

MỤC LỤC

DANH MỤC HÌNH VẼ	3
TỪ KHÓA VIẾT TẮT.....	4
LỜI CẢM ƠN	5
LỜI NÓI ĐẦU	6
CHƯƠNG 1 : TỔNG QUAN VỀ DỰ PHỤ TẢI NGẮN HẠN.....	7
1.1 Vai trò của dự báo phụ tải ngắn hạn trong công tác điều độ	7
1.2 Tầm quan trọng của dự báo phụ tải điện	9
1.3 Các yếu tố ảnh hưởng trực tiếp đến phụ tải ngày	12
1.4 Phương hướng nghiên cứu dự báo phụ tải.....	13
CHƯƠNG 2 : MẠNG NƠN.....	14
2.1 Giới thiệu về mạng nơron	14
2.1.1 Lịch sử phát triển.....	14
2.1.2 Ứng dụng	15
2.1.3 Căn nguyên sinh học	16
2.2 Nơron nhân tạo	17
2.3 Hàm xử lý	18
2.3.1 Hàm kết hợp	18
2.3.2 Hàm kích hoạt (hàm chuyển)	18
2.4 Cấu trúc của mạng nơron.....	20
2.4.1 Mạng truyền thẳng.....	20
2.4.2 Mạng hồi qui	21
2.5 Luật học	21
2.6 Hàm mục tiêu.....	22
CHƯƠNG 3 : ỨNG DỤNG MẠNG NƠN TRONG DỰ BÁO PHỤ TẢI ĐIỆN.....	24
3.1 Mạng nơron truyền thẳng	24
3.1.1 Kiến trúc cơ bản	24
3.1.2 Khả năng biểu diễn.....	25
3.1.3 Vấn đề thiết kế cấu trúc mạng	26
3.1.4 Số lớp ẩn.....	26

3.1.5 Số nơon trong lớp ẩn.....	26
3.2 Thuật toán lan truyền ngược(Back – Propagation Algorithm).....	28
3.2.1 Mô tả thuật toán.....	28
3.2.2 Sử dụng thuật toán lan truyền ngược	34
3.2.3 Một số biến thể của thuật toán lan truyền ngược	38
3.3 Chương trình dự báo đỉnh và đáy biểu đồ phụ tải	40
3.3.1 Cấu trúc mạng nơon dự báo đỉnh biểu đồ phụ tải	40
3.3.2 Dữ liệu dự báo phụ tải	41
3.3.3 Sơ đồ khối của chương trình	47
3.3.4 Hướng dẫn sử dụng chương trình.....	48
KẾT LUẬN	54
TÀI LIỆU THAM KHẢO	56

DANH MỤC HÌNH VẼ

Hình 1.1: Sản lượng trung bình ngày tháng 1 qua các năm a	10
Hình 1.2: Tốc độ tăng trưởng của phụ tải trung bình tháng 1 qua các năm b.....	11
Hình 2.1: Đơn vị xử lý thứ j c	17
Hình 2.2: Hàm tuyến tính d.....	18
Hình 2.3: Hàm ranh giới cứng e.....	19
Hình 2.4: Hàm Log – Sigmoid f.....	20
Hình 2.5: Hàm Sigmoid lưỡng cực g	21
Hình 2.6: Mạng nơron truyền thẳng nhiều lớp h.....	21
Hình 2.7: Mạng hồi qui i.....	22
Hình 2.8: Mô hình học có thầy j	24
Hình 3.1: Mạng nơron truyền thẳng nhiều lớp k.....	35
<i>a, Trường hợp $i=1$ l.....</i>	35
<i>b, Trường hợp $i=2$ m.....</i>	36
<i>c, Trường hợp $i=4$ n</i>	36
<i>d, Trường hợp $i=8$ o</i>	36
<i>Trường hợp $i=8$ và số nơron lớp ẩn là 8 p</i>	37
Hình 3.2: Kiến trúc mạng q.....	41
Hình 3.3: Biểu đồ cơ sở dữ liệu r	41
Hình 3.4: Sơ đồ khối của chương trình s	47
Hình 3.5: Giao diện chính t.....	48
Hình 3.6: Giao diện form dữ liệu u	49
Hình 3.7: Giao diện form đồ thị phụ tải v	50
Hình 3.8: Giao diện form dự báo phụ tải w	51
Hình 3.9: Giao diện form huấn luyện phụ tải x	52
Hình 3.9: Giao diện form dự báo phụ tải y	53

TỪ KHÓA VIẾT TẮT

HTĐ	HỆ THỐNG ĐIỆN
DBPT	DỰ BÁO PHỤ TẢI

LỜI CẢM ƠN

Trong quá trình thực hiện đồ án, em xin chân thành cảm ơn sự hướng dẫn, chỉ bảo tận tình của T.S Nguyễn Hữu Quỳnh – Trưởng khoa Công nghệ thông tin trường Đại học Điện Lực là người đã hướng dẫn trực tiếp em trong suốt quá trình làm đồ án, Thầy đã dành rất nhiều thời gian quý báu để giúp em định hướng nghiên cứu đồ án.

Em xin chân thành cảm ơn các Thầy, Cô giáo trong khoa Công nghệ thông tin, cùng toàn thể các Thầy, Cô giáo trong trường Đại học Dân lập Hải phòng. Trong suốt quá trình học tập, được sự quan tâm dạy dỗ và chỉ bảo tận tình của các Thầy các Cô em đã trưởng thành lên rất nhiều.

Xin cảm ơn các bạn sinh viên lớp CT1002 Trường Đại học Dân lập Hải phòng đã ủng hộ giúp đỡ chia sẻ với tôi những kiến thức, kinh nghiệm học tập trong suốt quá trình học tập tại trường.

Hải Phòng, ngày 01 tháng 07 năm 2010.

Người thực hiện

Đỗ Thanh Tùng

LỜI NÓI ĐẦU

Dự báo là bài toán quan trọng mang lại nhiều lợi ích thiết thực phục vụ con người, nó giúp con người nắm bắt được các qui luật vận động trong tự nhiên và trong đời sống kinh tế xã hội. Trong những năm gần đây, các mạng nơron truyền thẳng nhiều lớp được thực tiễn chứng minh là khá mạnh và hiệu quả trong các bài toán dự báo và phân tích số liệu, đặc biệt trong các bài toán dự báo phụ tải, dự báo kinh tế, dự báo thời tiết,...

Cho đến nay đã có khá nhiều phương pháp luận cho hoạt động dự báo, trong đó đa số các phương pháp luận đều mang tính chất kinh nghiệm thuần túy. Dùng cách giải quyết theo kinh nghiệm vào việc dự báo là chưa đầy đủ, vì cách làm ấy hoàn toàn chỉ dựa vào kinh nghiệm của giai đoạn quá khứ, mà các kinh nghiệm ấy không phải lúc nào cũng có thể vận dụng vào hoàn cảnh đã thay đổi so với trước.

Việc nghiên cứu ứng dụng các mô hình dự báo khác nhau cho phép ta có cơ sở tiếp cận tới việc lựa chọn các phương pháp dự báo, đánh giá mức độ chính xác của dự báo đồng thời xác định khoảng thời gian lớn nhất có thể dùng dự báo.

Dự báo phụ tải ngắn hạn (giờ, ngày hoặc tuần) có vai trò quan trọng trong vận hành, lập phương thức ngày, phương thức tuần. Nếu chúng ta dự báo phụ tải quá thừa so với nhu cầu sử dụng thì hậu quả là phải huy động nguồn dự phòng đắt tiền lớn hơn mức cần thiết. Ngược lại, nếu dự báo phụ tải quá thấp so với nhu cầu thì sẽ dẫn đến kết quả nguồn dự phòng thấp, giảm an toàn cung cấp điện, không đáp ứng đủ nhu cầu điện cho các hộ tiêu thụ, và làm thiệt hại cho nền kinh tế quốc dân.

Mục đích của luận văn này là tìm hiểu, làm sáng tỏ một số khía cạnh về mạng nơron truyền thẳng nhiều lớp, thuật toán lan truyền ngược, giới thiệu phương pháp luận và một số kết quả của nghiên cứu ứng dụng mạng nơron vào trong dự báo phụ tải ngắn hạn cho hệ thống điện.

Trong đồ án này gồm có:

Chương 1: Tổng quan về dự báo phụ tải ngắn hạn

Chương 2: Mạng nơron

Chương 3: Ứng dụng mạng nơron trong dự báo phụ tải điện

CHƯƠNG 1 : TỔNG QUAN VỀ DỰ PHỤ TẢI NGẮN HẠN

1.1 Vai trò của dự báo phụ tải ngắn hạn trong công tác điều độ

Điện năng là một sản phẩm đặc biệt quan trọng đối với sự phát triển kinh tế, đời sống dân sinh và môi trường của bất cứ quốc gia nào trên thế giới. Không giống như các doanh nghiệp sản xuất các sản phẩm khác trong nền kinh tế thị trường với mục tiêu là lợi nhuận khi không có lợi nhuận thì ngừng sản xuất, điện là sản phẩm đặc biệt không thể thiếu được, nên ngành điện được coi là một ngành thuộc cơ sở hạ tầng. Điện được phân biệt với các sản phẩm hàng hoá khác nhờ khả năng đáp ứng nhanh chóng những biến đổi của nhu cầu tại mọi thời điểm và tính hầu như không thể dự trữ được. Do đó vấn đề dự báo phụ tải một cách chính xác cũng như toàn bộ các dây truyền sản xuất, truyền tải, phân phối phải luôn ở tình trạng sẵn sàng đáp ứng nhu cầu phụ tải là yếu tố cơ bản ảnh hưởng đến hiệu quả phục vụ khách hàng của toàn hệ thống.

Trong công tác vận hành, lập phương thức ngày và phương thức tuần của người làm công tác điều độ, dự báo phụ tải ngắn hạn bao gồm dự báo trước một giờ, một ngày hoặc một tuần đóng một vai trò vô cùng quan trọng. Những phương thức vận hành cơ bản trong ngày như huy động nguồn, phối hợp thủy điện và nhiệt điện, truyền tải công suất giữa các miền, giải quyết các công tác sửa chữa trên lưới điện và đánh giá mức độ an toàn hệ thống điện (HTĐ) đều đòi hỏi phải có dự báo phụ tải chính xác.

