8 Kết quả mạng nơron Kohonen nhận dạng ký tự e và ký tự c	65
Hình 3.9 Đưa chữ ký vào mạng và gán tên	65
Hình 3.10 Kết quả mạng nơron Kohonen nhận dạng chữ ký.....	66

THUẬT NGỮ TIẾNG ANH

ANN	Mạng noron cần huấn luyện
BAM	Mạng BAM (Bidirectional Associative Memory)
SOM	Mạng noron tự tổ chức (Self Organizing Maps)
PE	Phần tử xử lý (Processing Element)
OCR	Nhận dạng ký tự quang (optical character recognition)

Lời cảm ơn

Chúng ta đều biết rằng, bộ não con người là một sản phẩm hoàn hảo của tạo hóa, nó có khả năng tư duy và sáng tạo. Hiện nay, con người đang nghiên cứu

phương thức hoạt động của bộ não, sau đó áp dụng cho những công nghệ hiện đại. Để tiếp cận khả năng học, người ta đưa ra mô hình mạng nơron gồm các nơron liên kết với nhau thành mạng phẳng theo cấu trúc mạng thần kinh của con người.

Mỗi nơron riêng lẻ có khả năng xử lý thông tin yếu, nhưng khi chúng được ghép với nhau thành mạng, thì khả năng xử lý thông tin sẽ mạnh hơn rất nhiều. Mỗi cấu trúc mạng đều có một ưu điểm đặc thù, chúng cho ta một công cụ mạnh trong các lĩnh vực kỹ thuật điều khiển và kỹ thuật thông tin. Một mạng nơron nhân tạo là tập hợp một số lớn các phần tử xử lý (các nút hay các khối), thường được tổ chức song song và được cấu hình theo kiến trúc đệ quy. Cách ứng sử trên mạng nơron nhân tạo giống như bộ não con người, nó chứng tỏ khả năng học, nhớ lại, và tổng quát hóa từ dữ liệu huấn luyện.

Mạng nơron nhân tạo là công cụ tốt trong việc giải quyết các bài toán như: hợp và phân lớp đối tượng, xấp xỉ hàm, tối ưu hóa, định lượng vector, phân cụm dữ liệu,... Nó thay thế hiệu quả các công cụ tính toán truyền thống để giải quyết các bài toán này.

Nhận dạng là một lĩnh vực đóng vai trò quan trọng trong khoa học kỹ thuật. Trong hầu hết các vấn đề kỹ thuật ngày nay, ta đều phải xác định, nhận dạng được các mô hình và đối tượng liên quan, để từ đó tìm ra giải pháp. Nhận dạng mô hình là bài toán rất quan trọng trong lý thuyết hệ thống. Lý do đơn giản là vì không thể phân tích, tổng hợp hệ thống khi không có mô hình toán học mô tả hệ thống. Trong quá trình xây dựng mô hình hệ thống trên phương diện lý thuyết, người ta thường không khảo sát được mọi ảnh hưởng của môi trường đến tính động học của hệ thống, cũng như những tác động qua lại bên trong hệ thống một cách chính xác tuyệt đối. Rất nhiều yếu tố đã bị bỏ qua, hoặc chỉ được xem xét đến như là một tác động ngẫu nhiên. Bởi vậy, nếu nói một cách chặt chẽ thì những hiểu biết lý thuyết ban đầu về hệ thống, mới chỉ có thể giúp ta khoanh được lớp các mô hình thích hợp. Để có thể có được một mô hình cụ thể có chất lượng phù hợp với bài cụ thể toán đặt ra trong lớp các mô hình thích hợp đó, thì phải sử dụng phương pháp nhận dạng. Còn bài toán nhận dạng, phân tích phân cụm dữ liệu, là các bài toán rất hay gặp

trong thực tế, khi chúng ta nhìn thấy một vật gì đó, thì câu hỏi thường trực của mỗi người là: vật đó có mấy loại, và nó thuộc loại nào trong các loại có thể có.

Để giải quyết các bài toán nhận dạng, người ta đã đưa vào các cách tiếp cận khác nhau, mỗi phương pháp tiếp cận trong những bài toán cụ thể đều có những ưu, nhược điểm riêng. Phương pháp ứng dụng mạng nơron trong nhận dạng là một cách tiếp cận mới và hiện đại. Nó có thể là công cụ rất mạnh để giải quyết các bài toán trong lĩnh vực này.

Nội dung của đề tài đi vào tìm hiểu và xây dựng các phần tử nơron cơ bản, xem xét và nghiên cứu cấu trúc một mạng nơron, giới thiệu về mạng nơron nhiều lớp với thuật toán lan truyền ngược. Trọng tâm của đề tài đi vào tìm hiểu về mạng nơron Kohonen (hay mạng nơron tự tổ chức – SOM).

Đề tài gồm ba chương

Chương 1, trình bày cấu trúc một phần tử nơron cơ bản, các cấu trúc mạng nơron nhân tạo thường gặp, thuật toán học, phân tích ưu nhược điểm của chúng, và giới thiệu về thuật toán lan truyền ngược.

Chương 2, tìm hiểu mạng nơron Kohonen.

Chương 3, nhận dạng ký tự quang sử dụng mạng nơron Kohonen.

Cuối cùng em xin cảm ơn các thầy cô giáo, đặc biệt là PGS.TSKH Bùi Công Cường đã tận tình chỉ dẫn cho em trong suốt thời gian làm đề tài. Xin cảm ơn các bạn cùng lớp đã tạo điều kiện cho tôi được học tập và nghiên cứu trong môi trường tốt.

Hà nội, tháng 12 năm 2009.

Chương 1. Giới thiệu về mạng nơron nhân tạo

Học máy là một ngành khoa học nghiên cứu các *thuật toán* cho phép máy tính có thể học được các khái niệm.

Phân loại: Có hai loại phương pháp học máy chính

- Phương pháp quy nạp: Máy học/phân biệt các khái niệm dựa trên dữ liệu đã thu thập được trước đó. Phương pháp này cho phép tận dụng được nguồn dữ liệu rất nhiều và sẵn có.
- Phương pháp suy diễn: Máy học/phân biệt các khái niệm dựa vào các luật. Phương pháp này cho phép tận dụng được các kiến thức chuyên ngành để hỗ trợ máy tính.

Hiện nay, các thuật toán đều cố gắng tận dụng được ưu điểm của hai phương pháp này.

Các ngành khoa học liên quan:

- Lý thuyết thống kê: các kết quả trong xác suất thống kê là tiền đề cho rất nhiều phương pháp học máy. Đặc biệt, lý thuyết thống kê cho phép ước lượng sai số của các phương pháp học máy.
- Các phương pháp tính: các thuật toán học máy thường sử dụng các tính toán số thực/số nguyên trên dữ liệu rất lớn. Trong đó, các bài toán như: tối ưu có/không ràng buộc, giải phương trình tuyến tính v.v... được sử dụng rất phổ biến.
- Khoa học máy tính: là cơ sở để thiết kế các thuật toán, đồng thời đánh giá thời gian chạy, bộ nhớ của các thuật toán học máy.

Ứng dụng: *Học máy* có ứng dụng rộng khắp trong các ngành khoa học/sản xuất, đặc biệt những ngành cần phân tích khối lượng dữ liệu khổng lồ. Một số ứng dụng thường thấy như:

- Xử lý ngôn ngữ tự nhiên: xử lý văn bản, giao tiếp người – máy, ...
- Nhận dạng: nhận dạng tiếng nói, chữ viết tay, vân tay, thị giác máy (Computer Vision) ...

-
-
- Tìm kiếm
 - Chẩn đoán trong y tế: phân tích ảnh X-quang, các hệ chuyên gia chẩn đoán tự động.
 - Tin sinh học: phân loại chuỗi gene, quá trình hình thành gene/protein
 - Vật lý: phân tích ảnh thiên văn, tác động giữa các hạt ...
 - Phát hiện gian lận tài chính (financial fraud): gian lận thẻ tín dụng
 - Phân tích thị trường chứng khoán (stock market analysis)
 - Chơi trò chơi: tự động chơi cờ, hành động của các nhân vật ảo
 - Rôbot: là tổng hợp của rất nhiều ngành khoa học, trong đó *học máy* tạo nên hệ thần kinh/bộ não của người máy.
 - **Các nhóm giải thuật học máy:** Học có giám sát, học không giám sát, học nửa giám sát, học tăng cường,...

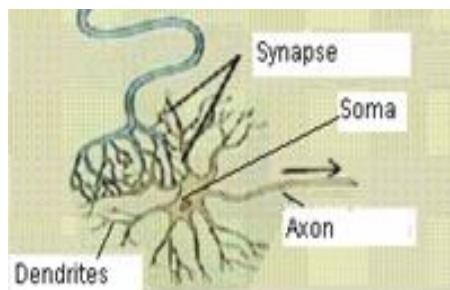
1.1 Cấu trúc và mô hình mạng nơron

1.1.1 Mô hình một nơron sinh học

Phần tử xử lý cơ bản của một mạng nơron sinh học là một nơron, phần tử này có thể chia làm bốn thành phần cơ bản như sau: dendrites, soma, axon, và synapses.