Việc dự báo phụ tải ngày HTĐ với sai số cao có ảnh hưởng rất lớn đến giá thành vận hành. Dự báo phụ tải cao hơn thực tế sẽ làm chi phí tăng cao, do huy động không cần thiết các nguồn điện dự phòng đắt tiền như diesel,.... Dự báo phụ tải thấp hơn thực tế sẽ dẫn đến kết quả nguồn dự phòng thấp, giảm độ an toàn cung cấp điện, có thể phải cắt điện do việc huy động thiếu nguồn, gây thiệt hại cả về kinh tế, an ninh xã hội, đời sống sinh hoạt của nhân dân và uy tín phục vụ của ngành đối với khách hàng.

Đối với thực tế vận hành HTĐ Việt Nam, do thành phần phụ tải sinh hoạt và dịch vụ còn chiếm tỷ lệ khá cao nên sự chênh lệch phụ tải cao điểm (phụ tải cao

nhất trong ngày P_{\max}) và phụ tải thấp điểm (phụ tải thấp nhất trong ngày P_{\min}) là rất lớn (khoảng 2.5 đến 3 lần). Đây chính là hai giá trị phụ tải đặc biệt trong đồ thị phụ tải ngày, và được người làm công tác điều độ quan tâm nhiều nhất trong quá trình vận hành. Về mặt vận hành với đồ thị phụ tải biến động lớn, thành phần thay đổi, quá trình khởi động máy diễn ra thường xuyên ảnh hưởng đến tuổi thọ và chỉ tiêu kinh tế kỹ thuật của nhà máy, tổn thất công suất và điện năng cũng sẽ gây nhiều bất lợi. Phụ tải cao điểm là nhân tố quyết định việc huy động các nguồn nhiệt điện, tua-bin khí, diesel dự phòng nhằm đảm bảo đủ nguồn phủ đỉnh và dự phòng nóng, tăng độ an toàn cung cấp điện. Phụ tải thấp điểm quyết định việc phối hợp và điều chỉnh các nguồn tua-bin khí, nhiệt điện, thủy điện nhằm đảm bảo vận hành kinh tế, tránh cho nhà máy thủy điện Hoà Bình, Thác Bà, Yaly, ... vận hành trong vùng công suất không cho phép về kỹ thuật. Trong đó phụ tải lúc cao điểm P_{\max} là khâu quan trọng nhất trong ngày.

Dự báo chính xác phụ tải cao điểm có hiệu quả lớn về kinh tế vào mùa lũ khi cần khai thác tối đa nguồn thủy điện, trong khi nước các hồ chứa đang xả thì các nguồn điện khác huy động chỉ nhằm mục đích an toàn. Nếu sai số dự báo phụ tải cao điểm giảm 2%, theo tính toán sơ bộ thì ngành điện tiết kiệm được khoảng 80MW nguồn tua-bin khí chạy dầu đắt tiền (khoảng 102.800.000đ). Việc hạn chế phụ tải lúc cao điểm cũng thường xảy ra trong mùa lũ khi thiếu nguồn phủ đỉnh.

Tính chính xác được lượng công suất thiếu để chủ động cắt phụ tải không quan trọng, tránh cắt thừa hoặc cắt thiếu, tránh sa thải phụ tải do bảo vệ tần số thấp tác động cũng có ý nghĩa lớn trong vận hành, góp phần giảm thiểu thiệt hại do việc cắt điện gây ra.

Hiện nay việc đảm bảo độ tin cậy cung cấp điện, vận hành an toàn và kinh tế HTĐ có ý nghĩa và ảnh hưởng lớn đối với ngành điện nói riêng và nền kinh tế quốc dân nói chung. Trong tương lai khi HTĐ Việt Nam liên kết với HTĐ các nước ASEAN và Trung Quốc, vấn đề tăng độ tin cậy cung cấp điện, vận hành an toàn và kinh tế HTĐ càng trở nên bức thiết. Đó không những là yếu tố đảm bảo việc nâng cao chất lượng phục vụ nhu cầu tiêu thụ điện năng trong sản xuất và đời sống sinh hoạt, góp phần vào sự tăng trưởng kinh tế của đất nước mà còn là yếu tố đảm bảo

cho Điện Lực Việt Nam hoàn thành những điều khoản cam kết trong văn bản ký kết với HTĐ các nước trong khu vực.

Giải quyết tốt vấn đề cung cấp điện có vai trò và ý nghĩa vô cùng to lớn, vì vậy việc áp dụng và nghiên cứu mô hình dự báo khác nhau nhằm tìm ra mô hình dự báo tối ưu cho vấn đề dự báo phụ tải là việc rất cần thiết.

1.2 Tầm quan trọng của dự báo phụ tải điện

Dự báo phụ tải điện (DBPT) đóng vai trò vô cùng quan trọng đối với công tác quy hoạch, đầu tư phát triển và vận hành hệ thống điện (HTĐ). Chúng ta biết rằng nhu cầu tiêu dùng điện năng phụ thuộc vào trình độ phát triển của nền kinh tế quốc dân. vì thế dự báo phụ tải điện là một bộ phận của dự báo phát triển kinh tế và khoa học kỹ thuật. Nếu chúng ta dự báo phụ tải quá thừa so với nhu cầu sử dụng thì hậu quả là phải huy động nguồn dự phòng đắt tiền lớn hơn mức cần thiết.

Ngược lại, nếu dự báo phụ tải quá thấp so với nhu cầu thì sẽ dẫn đến kết quả nguồn dự phòng thấp giảm an toàn cung cấp điện, không đáp ứng đủ nhu cầu điện cho các hộ tiêu thụ và làm thiệt hại cho nền kinh tế quốc dân.

Dự báo phụ tải dài hạn (khoảng từ 1-20 năm) có nhiệm vụ cung cấp số liệu cho công tác quy hoạch và đầu tư phát triển hệ thống điện. Còn dự báo phụ tải ngắn hạn (khoảng từ 1 -30 ngày) có nhiệm vụ đảm bảo vận hành hệ thống điện an toàn và kinh tế. Đối với dự báo dài hạn có tính chất chiến lược thì chỉ nêu nên những phương hướng phát triển chủ yếu mà không yêu cầu xác định chỉ tiêu cụ thể.

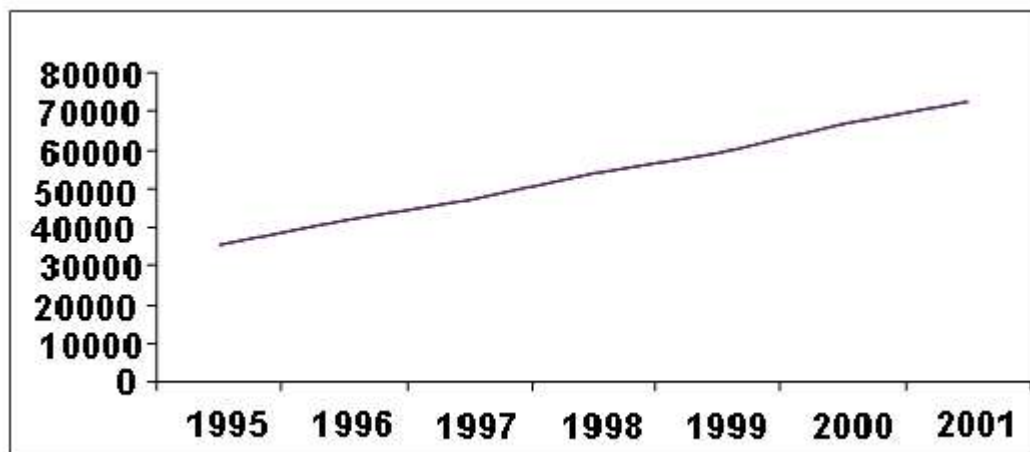
Dự báo phụ tải điện quá thừa so với nhu cầu thì dẫn đến việc huy động vốn đầu tư lớn để xây dựng nhiều nguồn phát điện, nhưng thực tế không dùng hết công suất của chúng gây lãng phí. Nếu dự báo lại quá nhỏ so với nhu cầu thực tế thì dẫn đến tình trạng thiếu nguồn điện.

Phụ tải của hệ thống điện là một hàm số phụ thuộc vào thời gian t , phụ thuộc vào thông số chế độ - tần số và mô đun điện áp thông qua những đặc tính thống kê.

Đặc tính ngẫu nhiên của phụ tải được xác định qua hai yếu tố:

- Tính ngẫu nhiên của phụ tải

- Tính ngẫu nhiên và tính chu kỳ trong hoạt động của con người trong một khoảng thời gian (ngày, tuần, tháng, năm...). Đặc điểm này tạo nên tính chu kỳ trong biến đổi của phụ tải, cho phép dự báo phụ tải với mức độ chính xác tương đối cao.



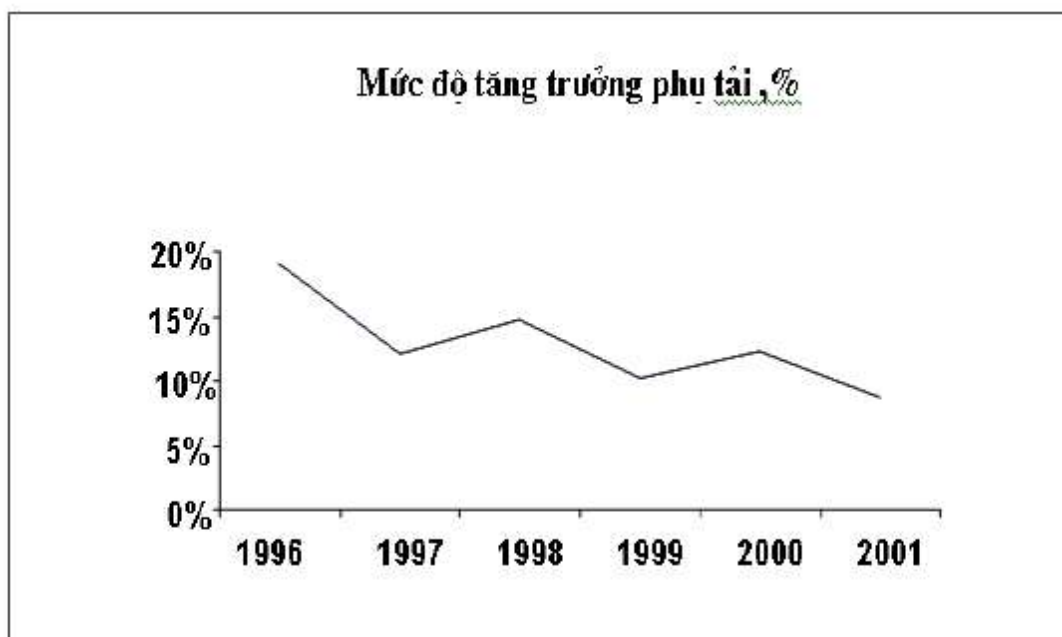
Hình 1.1: Sản lượng trung bình ngày tháng 1 qua các năm a

Chu kì tính toán T được xác định bằng các quy luật khách quan tiêu thụ điện năng. Những quy định này được xác định bằng một loạt các yếu tố :

- Thiên văn: chu kỳ quay của trái đất, mùa trong năm
- Xã hội: tuần 7 ngày với các ngày làm việc và ngày nghỉ hoặc ngày lễ, sinh hoạt trong ngày, 8 giờ làm việc, những giờ nghỉ chương trình, tivi, v.v..Nhu cầu phụ tải phụ thuộc rất nhiều vào yếu tố này.
- Thời tiết: nhiệt độ, không khí, nắng mưa. Mối quan hệ này là phi tuyến.