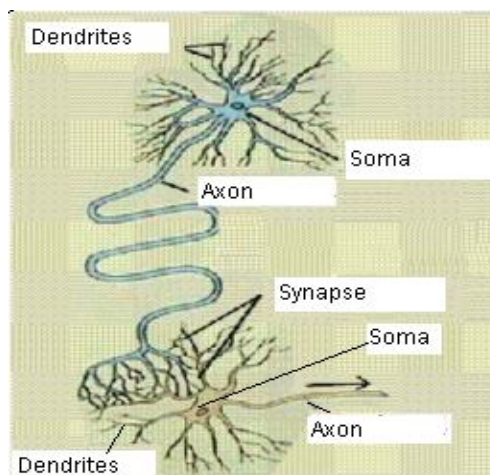
- Dendrites: là phần nhận tín hiệu đầu vào.
- Soma: là hạt nhân.
- Axon: là phần dẫn ra tín hiệu xử lý.
- Synapses: là đường tín hiệu điện hóa giao tiếp giữa các nơron.

Kiến trúc cơ sở này của bộ não con người có một vài đặc tính chung. Một cách tổng quát, thì một nơron sinh học nhận đầu vào từ các nguồn khác nhau, kết hợp chúng tại với nhau, thực thi tổ hợp phi tuyến chúng để cho ra kết quả cuối cùng ở đầu ra. Hình 1.1 chỉ ra mối quan hệ giữa bốn phần tử của một nơron sinh học.



Hình 1.1 Một nơron sinh học

Một nơron sinh học chỉ có một số chức năng cơ bản như vậy, ta nhận thấy khả năng xử lý thông tin của nó là rất yếu. Để có được khả năng xử lý thông tin hoàn hảo như bộ não con người, thì các nơron phải kết hợp và trao đổi thông tin với nhau. Ta hình dung sơ đồ liên kết, và trao đổi thông tin giữa hai nơron như hình 1.2.

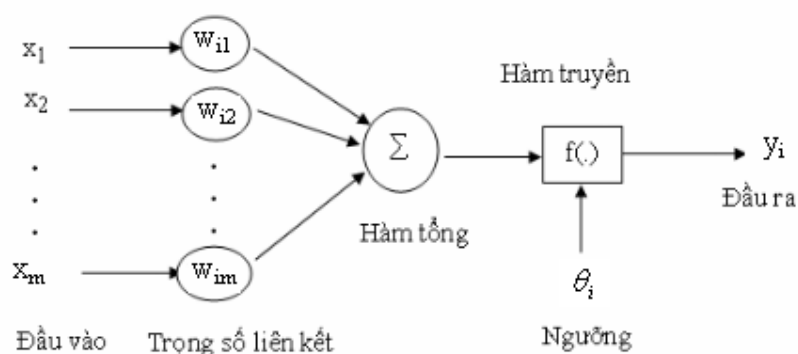


Hình 1.2. Sự liên kết các nơron

1.1.2 Cấu trúc và mô hình của một nơron nhân tạo

Mô hình toán học của mạng nơron sinh học được đề xuất bởi McCulloch và Pitts, thường được gọi là nơron M-P, ngoài ra nó còn được gọi là phân tử xử lý và được ký hiệu là PE (Processing Element).

Mô hình nơron có m đầu vào x_1, x_2, \dots, x_m , và một đầu ra y_i như sau:



Hình 1.3 Mô hình một nơron nhân tạo

Giải thích các thành phần cơ bản:

- Tập các đầu vào: Là các tín hiệu vào của nơron, các tín hiệu này thường được đưa vào dưới dạng một vector m chiều.
- Tập các liên kết (các trọng số): Mỗi liên kết được thể hiện bởi một trọng số (thường được gọi là trọng số liên kết). Trọng số liên kết giữa tín hiệu vào thứ j cho nơron i thường được ký hiệu là w_{ij} . Thông thường các trọng số này được khởi tạo ngẫu nhiên ở thời điểm khởi tạo mạng và được cập nhật liên tục trong quá trình học mạng.
- Bộ tổng (Hàm tổng): Thường dùng để tính tổng của tích các đầu vào với trọng số liên kết của nó.
- Ngưỡng: Ngưỡng này thường được đưa vào như một thành phần của hàm truyền.
- Hàm truyền: Hàm này dùng để giới hạn phạm vi đầu ra của mỗi nơron. Nó nhận đầu vào là kết quả của hàm tổng và ngưỡng đã cho. Thông thường, phạm vi đầu ra của mỗi nơron được giới hạn trong đoạn $[0,1]$ hoặc $[-1,1]$. Các hàm truyền rất đa dạng, có thể là các hàm tuyến tính hoặc phi tuyến. Việc lựa chọn hàm truyền tùy thuộc vào từng bài toán và kinh nghiệm của người thiết kế mạng.

- Đầu ra: Là tín hiệu đầu ra của một nơron, với mỗi nơron sẽ có tối đa một đầu ra.

Về mặt toán học, cấu trúc của một nơron i được mô tả bằng cặp biểu thức sau:

$$y_i = f(net_i - \theta_i) \text{ và } net_i = \sum_{j=1}^n w_{ij}x_j$$

trong đó: x_1, x_2, \dots, x_m là các tín hiệu đầu vào, còn $w_{i1}, w_{i2}, \dots, w_{im}$ là các trọng số kết nối của nơron thứ i , net_i là hàm tổng, f là hàm truyền, θ_i là một ngưỡng, y_i là tín hiệu đầu ra của nơron.

Như vậy, tương tự như nơron sinh học, nơron nhân tạo cũng nhận các tín hiệu đầu vào, xử lý (nhân các tín hiệu này với trọng số liên kết, tính tổng các tích thu được rồi gửi kết quả đến hàm truyền), và cho một tín hiệu đầu ra (là kết quả của hàm truyền).

- Hàm truyền có thể có các dạng sau:

- Hàm bước
$$y = \begin{cases} 1 & \text{khi } x \geq 0 \\ 0 & \text{khi } x < 0 \end{cases} \quad (1.6)$$

- Hàm giới hạn chặt (hay còn gọi là hàm bước)

$$y = \text{sgn}(x) = \begin{cases} 1 & \text{khi } x \geq 0 \\ -1 & \text{khi } x < 0 \end{cases} \quad (1.7)$$

- Hàm bậc thang

$$y = \text{sgn}(x) = \begin{cases} 1 & \text{khi } x > 1 \\ x & \text{khi } 0 \leq x \leq 1 \\ 0 & \text{khi } x < 0 \end{cases} \quad (1.8)$$

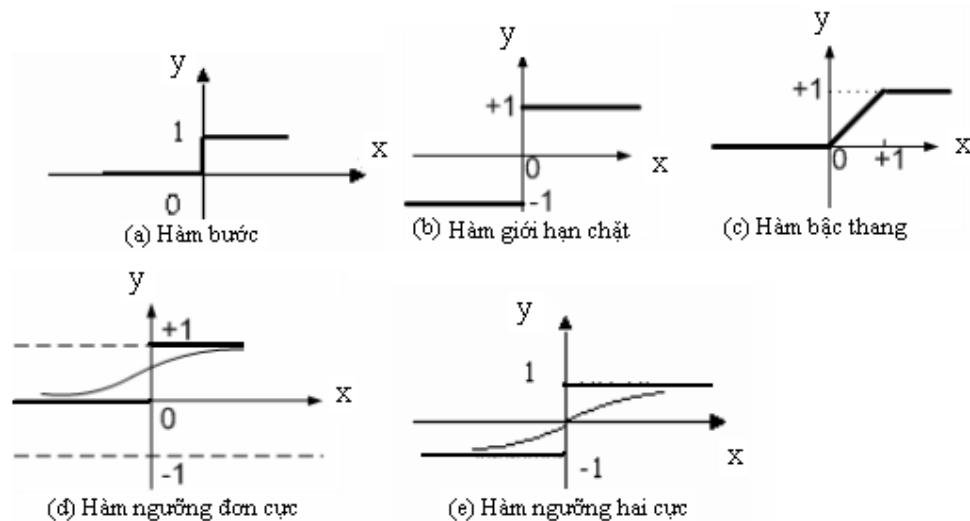
- Hàm ngưỡng đơn cực

$$y = \frac{1}{1 + e^{-\lambda x}} \text{ với } \lambda > 0 \quad (1.9)$$

- Hàm ngưỡng hai cực

$$y = \frac{2}{1 + e^{-\lambda x}} - 1 \text{ với } \lambda > 0 \quad (1.10)$$

- Đồ thị các dạng hàm truyền được biểu diễn như sau:



Hình 1.4 Đồ thị các dạng hàm truyền

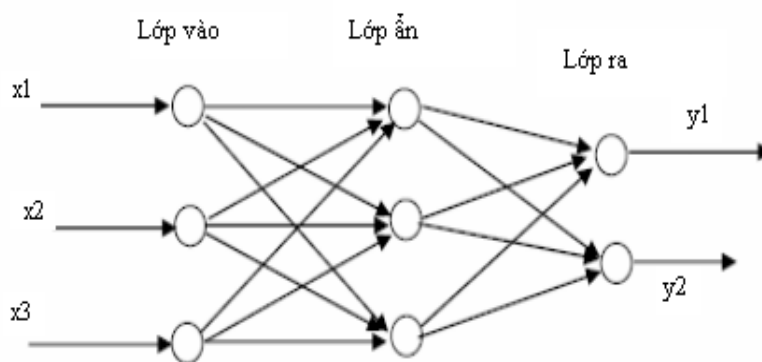
1.2 Cấu tạo và phương thức làm việc của mạng nơron

Dựa trên những phương pháp xây dựng nơron đã trình bày ở mục trên, ta có thể hình dung mạng nơron như là một hệ truyền đạt và xử lý tín hiệu. Đặc tính truyền đạt của nơron phần lớn là đặc tính truyền đạt tĩnh.