Tương tự như vậy chu kỳ T có thể lấy từ chu kỳ hoạt động của con người: ngày, đêm, tháng, năm. Tuy nhiên không thể coi một chu kỳ nào trong những chu kỳ trên là lặp lại của chu kỳ trước: do sự tăng trưởng của phụ tải theo thời gian, do những biến đổi bất thường của phụ tải. Tóm lại các yếu tố ảnh hưởng đến phụ tải bao gồm:

Qua đồ thị trên ta thấy tốc độ tăng trưởng của phụ tải qua các năm là khác nhau:



Hình 1.2: Tốc độ tăng trưởng của phụ tải trung bình tháng 1 qua các năm b

Để dự báo phụ tải cần phải mô tả các thông số phụ tải đã có bằng một hoặc nhiều hàm số nào đó. Các hàm số này mô tả dãy thông số quan sát một cách gần đúng. Sai số tiệm cận hoá thông số quan sát bằng hàm số được đánh giá khi tính giá trị của hàm số theo chính những đối số đó. Những sai số này có thể coi là ngẫu nhiên vì chúng không phụ thuộc vào đối số của hàm số. Phương trình mô tả dãy thông số quan sát không phải là một quan hệ gần đúng mà là quan hệ hồi quy. Như vậy sai số dự báo phụ thuộc vào chất lượng của mô hình dự báo phụ tải, nghĩa là tùy theo mô hình tính toán được lựa chọn, tùy theo mức độ chính xác đạt được khi xác định các hệ số của mô hình. Hay nói một cách khác sai số của dự báo phụ thuộc vào khả năng của mô hình tính toán có thể mô tả diễn biến của phụ tải chính xác đến mức nào. Khi xây dựng mô hình tính toán, nếu có thể xác định mối liên quan của sai số dự báo ở những thời điểm khác nhau hoặc liên quan đến một đại lượng khác tức là sai số mang tính hệ thống, thì sau khi xác định được sai số ở một thời điểm nào đó có thể ước lượng được sai số trong tương lai. Sai số này còn có thể gọi là sai số tính toán hay sai số tương đối. Phân biệt giữa sai số này và sai số quan sát (gọi là sai số thực tế hay sai số tuyệt đối). Nếu không ước lượng được sai số dự báo thì khó có thể tăng được độ chính xác của dự báo.

Sai số của dự báo phụ thuộc vào đặc tính của phụ tải, phụ thuộc vào khoảng thời gian đón trước cần tính toán. Phụ tải tương đương cần phải tính dự báo càng lớn (bao gồm nhiều phụ tải nhỏ, có tính chất khác nhau) thì sai số tương đối càng nhỏ. Sai số tính toán theo phần trăm của phụ tải tổng sẽ nhỏ hơn sai số cũng tính theo phần trăm của bất kỳ phụ tải thành phần nào.

Phụ tải công nghiệp thường ít thay đổi trong một chu kỳ và qua các chu kỳ, phụ tải sinh hoạt có thể thay đổi nhiều trong một chu kỳ và ít thay đổi qua các chu kỳ.

1.3 Các yếu tố ảnh hưởng trực tiếp đến phụ tải ngày

Mục tiêu cuối cùng của tất cả các dự báo là kết quả dự báo phải bám sát giá trị thực tế. Muốn có dự báo ngày với sai số nhỏ, phải chọn phương pháp dự báo tối ưu.

Qua các số liệu thống kê và thực tế vận hành cho thấy phụ tải HTĐ Việt Nam phụ thuộc chủ yếu vào các yếu tố sau:

- Thứ của ngày trong tuần: giá trị phụ tải giờ trong các ngày làm việc thường cao hơn phụ tải giờ của các ngày nghỉ cuối tuần.
- Thời tiết trong ngày: Thời tiết trong ngày bao gồm các thông số: nhiệt độ, độ ẩm, tốc độ gió và hướng gió, cường độ ánh sáng... trong các thông số đó nhiệt độ là yếu tố ảnh hưởng nhanh nhất đến phụ tải. Có thể xem phụ tải là một hàm số phụ thuộc vào nhiệt độ.

Giữa phụ tải và nhiệt độ có mối quan hệ là nhiệt độ tăng thì phụ tải tăng và ngược lại. Chỉ có trường hợp đặc biệt ở miền Bắc khi thời tiết quá lạnh phụ tải có thể tăng do việc sử dụng các thiết bị sưởi ấm.

- Ngày đặc biệt trong năm: Ngày đặc biệt trong năm là các ngày lễ, ngày tết, như ngày tết âm lịch, tết dương lịch, ngày 30/4 và 01/5, ngày 02/9,..Phụ tải những ngày này trên toàn quốc giảm rất nhiều so với ngày thường. Dự báo các ngày đặc biệt này không áp dụng như ngày thường mà phải theo phương pháp riêng.

- Kế hoạch sửa chữa lớn trong ngày có cắt điện phụ tải: Hiện nay, trong quá trình thực hiện các kế hoạch sửa chữa lớn, như cải tạo các trạm biến áp 220KV, cắt

điện đường dây đang vận hành để kéo đường dây mới..., vẫn phải cắt phụ tải diện rộng, dẫn đến điện năng tiêu thụ vào các ngày (giờ) này giảm một cách đáng kể. Dự báo phụ tải những ngày này phải tính toán khấu trừ phụ tải giảm do ngừng cung cấp điện. Dạng của biểu đồ phụ tải trong ngày có cắt điện diện rộng cũng mang đặc thù riêng, thay đổi tùy thuộc vào thời gian, vùng, miền cắt điện.

Qua các số liệu thống kê, quá trình nghiên cứu thực tế và kinh nghiệm vận hành, các nhà nghiên cứu, các chuyên gia điều độ đi đến kết luận: nhiệt độ môi trường (thông số quan trọng nhất của thời tiết trong ngày), và đặc thù ngày đang xét (ngày đặc biệt, ngày nghỉ cuối tuần, ngày lễ, ngày đặc biệt) là các biến ảnh hưởng mạnh nhất đến giá trị phụ tải ngày.

1.4 Phương hướng nghiên cứu dự báo phụ tải

Dự báo luôn giữ vai trò quan trọng trong nhiều lĩnh vực, quyết định kế hoạch sản xuất, hướng đầu tư phát triển trong tương lai, do đó có rất nhiều các mô hình toán học áp dụng cho dự báo. Cũng như các dự báo khác, dự báo phụ tải ngày cũng phải dựa vào số liệu thống kê, phân tích và áp dụng thuật toán để xác định mối quan hệ giữa phụ tải và các yếu tố ảnh hưởng, từ đó dự báo phụ tải dựa trên các yếu tố ảnh hưởng đó.

Trước đây các mô hình toán học truyền thống hay được áp dụng cho dự báo như: Hồi qui tuyến tính (linear regression), san trung bình (moving averages), hệ số ngẫu nhiên (stochastic), san hàm mũ (expodential smoothing), AGRIMA (Autoregressive Intergarated Moving Average)...

Nghiên cứu phương pháp dự báo phụ tải ngày là nghiên cứu và áp dụng các phương pháp để tìm quan hệ giữa phụ tải ngày và các yếu tố chủ yếu ảnh hưởng đến phụ tải ngày. Việc áp dụng mạng noron nhân tạo cho dự báo phụ tải của HTĐ Việt Nam dựa trên các yêu cầu: phương pháp đơn giản, có tính đến ảnh hưởng của biến nhiệt độ và đặc thù ngày, cho dự báo với sai số nhỏ là phương hướng của nghiên cứu này.

CHƯƠNG 2 : MẠNG NƠN

2.1 Giới thiệu về mạng nơon

2.1.1 Lịch sử phát triển

Dưới đây là các mốc đáng chú ý trong lịch sử phát triển của mạng nơon :

Cuối thế kỷ 19 đầu thế kỷ 20 sự phát triển chủ yếu chỉ là những công việc có sự tham gia của cả ba ngành Vật lý học, Tâm lý học, và thần kinh học, bởi các nhà khoa học như Hermann von Hemholtz, Ernst Mach, Ivan Pavlov. Các công trình nghiên cứu của họ chủ yếu đi sâu vào các lý thuyết tổng quát về HỌC, NHÌN, và lập luận,..và không đưa ra các mô hình toán học cụ thể mô tả hoạt động của các nơon.

Mọi chuyện thực sự bắt đầu vào những năm 1940 với công trình của Warren McCulloch và Walter Pitts. Họ chỉ ra rằng về nguyên tắc, mạng của các nơon nhân tạo có thể tính toán bất kỳ một hàm số học hay logic nào.

Tiếp theo là Donald Hebb, ông đã phát biểu rằng việc thuyết lập luận cổ điển (như Pavlov) đưa ra là hiện thực bởi do các thuộc tính của từng nơon riêng biệt, ông cũng nêu ra một phương pháp học của các nơon nhân tạo

Ứng dụng thực nghiệm đầu tiên của các nơon nhân tạo có được vào cuối những năm 50 cùng với phát minh của mạng nhận thức và luật học tương ứng bởi Frank Rosenblatt. Mạng này có khả năng nhận dạng các mẫu. Điều này đã mở ra rất nhiều hy vọng cho việc nghiên cứu mạng nơon. Tuy nhiên nó có hạn chế là chỉ có thể giải quyết một số lớp hữu hạn các bài toán.