Khi liên kết các đầu vào/ra của nhiều nơron với nhau, ta thu được một mạng nơron, việc ghép nối các nơron trong mạng với nhau có thể là theo một nguyên tắc bất kỳ. Vì mạng nơron là một hệ truyền đạt và xử lý tín hiệu, nên có thể phân biệt các loại nơron khác nhau, các nơron có đầu vào nhận thông tin từ môi trường bên ngoài khác với các nơron có đầu vào được nối với các nơron khác trong mạng, chúng được phân biệt với nhau qua vector hàm trọng số ở đầu vào w .

Nguyên lý cấu tạo của mạng nơron bao gồm nhiều lớp, mỗi lớp bao gồm nhiều nơron có cùng chức năng trong mạng. Hình 1.5 là mô hình hoạt động của một mạng nơron 3 lớp với 8 phần tử nơron. Mạng có ba đầu vào là x_1, x_2, x_3 và hai đầu ra y_1, y_2 . Các tín hiệu đầu vào được đưa đến 3 nơron đầu vào, 3 nơron này làm thành lớp đầu vào của mạng. Các nơron trong lớp này được gọi là nơron đầu vào. Đầu ra của

các nơron này được đưa đến đầu vào của 3 nơron tiếp theo, 3 nơron này không trực tiếp tiếp xúc với môi trường bên ngoài mà làm thành lớp ẩn, hay còn gọi là lớp trung gian. Các nơron trong lớp này có tên là nơron nội hay nơron ẩn. Đầu ra của các nơron này được đưa đến 2 nơron đưa tín hiệu ra môi trường bên ngoài. Các nơron trong lớp đầu ra này được gọi là nơron đầu ra.



Hình 1.5 Mạng nơron ba lớp

Mạng nơron được xây dựng như trên là mạng gồm 3 lớp mắc nối tiếp nhau đi từ đầu vào đến đầu ra. Trong mạng không tồn tại bất kỳ một mạch hồi tiếp nào. Một mạng nơron có cấu trúc như vậy gọi là mạng một hướng hay mạng truyền thẳng một hướng (Feed forward network), và có cấu trúc mạng ghép nối hoàn toàn (vì bất cứ một nơron nào trong mạng cũng được nối với một hoặc vài nơron khác). Mạng nơron bao gồm một hay nhiều lớp trung gian được gọi là mạng Multilayer Perceptrons (MLP-Network).

Mạng nơron khi mới được hình thành thì chưa có tri thức, tri thức của mạng sẽ được hình thành dần dần sau một quá trình học. Mạng nơron được học bằng cách đưa vào những kích thích, và mạng hình thành những đáp ứng tương ứng, những đáp ứng tương ứng phù hợp với từng loại kích thích sẽ được lưu trữ. Giai đoạn này được gọi là giai đoạn học của mạng. Khi đã hình thành tri thức mạng, mạng có thể giải quyết các vấn đề một cách đúng đắn. Đó có thể là vấn đề ứng dụng rất khác

nhau, được giải quyết chủ yếu dựa trên sự tổ chức hợp nhất giữa các thông tin đầu vào của mạng và các đáp ứng đầu ra.

- Nếu nhiệm vụ của một mạng là hoàn chỉnh hoặc hiệu chỉnh các thông tin thu được không đầy đủ hoặc bị tác động của nhiễu. Mạng nơron kiểu này được ứng dụng trong lĩnh vực hoàn thiện mẫu, trong đó có một ứng dụng cụ thể là nhận dạng chữ viết.
- Nhiệm vụ tổng quát của một mạng nơron là lưu giữ động các thông tin. Dạng thông tin lưu giữ này chính là quan hệ giữa các thông tin đầu vào và các đáp ứng đầu ra tương ứng, để khi có một kích thích bất kỳ tác động vào mạng, mạng có khả năng suy diễn và đưa ra một đáp ứng phù hợp. Đây chính là chức năng nhận dạng theo mẫu của mạng nơron. Để thực hiện chức năng này, mạng nơron đóng vai trò như một bộ phận tổ chức các nhóm thông tin đầu vào, và tương ứng với mỗi nhóm là một đáp ứng đầu ra phù hợp. Như vậy, một nhóm bao gồm một loại thông tin đầu vào và một đáp ứng đầu ra. Các nhóm có thể được hình thành trong quá trình học, và cũng có thể không hình thành trong quá trình học.

Hình 1.6 là một số liên kết đặc thù của mạng nơron. Nơron được vẽ là các vòng tròn xem như một tế bào thần kinh, chúng có các mối liên hệ đến các nơron khác nhờ các trọng số liên kết. Tập hợp các trọng số liên kết này sẽ lập thành các ma trận trọng số tương ứng.

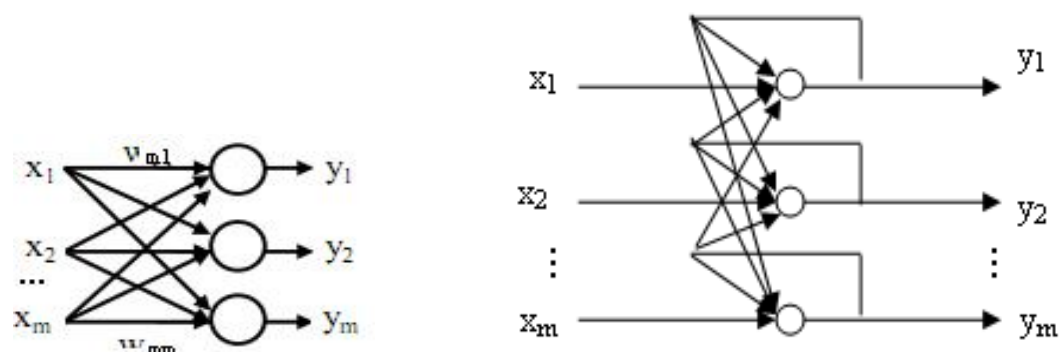
1.2.1 Mạng nơron một lớp

Mỗi một nơron có thể phối hợp với các nơron khác tạo thành một lớp các trọng số. Mạng một lớp truyền thẳng như hình 1.6a. Một lớp nơron là một nhóm các nơron mà chúng đều có cùng trọng số, nhận cùng một tín hiệu đầu vào đồng thời.

Trong ma trận trọng số, các hàng là thể hiện nơron, hàng thứ j có thể đặt nhãn như một vector w_j của nơron thứ j gồm m trọng số w_{ji} . Các trọng số trong cùng một cột thứ j ($j=1,2,...,n$) đồng thời cùng nhận một tín hiệu đầu vào x_j .

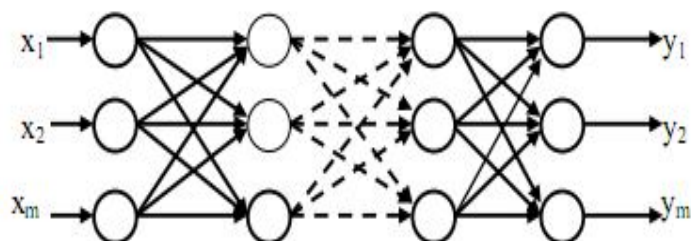
$$w_j = [w_{j1}, w_{j2}, \dots, w_{jm}]$$

Tại cùng một thời điểm, vector đầu vào $x = [x_1, x_2, \dots, x_n]$ có thể là một nguồn bên ngoài là cảm biến hoặc thiết bị đo lường đưa tới mạng.

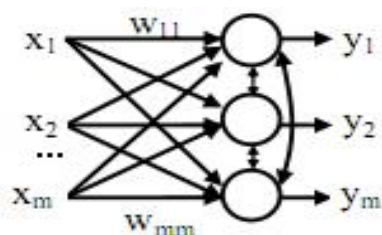


(a) Mạng truyền thẳng một lớp

(b) Mạng hồi tiếp một lớp



(c) Mạng truyền thẳng nhiều lớp



(d) Mạng nơron hồi quy

Hình 1.6 Một số dạng mạng nơron

1.2.2 Mạng nơron truyền thẳng nhiều lớp

Mạng nơron nhiều lớp (Hình 1.6.c) có các lớp được phân chia thành 3 loại sau đây:

- Lớp vào là lớp nơron đầu tiên nhận tín hiệu vào x_i ($i = 1, 2, \dots, n$). Mỗi tín hiệu x_i được đưa đến tất cả các nơron của lớp đầu vào. Thông thường, các nơron đầu vào không làm biến đổi các tín hiệu vào x_i , tức là chúng không có các trọng số hoặc không có các loại hàm chuyển đổi nào, chúng chỉ đóng vai trò phân phối các tín hiệu.
- Lớp ẩn là lớp nơron sau lớp vào, chúng không trực tiếp liên hệ với thế giới bên ngoài như các lớp nơron vào/ra.
- Lớp ra là lớp nơron tạo ra các tín hiệu ra cuối cùng.