Cùng thời gian đó, Bernard Widrow và Marcian Hoff đã đưa ra một thuật toán học mới và sử dụng nó để huấn luyện cho các mạng nơon tuyến tính thích nghi, mạng có cấu trúc và chức năng tương tự như mạng của Rosenblatt. Luật học Widrow – Hoff vẫn còn được sử dụng cho đến ngày nay.

Tuy nhiên cả Rosenblatt và Widrow – Hoff đều cùng vấp phải một vấn đề do Marvin Minsky và Seymour Papert phát hiện ra, đó là các mạng nhận thức chỉ có khả năng giải quyết được các bài toán tách được tuyến tính. Họ cố gắng cải tiến luật học và mạng để có thể vượt qua được hạn chế này nhưng họ đã không thành công trong việc cải tiến luật học để có thể huấn luyện được các mạng có cấu trúc phức tạp hơn.

Do những kết quả của Minsky Papert nên việc nghiên cứu về mạng nơron gần như bị chững lại trong suốt một thập kỷ 70 do nguyên nhân là không có được các máy tính đủ mạnh để có thể thực nghiệm.

Mặc dù vậy, vẫn có một vài phát kiến quan trọng vào những năm 70. Năm 1972, Teuvo Kohonen và James Anderson độc lập nhau phát triển một loại mạng mới có thể hoạt động như một bộ nhớ. Stephen Grossberg cũng rất tích cực trong việc khảo sát các mạng tự tổ chức.

Vào những năm 80, việc nghiên cứu mạng nơron phát triển rất mạnh mẽ cùng với sự ra đời của PC. Có hai khái niệm mới liên quan đến sự hồi sinh này, đó là:

- Việc sử dụng các phương pháp thống kê để giải thích hoạt động của một lớp các mạng hồi qui có thể được dùng như bộ nhớ liên hợp trong công trình của nhà vật lý học John Hopfield.

- Sự ra đời của thuật toán lan truyền ngược để luyện các mạng nhiều lớp được một vài nhà nghiên cứu độc lập tìm ra như: David Rumelhart, James McClelland,...đó cũng là câu trả lời cho Minsky-Papert.

2.1.2 Ứng dụng

Trong quá trình phát triển, mạng nơron đã được ứng dụng trong rất nhiều lĩnh vực:

- Tài chính: Định giá bất động sản, cho vay, kiểm tra tài sản cầm cố, đánh giá mức độ hợp tác, phân tích đường tín dụng, chương trình thương mại qua giấy tờ, phân tích tài chính liên doanh, dự báo tỷ giá tiền tệ.

- Ngân hàng: Bộ đọc séc và các tài liệu, tính tiền của thẻ tín dụng

- Giải trí: Hoạt hình, các hiệu ứng đặc biệt.

- Bảo hiểm: Đánh giá việc áp dụng chính sách, tối ưu hoá sản phẩm.

- Điện tử học: Dự báo mã tuần tự, sơ đồ chip IC, điều khiển tiến trình, phân tích nguyên nhân hỏng chip, nhận dạng tiếng nói, mô hình phi tuyến.

- Quốc phòng: Định vị - phát hiện vũ khí, dò mục tiêu, phát hiện đối tượng, nhận dạng nét mặt, các bộ cảm biến thế hệ mới, xử lý ảnh radar,..

- Tự động: Các hệ thống dẫn đường tự động cho ô tô, các bộ phân tích hoạt động của xe.
- Hàng không: Phi công tự động, giả lập đường bay, các hệ thống điều khiển lái máy bay, bộ phát hiện lỗi.
-

2.1.3 Căn nguyên sinh học

Bộ não con người chứa khoảng 1011 các phân tử liên kết chặt chẽ với nhau (khoảng 104 liên kết đôi với mỗi phân tử) gọi là các nơron. Dưới con mắt của những người làm công tác tin học, một nơron được cấu tạo bởi các thành phần : tế bào hình cây(dendrite) - tế bào thân (cell body) và sợi trục thần kinh(axon). Tế bào hình cây có nhiệm vụ mang các tín hiệu điện tới tế bào thân, tế bào thân sẽ thực hiện gộp và phân ngưỡng các tín hiệu đến. Sợi trục thần kinh làm nhiệm vụ đưa tín hiệu từ tế bào thân ra bên ngoài.

Điểm tiếp xúc giữa một sợi trục thần kinh của nơron này và tế bào hình cây của một nơron khác được gọi là khớp thần kinh (synapse). Sự sắp xếp của các nơron và mức độ mạnh yếu của các khớp thần kinh được quyết định bởi các quá trình hoá học phức tạp, sẽ thiết lập chức năng của mạng nơron.

Một vài nơron có sẵn từ khi sinh ra, các phần khác được phát triển thông qua việc học, ở đó có sự thiết lập các liên kết mới và loại bỏ các liên kết cũ.

Cấu trúc của mạng nơron luôn luôn phát triển. Các thay đổi sau này có khuynh hướng bao gồm chủ yếu là việc làm tăng hay giảm độ mạnh của các mối liên kết thông qua các khớp thần kinh.

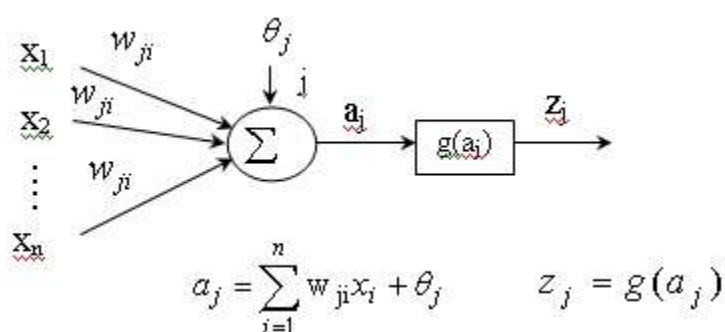
Mạng nơron nhân tạo không tiếp cận đến sự phức tạp của bộ não. Mặc dù vậy, có hai sự tương quan cơ bản giữa mạng nơron nhân tạo và sinh học. Thứ nhất, cấu trúc khối tạo thành chúng đều là các thiết bị tính toán đơn giản, được liên kết chặt chẽ với nhau. Thứ hai, các liên kết giữa các nơron quyết định chức năng của mạng.

Cần chú ý rằng mặc dù mạng nơron sinh học hoạt động rất chậm so với các linh kiện điện tử (10-3 giây so với 10-9 giây), nhưng bộ não có khả năng thực hiện nhiều công việc nhanh hơn nhiều so với các máy tính thông thường. Đó một phần là

do cấu trúc song song của mạng nơron sinh học: toàn bộ các nơron hoạt động một cách đồng thời tại một thời điểm. Mạng nơron nhân tạo cũng chia sẻ đặc điểm này. Mặc dù hiện nay, các mạng nơron chủ yếu được thực nghiệm trên các máy tính số, nhưng cấu trúc song song của chúng khiến chúng ta có thể thấy cấu trúc phù hợp nhất là thực nghiệm chúng trên các vi mạch tích hợp lớn, các thiết bị quang và các bộ xử lý song song.

2.2 Nơron nhân tạo

Một nơron nhân tạo (Hình 2.1), còn được gọi là một đơn vị xử lý hay một nút, thực hiện một chức năng : nhận tín hiệu vào từ một nguồn bên ngoài hay từ các đơn vị phía trước và tính tín hiệu ra từ các tín hiệu vào sau đó lan truyền sang các đơn vị khác.



Hình 2.1: Đơn vị xử lý thứ j c

Ở đây:

x_i : là các đầu vào

w_{ji} : các trọng số tương ứng với các đầu vào

θ_j : ngưỡng của nơron thứ j

a_j : tổng đầu vào của nơron thứ j (net input)

z_j : đầu ra của nơron thứ j

$g(\cdot)$: hàm chuyển (hàm kích hoạt)

Một nơron trong mạng có thể có nhiều đầu vào (x_1, x_2, \dots, x_n) nhưng chỉ có một đầu ra z_j . Đầu vào của một nơron có thể là từ bên ngoài mạng, hoặc đầu ra của một nơron khác, hay là đầu ra của chính nó.

Trong mạng nơron có ba kiểu nơron:

- Nơron đầu vào, nhận tín hiệu từ bên ngoài
- Nơron ẩn, tín hiệu vào và ra của nó nằm trong mạng
- Nơron đầu ra, gửi tín hiệu ra bên ngoài

2.3 Hàm xử lý

2.3.1 Hàm kết hợp

Hàm kết hợp thực hiện nhiệm vụ kết hợp các giá trị đưa vào nơron thông qua các liên kết với các nơron khác, sinh ra một giá trị gọi là net input. Tổng đầu vào nơron j là tổng trọng số của các đầu ra từ các nơron cộng thêm ngưỡng θ_j :

$$\sum_{i=0}^n w_{ji}x_i + \theta_j$$

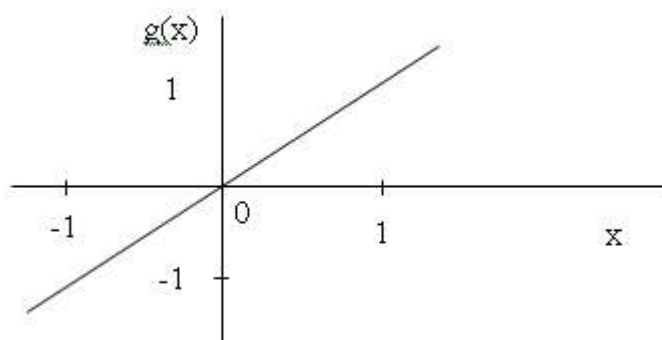
Nếu $w_{ji} > 0$ nơron được coi là ở trạng thái kích thích. Nếu $w_{ji} < 0$ nơron được coi là ở trạng thái kiềm chế.

2.3.2 Hàm kích hoạt (hàm chuyển)

Quá trình biến đổi giá trị a_j thành tín hiệu đầu ra z_j được thực hiện bởi hàm kích hoạt $z_j = g(a_j)$. Hàm kích hoạt thường bị ép vào một khoảng giá trị xác định. Các hàm kích hoạt hay được sử dụng là:

a, Hàm tuyến tính (Linear function):

$$g(x) = x$$



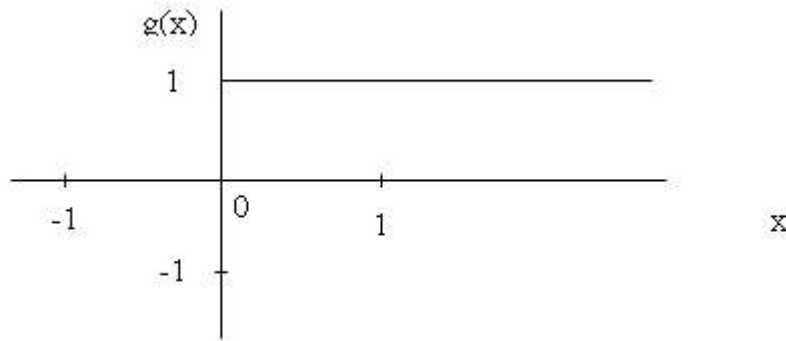
Hình 2.2: Hàm tuyến tính d

b, Hàm ranh giới cứng(Hard limiter function)

Hàm này chỉ có 2 giá trị là 0 và 1

$$g(x) = \begin{cases} 1 & \text{if } (x \geq \theta) \\ 0 & \text{if } (x < \theta) \end{cases}$$

Trong hình sau θ được chọn bằng 0

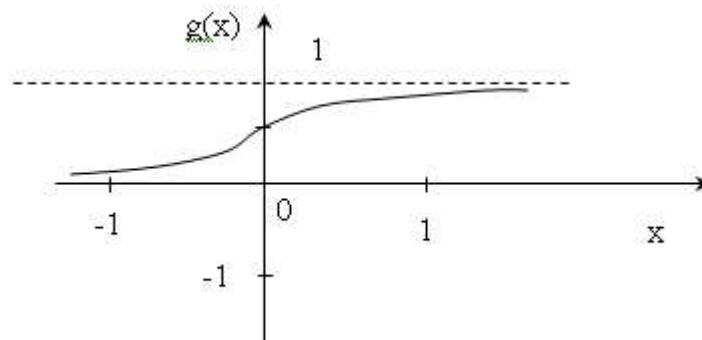


Hình 2.3: Hàm ranh giới cứng e

c, Hàm Sigmoid (Sigmoid function)

$$g(x) = \frac{1}{1 + e^{-x}}$$

Hàm này rất thuận lợi khi sử dụng cho các mạng được huấn luyện bằng thuật toán Lan truyền ngược BP, bởi lẽ nó dễ lấy đạo hàm, do đó nó có thể giảm đáng kể tính toán trong quá trình huấn luyện. Hàm này được ứng dụng cho các bài toán mà các đầu ra mong muốn rơi vào khoảng $[0,1]$.

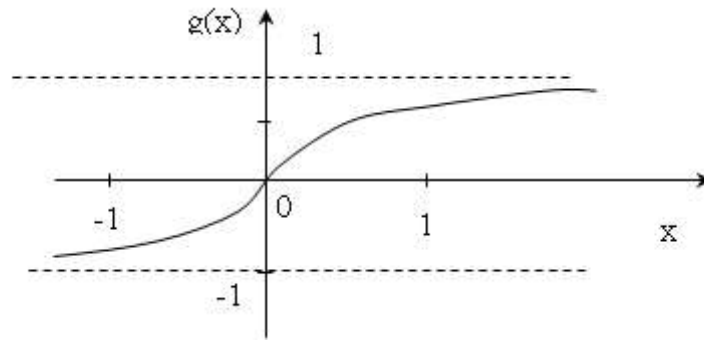


Hình 2.4: Hàm Log – Sigmoid f

d, Hàm Sigmoid lưỡng cực (Bipolar Sigmoid function)

Hàm này có các đặc trưng tương tự hàm Sigmoid. Nhưng được ứng dụng cho các bài toán có đầu ra mong muốn rơi vào khoảng $[-1,1]$

$$g(x) = \frac{1 - e^{-x}}{1 + e^{-x}}$$



Hình 2.5: Hàm Sigmoid lưỡng cực g

Hàm kích hoạt của các nơron ẩn là cần thiết để biểu diễn sự phi tuyến vào trong mạng. Lý do là hợp của các hàm đồng nhất là một hàm đồng nhất. Mặc dù vậy nó mang tính chất phi tuyến làm cho các mạng nhiều tầng có khả năng biểu diễn các ánh xạ phi tuyến rất tốt. Tuy nhiên đối với luật học lan truyền ngược, hàm phải khả vi và sẽ có ích nếu hàm được gán trong một khoảng nào đó. Do vậy hàm Sigmoid là một lựa chọn thông dụng nhất.

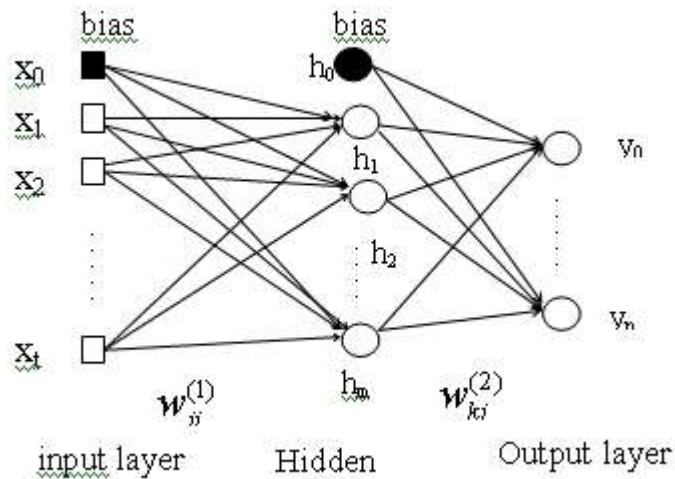
Hàm kích hoạt của các nơron đầu ra được chọn sao cho phù hợp với sự phân phối của các giá trị đích mong muốn.

2.4 Cấu trúc của mạng nơron

Cấu trúc của mạng nơron được xác định bởi : số lớp, số nơron trên mỗi lớp, và sự liên kết giữa các nơron. Dựa trên cách thức liên kết các nơron người ta chia làm hai loại:

2.4.1 Mạng truyền thẳng

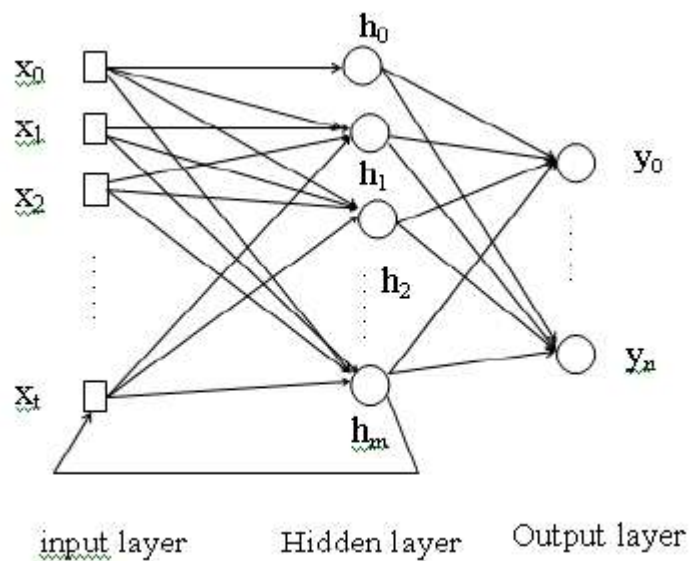
Dòng dữ liệu từ nơron đầu vào đến nơron đầu ra chỉ được truyền thẳng. Không có các liên kết từ nơron đầu ra đến các nơron đầu vào trong cùng một lớp hay của các lớp trước đó.



Hình 2.6: Mạng nơ-ron truyền thẳng nhiều lớp h

2.4.2 Mạng hồi qui

Có chứa các liên kết ngược, nghĩa là các nơ-ron lớp ra tới nơ-ron lớp vào trong cùng một lớp hoặc các lớp trước đó



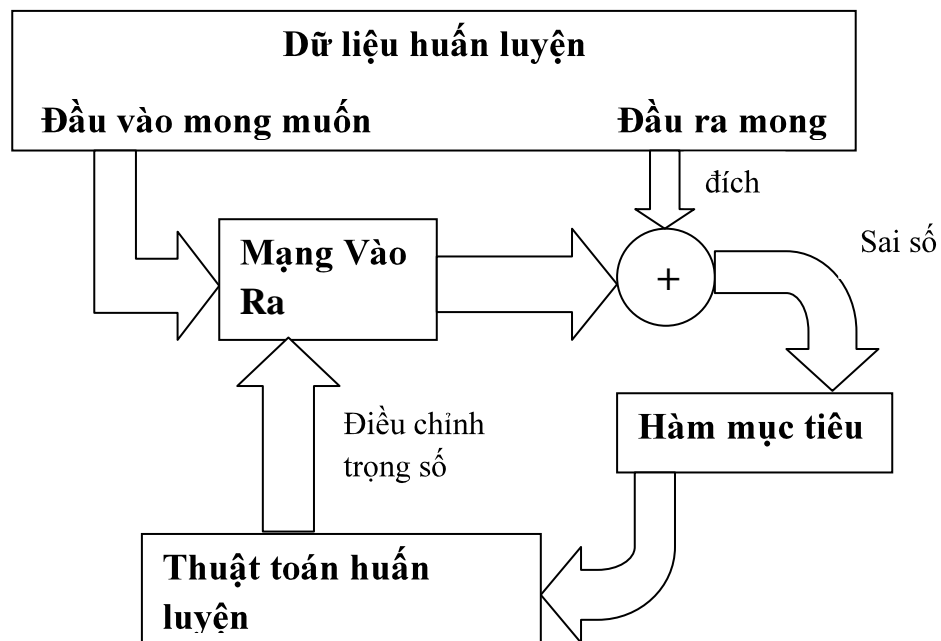
Hình 2.7: Mạng hồi qui i

2.5 Luật học

Thành phần quan trọng nhất và không thể thiếu của mạng nơ-ron đó là luật học, được dùng để điều chỉnh các trọng số để mạng nhận biết được quan hệ giữa đầu vào và đích mong muốn. Có nhiều thuật toán học đã được phát minh để tìm ra

tập trọng số tối ưu để làm giải pháp cho các bài toán. Các thuật toán đó có thể chia làm hai nhóm chính: học có thầy và học không có thầy.