1.2.3 Mạng nơron phản hồi

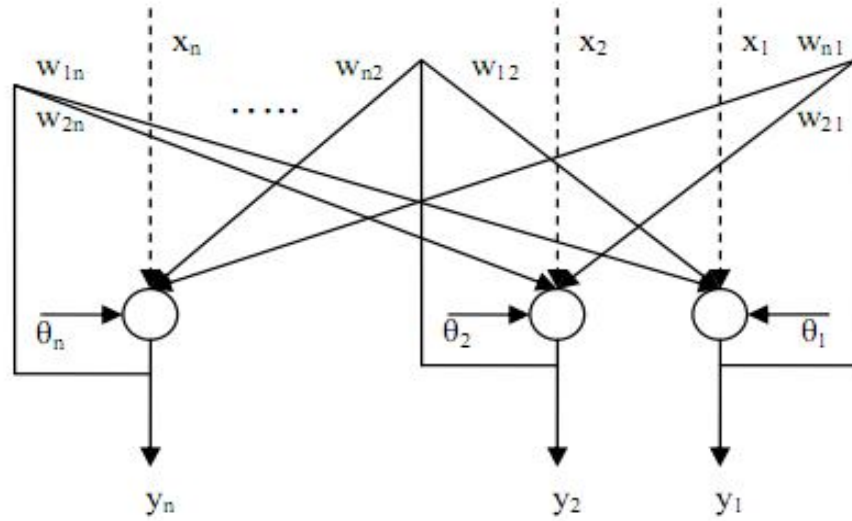
Mạng nơron phản hồi là mạng mà đầu ra của mỗi nơron được quay trở lại nối với đầu vào của các nơron cùng lớp được gọi là mạng Laeral như hình 1.6b

1.2.4 Mạng nơron hồi quy

Mạng nơron phản hồi có thể thực hiện đóng vòng được gọi là mạng nơron hồi quy như hình 1.6d. Mạng nơron hồi quy có trọng số liên kết đối xứng như mạng Hopfield, mạng luôn hội tụ về trạng thái ổn định (Hình 1.6.b). Mạng BAM thuộc nhóm mạng nơron hồi quy, gồm 2 lớp liên kết 2 chiều, không được gắn với tín hiệu vào/ra. Nghiên cứu mạng nơron hồi quy mà có trọng số liên kết không đối xứng, thì sẽ gặp phải vấn đề phức tạp nhiều hơn so với mạng truyền thẳng và mạng hồi quy có trọng số liên kết đối xứng.

1.2.5 Mạng Hopfield

Mạng Hopfield là mạng phản hồi một lớp, được chỉ ra trong hình 1.6.b. Cấu trúc chi tiết của nó được thể hiện trong hình 1.7. Khi hoạt động với tín hiệu rời rạc, nó được gọi là mạng Hopfield rời rạc, và cấu trúc của nó cũng được gọi là mạng hồi quy.



Hình 1.7 Cấu trúc của mạng Hopfield

Như mạng Hopfield đã vẽ ở trên, ta thấy nút có một đầu vào bên ngoài x_j và một giá trị ngưỡng θ_j ($j = 1, 2, \dots, n$). Một điều quan trọng cần nói ở đây là mỗi nút không có đường phản hồi về chính nó. Nút đầu ra thứ j được nối tới mỗi đầu vào của nút khác qua trọng số w_{ij} , với $i \neq j$, ($i = 1, 2, \dots, n$), hay nói cách khác $w_{ii} = 0$, (với $i = 1, 2, \dots, n$).

Một điều quan trọng nữa là trọng số của mạng Hopfield là đối xứng, tức là $w_{ij} = w_{ji}$, (với $i, j = 1, 2, \dots, n$). Khi đó, luật cập nhật cho mỗi nút mạng là như sau:

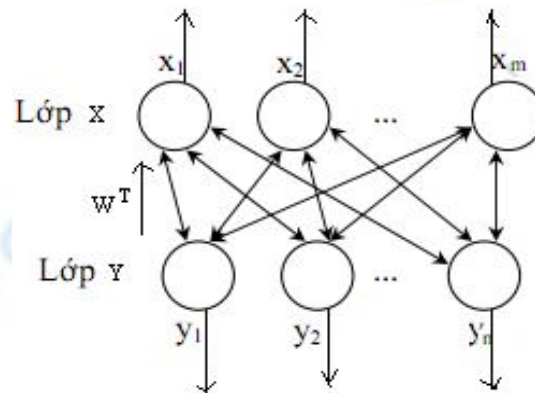
$$y_i^{(k+1)} = \text{sgn} \left(\sum_{\substack{j=1 \\ j \neq i}}^n w_{ij} y_j^{(k)} + x_i - \theta \right), \quad i = 1, 2, \dots, n \quad (1.11)$$

Luật cập nhật trên được tính toán trong cách thức không đồng bộ. Điều này có nghĩa là, với một thời gian cho trước, chỉ có một nút mạng cập nhật được đầu ra của nó. Sự cập nhật tiếp theo trên một nút sẽ sử dụng chính những đầu ra đã được cập nhật. Nói cách khác, dưới hình thức hoạt động không đồng bộ của mạng, mỗi đầu ra được cập nhật độc lập.