- **Học có thầy:** Mạng được huấn luyện dựa trên tập ví dụ huấn luyện (các cặp mẫu đầu vào x và đầu ra mong muốn t). Sự khác biệt giữa các đầu ra thực tế với đầu ra mong muốn được thuật toán sử dụng để điều chỉnh các trọng số. Điều này thường được đưa ra như một bài toán xấp xỉ hàm số - cho dữ liệu huấn luyện, mục đích là tìm ra hàm $f(x)$ thỏa mãn tất cả các mẫu học đầu vào.



Hình 2.8: Mô hình học có thầy j

- **Học không có thầy:**

Với phương pháp học không có thầy không có phản hồi từ môi trường để chỉ ra rằng đầu ra của mạng là đúng hay sai. Mạng sẽ phải khám phá các đặc trưng, các điều chỉnh, các mối tương quan, hay các lớp trong dữ liệu vào một cách tự động. Trong thực tế, phần lớn các biến thể của học không có thầy, các đích trùng với đầu vào. Nói cách khác, học không có thầy luôn thực hiện một công việc tương tự như một mạng tự liên hợp, cô đọng thông tin từ dữ liệu vào.

2.6 Hàm mục tiêu

Để huấn luyện một mạng và xét xem nó thực hiện tốt đến đâu, ta cần xây dựng một hàm mục tiêu để cung cấp cách thức đánh giá khả năng của hệ thống. Việc chọn hàm mục tiêu là rất quan trọng bởi vì hàm này thể hiện các mục tiêu thiết

kế và quyết định thuật toán huấn luyện nào có thể được sử dụng. Để xây dựng một hàm mục tiêu đo được chính xác cái chúng ta muốn không phải là việc dễ dàng. Trong số những hàm cơ bản được sử dụng rộng rãi là hàm tổng bình phương lỗi (sum of squares error function)

$$E = \frac{1}{NQ} \sum_{q=1}^Q \sum_{i=1}^N (t_{qi} - y_{qi})^2$$

Ở đây:

Q : số ví dụ trong tập ví dụ huấn luyện

N : số đơn vị đầu ra

t_{qi} : đầu ra mong muốn của neuron ra i trên ví dụ thứ q

y_{qi} : đầu ra thực tế của neuron ra i trên ví dụ thứ q

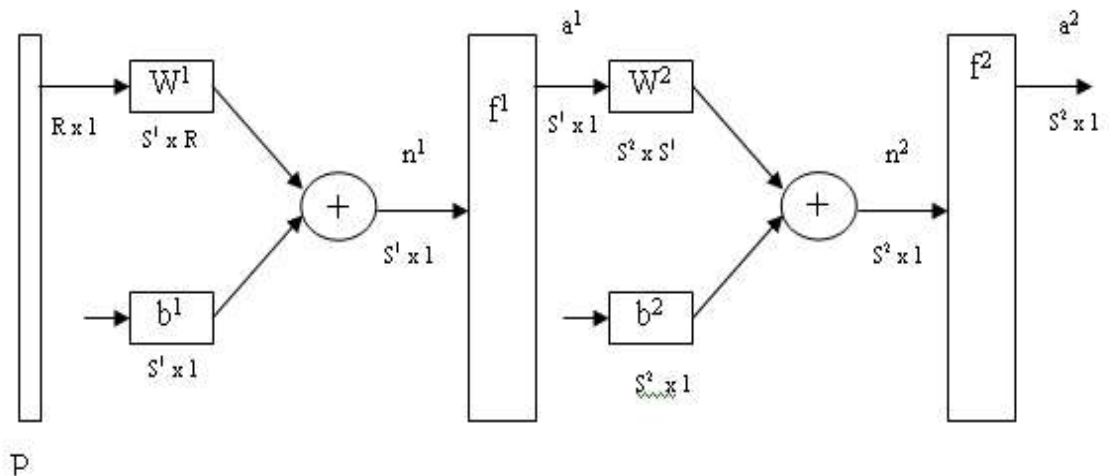
CHƯƠNG 3 : ỨNG DỤNG MẠNG NƠON TRONG DỰ BÁO PHỤ TẢI ĐIỆN

3.1 Mạng nơon truyền thẳng

3.1.1 Kiến trúc cơ bản

Kiến trúc mạng truyền thẳng nhiều lớp là kiến trúc chủ đạo của các mạng nơon hiện tại. Mặc dù có khá nhiều biến thể nhưng đặc trưng của kiến trúc này là cấu trúc và thuật toán học là đơn giản và nhanh (Masters 1993).

Mạng truyền thẳng nhiều lớp bao gồm một lớp vào, một hoặc nhiều lớp ẩn và một lớp ra. Các nơon đầu vào thực chất không phải là nơon theo đúng nghĩa, bởi lẽ chúng không thực hiện bất kỳ một tính toán nào trên dữ liệu vào, đơn giản nó chỉ tiếp nhận các dữ liệu vào và chuyển cho các lớp kế tiếp. Các nơon ở lớp ẩn và lớp ra mới thực sự thực hiện các tính toán, kết quả được định dạng bởi hàm kích hoạt của nơon đầu ra.



Hình 3.1: Mạng nơon truyền thẳng nhiều lớp k

Ở đây:

P : ma trận đầu vào (R hàng, 1 cột)

W^i : ma trận trọng số của các nơon lớp thứ i

b^i : ma trận độ lệch của lớp thứ i

n^i : tổng liên kết đầu vào (net input)

f^i : hàm kích hoạt của lớp thứ i

a^i : ma trận đầu ra của lớp thứ i

\oplus : Hàm tính tổng liên kết đầu vào

Mỗi liên kết gắn với một trọng số, trọng số này được điều chỉnh trong quá trình tín hiệu đi qua liên kết đó. Trọng số dương thể hiện trạng thái kích thích, các trọng số âm thể hiện trạng thái kiềm chế. Mỗi nơron tính toán mức kích hoạt của chúng bằng cách tính tổng các đầu vào và đưa đến hàm kích hoạt. Khi tất cả các nơron trong một lớp mạng cụ thể đã tính toán xong đầu ra thì lớp kế tiếp có thể bắt đầu thực hiện, bởi vì đầu ra của lớp hiện tại chính là đầu vào của lớp kế tiếp. Khi tất cả các nơron của mạng đã thực hiện xong tính toán thì đầu ra của nơron lớp ra là kết quả của mạng. Tuy nhiên, có thể chưa là kết quả mong muốn, khi đó ta cần điều chỉnh các trọng số bằng cách áp dụng một thuật toán huấn luyện.

Mạng có nhiều lớp có khả năng tốt hơn là các mạng chỉ có một lớp, chẳng hạn như mạng hai lớp với lớp thứ nhất sử dụng hàm Sigmoid và lớp thứ hai sử dụng hàm tuyến tính có thể áp dụng để xấp xỉ các hàm toán học khá tốt, trong khi các mạng chỉ có một lớp thì không có khả năng này

Chẳng hạn, đầu ra của lớp thứ 2 là:

$$a^2 = f^2(n^2) = f^2(a^1 * W^2 + b^2) = f^2(W^2(f^1(PW^1 + b^1)) + b^2)$$

3.1.2 Khả năng biểu diễn

Mạng một lớp có khả năng biểu diễn các hàm tách được tuyến tính, chẳng hạn hàm logic AND có thể phân chia được bằng một đường thẳng trong khi miền giá trị của hàm XOR thì không.

Mạng có hai lớp ẩn có khả năng biểu diễn một đường phân chia tùy ý với một độ chính xác bất kỳ với các hàm kích hoạt phân ngưỡng và có thể xấp xỉ bất kỳ ánh xạ mịn nào với độ chính xác bất kỳ với các hàm kích hoạt có dạng Sigmoid.

Mạng có một lớp ẩn có thể xấp xỉ tốt bất kỳ một ánh xạ liên tục nào từ một không gian hữu hạn sang một không gian hữu hạn khác, chỉ cần cung cấp đủ số nơron trong lớp ẩn.

3.1.3 Vấn đề thiết kế cấu trúc mạng

Về mặt lý thuyết có tồn tại một mạng có thể mô phỏng một bài toán với độ chính xác bất kỳ. Nhưng không phải đơn giản để có thể tìm ra được mạng này. Với một bài toán cụ thể để xác định chính xác một kiến trúc mạng như: cần bao nhiêu lớp ẩn, mỗi lớp ẩn cần bao nhiêu nơron là một công việc hết sức khó khăn.

3.1.4 Số lớp ẩn

Về mặt lý thuyết, các mạng có hai lớp ẩn có thể biểu diễn các hàm với dáng điệu bất kỳ, nên không có lý do nào sử dụng mạng có nhiều hơn hai lớp ẩn. Người ta đã xác định rằng đối với phần lớn các bài toán cụ thể, chỉ cần sử dụng một lớp ẩn là đủ. Các bài toán sử dụng hai lớp ẩn hiếm khi xảy ra trong thực tế. Việc huấn luyện mạng thường rất chậm khi sử dụng nhiều lớp ẩn. Bởi lẽ:

Phần lớn các thuật toán luyện mạng cho các mạng nơron truyền thẳng đều dựa trên phương pháp gradient. Các lớp thêm vào sẽ thêm việc phải lan truyền các lỗi làm cho vectơ gradient rất không ổn định. Sự thành công của bất kỳ một thuật toán tối ưu theo gradient phụ thuộc vào độ không thay đổi của hướng khi mà các tham số thay đổi.

Số các cực trị địa phương tăng lên rất lớn khi có nhiều lớp ẩn. Phần lớn các thuật toán tối ưu dựa trên gradient chỉ có thể tìm ra các cực trị địa phương, do vậy chúng có thể không tìm ra cực trị toàn cục. Mặc dù thuật toán luyện mạng có thể tìm ra cực trị toàn cục, nhưng xác suất khá cao là chúng ta sẽ bị mắc trong một cực trị địa phương.

Tuy nhiên, có thể đối với một bài toán cụ thể, sử dụng nhiều hơn một lớp ẩn với số ít nơron thì tốt hơn là sử dụng ít lớp ẩn với số nơron là lớn, đặc biệt đối các mạng cần phải học các hàm không liên tục. Về tổng thể người ta cho rằng việc đầu tiên là nên xem xét mạng chỉ có một lớp ẩn. Nếu dùng một lớp ẩn với một lượng lớn các nơron mà không có hiệu quả thì nên sử dụng thêm một lớp ẩn nữa với một số ít các nơron.

3.1.5 Số nơron trong lớp ẩn

Vấn đề quan trọng trong thiết kế một mạng là mỗi tầng cần có bao nhiêu

noron. Số noron trong mỗi lớp quá ít có thể dẫn đến việc không thể nhận dạng được tín hiệu đầy đủ trong một tập dữ liệu phức tạp, hay thiếu ăn khớp. Sử dụng quá nhiều noron sẽ tăng thời gian luyện mạng, có lẽ là quá nhiều để luyện khi mà không thể luyện mạng trong một thời gian hợp lý. Số lượng lớn các noron có thể dẫn đến tình trạng quá khớp, trong trường hợp này mạng có quá nhiều thông tin, hoặc lượng thông tin trong tập ví dụ huấn luyện không đủ các dữ liệu đặc trưng để huấn luyện mạng.

Số lượng thích hợp các noron ẩn phụ thuộc vào rất nhiều yếu tố như: số đầu vào, số đầu ra của mạng, số trường hợp trong tập mẫu, độ nhiễu của dữ liệu đích, độ phức tạp của hàm lỗi, kiến trúc mạng và thuật toán luyện mạng.

Có rất nhiều luật để lựa chọn số noron trong các lớp ẩn, chẳng hạn:

$m \in [t, n]$: giữa khoảng kích thước lớp vào, lớp ra

$m = \frac{2(t+n)}{3}$: 2/3 tổng kích thước lớp vào và lớp ra

$m < 2t$: nhỏ hơn hai lần kích thước lớp vào

$m = \sqrt{t \cdot n}$: căn bậc hai của tích kích thước lớp vào và lớp ra

Các luật này chỉ được coi như là các lựa chọn thô khi chọn lựa kích thước của các lớp. Chúng không phản ánh được thực tế, bởi lẽ chúng chỉ xem xét đến nhân tố kích thước đầu vào, đầu ra mà bỏ qua các nhân tố quan trọng khác như: số trường hợp đưa vào huấn luyện, độ nhiễu ở các đầu ra mong muốn, độ phức tạp của hàm lỗi, kiến trúc của mạng và thuật toán học.

Trong phần lớn các trường hợp, không có một cách để có thể dễ dàng xác định được số tối ưu các đơn vị trong lớp ẩn mà cách tốt nhất là dùng phương pháp Thử - Sai. Trong thực tế có thể lựa chọn phương pháp Lựa chọn tiến hay Lựa chọn lùi để xác định số đơn vị trong lớp ẩn.

Lựa chọn tiến bắt đầu với việc chọn một luật hợp lý cho việc đánh giá hiệu năng của mạng. Sau đó, ta chọn một số nhỏ các noron ẩn, luyện và thử mạng, ghi lại hiệu năng của mạng. Tiếp theo tăng một số ít các noron ẩn, luyện và thử lại cho đến khi lỗi là chấp nhận được, hoặc không có tiến triển đáng kể so với trước.

Ngược lại, Lựa chọn lùi bắt đầu với một số lớn nơron trong lớp ẩn, sau đó giảm dần. Quá trình này rất mất thời gian nhưng sẽ giúp ta tìm được số lượng nơron phù hợp cho lớp ẩn.

3.2 Thuật toán lan truyền ngược(Back – Propagation Algorithm)

Phần này mô tả thuật toán học sử dụng để điều chỉnh hiệu năng của mạng sao cho mạng có khả năng sinh ra được các kết quả mong muốn.

Các mạng nơron truyền thẳng nhiều lớp được huấn luyện bằng phương pháp học có thầy. Phương pháp này căn bản dựa trên việc yêu cầu mạng thực hiện chức năng của nó và sau đó trả lại kết quả, kết hợp kết quả này với các đầu ra mong muốn để điều chỉnh các tham số của mạng.

Về căn bản, thuật toán lan truyền ngược là dạng tổng quát của thuật toán trung bình bình phương tối thiểu(Least Means Square - LMS) bởi lẽ nó cùng sử dụng kỹ thuật giảm theo hướng véctơ gradient nhưng với độ phức tạp của hàm lỗi lớn hơn. Thuật toán này thuộc dạng thuật toán xấp xỉ để tìm các điểm mà tại đó hiệu năng của mạng là tối ưu. Chỉ số tối ưu (performance index) thường được xác định bởi một hàm số của ma trận trọng số và các đầu vào nào đó mà trong quá trình tìm hiểu bài toán đặt ra.

3.2.1 Mô tả thuật toán

Ta sẽ sử dụng dạng tổng quát của mạng nơron truyền thẳng nhiều lớp như ở hình 3.1. Khi đó đầu ra của một lớp trở thành đầu vào của lớp kế tiếp:

$$a^{m+1} = f^{m+1}(W^{m+1} a^m + b^{m+1}) \quad \text{với } m=0,1,\dots,M-1$$

ở đây M là số lớp trong mạng, các nơron lớp thứ nhất nhận tín hiệu từ bên ngoài:

$$a^0 = p$$

đầu ra của lớp cuối cùng là đầu ra của mạng

$$a = a^m$$

Chỉ số hiệu năng (performance index)

Thuật toán lan truyền ngược sử dụng chỉ số hiệu năng là trung bình bình phương lỗi của đầu ra so với giá trị đích. Đầu vào của mạng là tập ví dụ huấn luyện:

$$\{(p_1, t_1), (p_2, t_2), \dots, (p_q, t_q)\}$$

ở đây p_i là đầu vào và t_i là đầu ra đích tương ứng (với $i=1, 2, \dots, q$). Mỗi đầu vào đưa vào mạng, sau khi tính toán cho đầu ra, đầu ra này được đem so sánh với đầu ra mong muốn. Thuật toán sẽ điều chỉnh các tham số của mạng để tối thiểu hoá trung bình bình phương lỗi:

$$F(x) = E|e^2| = E|(t-a)^2|$$

ở đây x là biến được tạo thành bởi các trọng số và độ lệch, E là ký hiệu kỳ vọng toán học. Nếu như mạng có nhiều đầu ra, ta có thể viết lại phương trình trên ở dạng ma trận như sau:

$$F(x) = E|e^T e| = E|(t - a)^T (t - a)|$$

Ký hiệu $\hat{F}(x)$ là giá trị xấp xỉ của $F(x)$ thì ta có xấp xỉ của trung bình bình phương lỗi như sau:

$$\hat{F}(x) = (t(k) - a(k))^T (t(k) - a(k)) = e^T(k)e(k)$$

Trong đó kỳ vọng toán học của bình phương lỗi được thay bởi bình phương lỗi tại bước thứ k .

Thuật toán giảm theo hướng cho trung bình bình phương lỗi xấp xỉ là:

$$w_{ji}^m(k+1) = w_{ji}^m(k) - \alpha \frac{\partial \hat{F}}{\partial w_{ji}^m} \quad (1)$$

$$b_j^m(k+1) = b_j^m(k) - \alpha \frac{\partial \hat{F}}{\partial b_j^m} \quad (2)$$

ở đây α là hệ số học

Luật xích (Chain Rule)

Đối với mạng nơron truyền thẳng nhiều lớp, lỗi không chỉ là một hàm của chỉ các trọng số trong lớp ẩn, do vậy việc tính các đạo hàm từng phần này là không đơn giản. Chính vì lý do đó mà ta phải sử dụng luật xích để tính. Luật này được mô tả như sau:

Giả sử ta có f là hàm của biến n , ta muốn đạo hàm của f có liên quan đến một biến w khác, luật xích này như sau:

$$\frac{df(n(w))}{dw} = \frac{df(n)}{dn} \cdot \frac{dn(w)}{dw}$$

Vậy đạo hàm trong (1) và (2) sẽ là :

$$\frac{\partial \hat{F}}{w_{ji}^m} = \frac{\partial \hat{F}}{\partial n_j^m} \cdot \frac{\partial n_j^m}{\partial w_{ji}^m} \quad (1')$$

$$\frac{\partial \hat{F}}{b_j^m} = \frac{\partial \hat{F}}{\partial n_j^m} \cdot \frac{\partial n_j^m}{\partial b_j^m} \quad (2')$$

Trong đó hạng thức thứ hai ở vế phải của phương trình (1') và (2') được tính một cách dễ dàng, bởi vì đầu vào của lớp m của mạng là một hàm của trọng số và độ lệch trong tầng đó:

$$n_j^m = \sum_{i=1}^{s^{m-1}} w_{ji}^m a_i^{m-1} + b_j^m$$

Do vậy

$$\frac{\partial n_j^m}{\partial w_{ji}^m} = a_i^{m-1}; \quad \frac{\partial n_j^m}{\partial b_j^m} = 1$$

Ta ký hiệu $S_j^m = \frac{\partial \hat{F}}{\partial n_j^m}$ là độ nhạy cảm của \hat{F} đối với các thay đổi của phần tử thứ j của đầu vào của mạng tại lớp thứ m . Khi đó ta có:

$$\frac{\partial \hat{F}}{w_{ji}^m} = \frac{\partial \hat{F}}{\partial n_j^m} \cdot \frac{\partial n_j^m}{\partial w_{ji}^m} = S_j^m a_i^{m-1}$$

$$\frac{\partial \hat{F}}{b_j^m} = \frac{\partial \hat{F}}{\partial n_j^m} \cdot \frac{\partial n_j^m}{\partial b_j^m} = S_j^m$$

Thuật toán giảm theo hướng được biểu diễn như sau:

$$w_{ji}^m(k+1) = w_{ji}^m(k) - \alpha S_j^m a_i^{m-1}$$

$$b_j^m(k+1) = b_j^m(k) - \alpha S_j^m$$

Ở dạng ma trận :

$$W^m(k+1) = W^m(k) - \alpha s^m (a^{m-1})^T$$

$$b^m(k+1) = b^m(k) - \alpha s^m$$

Ở đây :

$$s^m = \frac{\partial \hat{F}}{\partial n^m} = \begin{bmatrix} \frac{\partial \hat{F}}{\partial n_1^m} \\ \frac{\partial \hat{F}}{\partial n_2^m} \\ \vdots \\ \frac{\partial \hat{F}}{\partial n_{s^m}^m} \end{bmatrix}$$

Lan truyền ngược độ nhạy cảm

Bây giờ ta cần tính toán nốt ma trận độ nhạy cảm s^m . Để thực hiện điều này cần sử dụng một áp dụng khác của luật xích. Quá trình này cho ta khái niệm về sự “Lan truyền ngược” bởi vì nó mô tả mối quan hệ hồi qui trong đó độ nhạy cảm s^m được tính qua độ nhạy cảm s^{m+1} của lớp $m+1$.