Có sự khác biệt giữa luật cập nhật đồng bộ và luật cập nhật không đồng bộ. Với luật cập nhật không đồng bộ thì sẽ chỉ có một trạng thái cân bằng của hệ (với giá trị đầu đã được xác định trước). Trong khi đó, với luật cập nhật đồng bộ thì có thể làm mạng hội tụ ở mỗi điểm cố định hoặc một vòng giới hạn.

1.2.6 Mạng BAM

Mạng BAM bao gồm hai lớp và được xem như là trường hợp mở rộng của mạng Hopfield. Ở đây ta chỉ xét mạng rời rạc, vì nó đơn giản và dễ hiểu.



Hình 1.8 Cấu trúc của BAM

Khi mạng nơron được tích cực với giá trị đầu vào của vector tại đầu vào của một lớp, mạng sẽ có hai mẫu trạng thái ổn định, với mỗi mẫu tại đầu ra của nó là một lớp. Tính động học của mạng thể hiện dưới dạng tác động qua lại giữa hai lớp. Cụ thể hơn, giả sử một vector đầu vào x được cung cấp cho đầu vào của lớp nơron y . Đầu vào được xử lý và truyền tới đầu ra của lớp y như sau:

$$y' = a(wx) \quad ; \quad y'_i = a\left(\sum w_{ij}x_j\right) \quad ; \quad \text{với } i = 1, 2, \dots, n \quad (1.12)$$

Ở đó $a(\cdot)$ là hàm truyền, vector y' bây giờ lại nuôi trở lại lớp nơron X và tạo nên đầu ra như sau:

$$x' = a(w^T y') \quad ; \quad x_j = a\left(\sum_{i=1}^n w_{ij}y'_i\right) \quad ; \quad \text{với } j = 1, 2, \dots, m \quad (1.13)$$

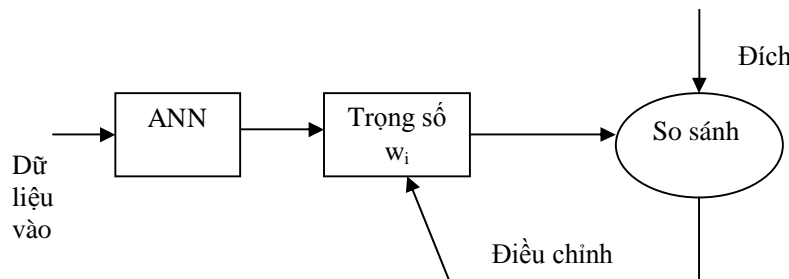
Sau đó x' nuôi trở lại đầu vào của lớp y và tạo ra hàm y'' theo phương trình (1.12). Quá trình này cứ tiếp tục, bao gồm các bước như sau:

$$\begin{aligned} y^{(1)} &= a(wx^{(0)}) && \text{(truyền thẳng lần thứ nhất)} \\ x^{(2)} &= a(w^{(T)}y^{(1)}) && \text{(truyền ngược lần thứ nhất)} \\ y^{(3)} &= a(wx^{(2)}) && \text{(truyền thẳng lần thứ hai)} \\ x^{(4)} &= a(w^{(T)}y^{(3)}) && \text{(truyền ngược lần thứ hai)} \\ &\vdots \\ y^{(k-1)} &= a(wx^{(k-2)}) && \text{(truyền thẳng lần thứ } k/2) \\ x^{(k)} &= a(w^{(T)}y^{(k-1)}) && \text{(truyền ngược lần thứ } k/2) \end{aligned} \quad (1.14)$$

Chú ý rằng trạng thái cập nhật trong phương trình (1.14) là đồng bộ theo phương trình (1.12) và (1.13). Trạng thái cập nhật cũng có thể không đồng bộ theo phương trình (1.12) và (1.13) với các nút i, j được chọn tự do. Người ta đã chỉ ra rằng, hệ thống ổn định cho cả hai chế độ đồng bộ và không đồng bộ. Tuy nhiên, chế độ đồng bộ sẽ làm cho hệ thống hội tụ nhanh hơn nhiều.

1.3 Các luật học

Thông thường, mạng nơron được điều chỉnh hoặc được huấn luyện để hướng các đầu vào riêng biệt đến đích ở đầu ra. Cấu trúc huấn luyện mạng được chỉ ra ở hình dưới. Ở đây, hàm trọng số của mạng được điều chỉnh trên cơ sở so sánh đầu ra với đích mong muốn (taget), cho tới khi đầu ra của mạng phù hợp với đích. Những cặp vào/đích (input/taget) được dùng để giám sát cho sự huấn luyện mạng.



Hình 1.9: Cấu trúc huấn luyện mạng nơron