Để dẫn đến quan hệ đó, ta sẽ sử dụng ma trận Jacobian:

$$\frac{\partial n^{m+1}}{\partial n^m} = \begin{bmatrix} \frac{\partial n_1^{m+1}}{\partial n_1^m} & \frac{\partial n_1^{m+1}}{\partial n_2^m} & \cdots & \frac{\partial n_1^{m+1}}{\partial n_{s^m}^m} \\ \frac{\partial n_2^{m+1}}{\partial n_1^m} & \frac{\partial n_2^{m+1}}{\partial n_2^m} & \cdots & \frac{\partial n_2^{m+1}}{\partial n_{s^m}^m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial n_{s^{m+1}}^{m+1}}{\partial n_1^m} & \frac{\partial n_{s^{m+1}}^{m+1}}{\partial n_2^m} & \cdots & \frac{\partial n_{s^{m+1}}^{m+1}}{\partial n_{s^m}^m} \end{bmatrix}$$

Xét thành phần thứ j,i của ma trận trên :

$$\begin{aligned} \frac{\partial n_j^{m+1}}{\partial n_i^m} &= \frac{\partial (\sum_{j=1}^{s^m} w_{ji}^{m+1} a_j^m + b_j^{m+1})}{\partial n_i^m} = w_{ji}^{m+1} \frac{\partial a_i^m}{\partial n_i^m} = w_{ji}^{m+1} \frac{\partial f^m(n_i^m)}{\partial n_i^m} \\ &= w_{ji}^{m+1} f^m(n_i^m) \end{aligned}$$

Ở đây :

$$f^m(n_i^m) = \frac{\partial f^m(n_i^m)}{\partial_i^m}$$

Như vậy ma trận Jacobian có thể viết lại như sau:

$$\frac{\partial n^{m+1}}{\partial n^m} = W^{m+1} F^m(n^m)$$

Ở đây :

$$F^m(n^m) = \begin{bmatrix} f^m(n_1^m) & 0 \dots & 0 \\ 0 & f^m(n_2^m) & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & f^m(n_{s^m}^m) \end{bmatrix}$$

Bây giờ ta viết lại quan hệ hồi qui cho độ nhạy cảm dưới dạng ma trận:

$$\begin{aligned} s^m &= \frac{\partial \hat{F}}{\partial n^m} = \left(\frac{\partial n^{m+1}}{\partial n^m} \right)^T \frac{\partial \hat{F}}{\partial n^{m+1}} = \hat{F}^m(n^m) (W^{m+1})^T \frac{\partial \hat{F}}{\partial n^{m+1}} \\ &= F^m(n^m) (W^{m+1})^T s^{m+1} \end{aligned}$$

Bây giờ ta có thể thấy độ nhạy cảm được lan truyền ngược qua mạng từ lớp cuối cùng trở về lớp đầu tiên:

$$S^M \rightarrow S^{M-1} \rightarrow \dots \rightarrow S^2 \rightarrow S^1$$

Cần nhấn mạnh rằng ở đây thuật toán lan truyền ngược lỗi sử dụng cùng một kỹ thuật giảm theo hướng như thuật toán LMS. Sự phức tạp duy nhất là ở chỗ để tính gradient ta cần phải lan truyền ngược độ nhạy cảm từ các lớp về các lớp trước.

Bây giờ ta cần biết điểm bắt đầu lan truyền ngược, xét độ nhạy cảm s^m tại lớp cuối cùng:

$$s_j^m = \frac{\partial \hat{F}}{\partial n_j^m} = \frac{\partial (t - a)^T (t - a)}{\partial n_j^m} = \frac{\partial \sum_{i=1}^{s^M} (t_i - a_i)^2}{\partial n_j^M} = -2(t_j - a_j) \frac{\partial a_j}{\partial n_j^M}$$

Do

$$\frac{\partial a_j}{\partial n_j^M} = \frac{\partial a_j^M}{\partial n_j^M} = \frac{\partial f^M(n_j^M)}{\partial n_j^M} = f^M(n_j^M)$$

Nên ta có thể viết :

$$S_j^M = -2(t_j - a_j)f^M(n_j^M)$$

Biểu diễn dưới dạng ma trận :

$$S^M = -2F^M(n^M)(t - a)$$

Tóm lại, thuật toán lan truyền ngược có thể biểu diễn như sau :

**THUẬT TOÁN LAN TRUYỀN NGƯỢC – BACK PROPAGATION
ALGORITHM**

Bước 1: Lan truyền xuôi đầu vào qua mạng

$$a^0 = p$$

$$a^{m+1} = f^{m+1} (W^{m+1}a^m + b^{m+1}) \quad \text{với } m = 0, 1, \dots, M-1$$

$$a = a^M$$

Bước 2: Lan truyền độ nhạy cảm (lỗi) ngược lại qua mạng:

$$s^M = -2F^M (n^M)(t - a)$$

$$s^m = F^m (n^m)(W^{m+1})^T s^{m+1} \quad \text{với } m = M-1, M-2, \dots, 1$$

Bước 3: Các trọng số và độ lệch được cập nhật bởi công thức sau:

$$W^m(k+1) = W^m(k) - \alpha s^m (a^{m-1})^T$$

$$b^m(k+1) = b^m(k) - \alpha s^m$$

3.2.2 Sử dụng thuật toán lan truyền ngược

Ở phần này ta sẽ bàn về các khía cạnh ứng dụng của thuật toán lan truyền ngược như chọn lựa cấu trúc mạng, sự hội tụ và khả năng tổng quát hoá.

Chọn lựa cấu trúc mạng

Thuật toán lan truyền ngược có thể được sử dụng để xấp xỉ bất kỳ một hàm số học nào nếu như ta có đủ số nơron trong các lớp ẩn. mặc dù vậy, phát biểu trên chưa cho ta được một số cụ thể các lớp và số nơron trong mỗi lớp cần sử dụng.

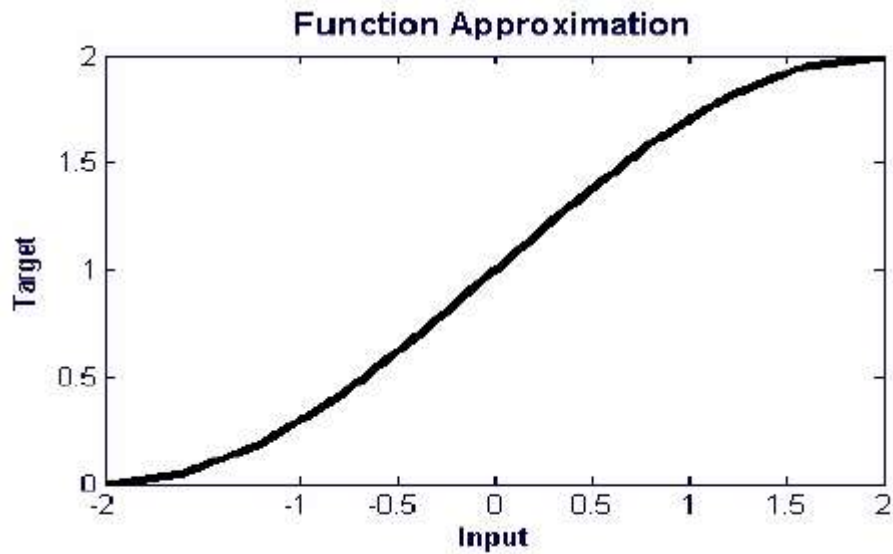
Ta sẽ dùng một ví dụ để có cái nhìn chi tiết hơn về vấn đề này. Ví dụ, ta muốn xấp xỉ hàm số sau:

$$f(x) = 1 + \sin\left(\frac{i\pi}{4}x\right) \quad \text{với } -2 \leq x \leq 2$$

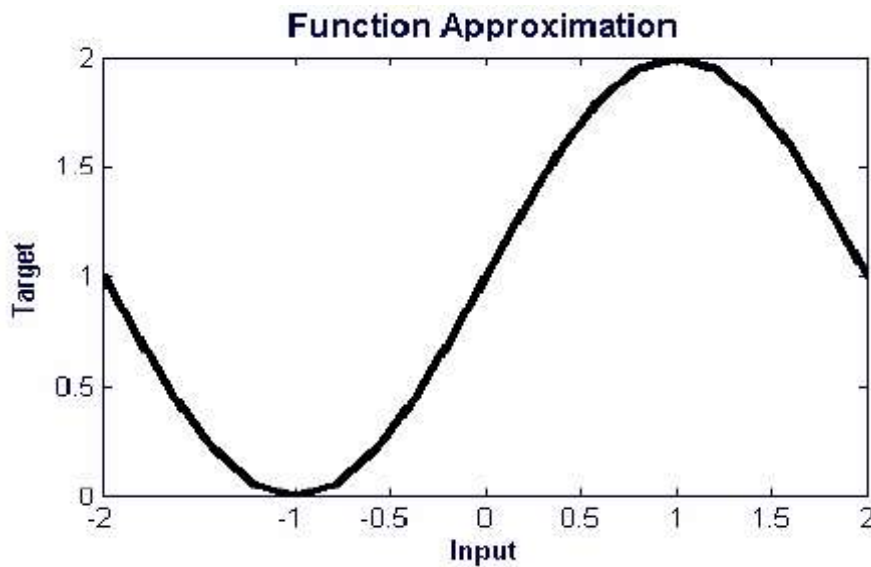
Trong đó i nhận các giá trị 1, 2, 4 và 8. Khi i tăng thì hàm số cần xét sẽ trở nên phức tạp hơn do ta sẽ nhận được nhiều chu kỳ của hình sin trong phạm vi $[-2, 2]$. Khi đó, mạng nơron với một số nơron cố định sẽ khó có thể xấp xỉ được hàm nếu i

tăng. Ta sẽ sử dụng ví dụ Function approximation trong thư viện của bộ Matlab 6.0 (tệp `nn11fa.m`). Ở đây mạng sử dụng có một lớp ẩn, một lớp ra, mạng có một đầu vào và một đầu ra.

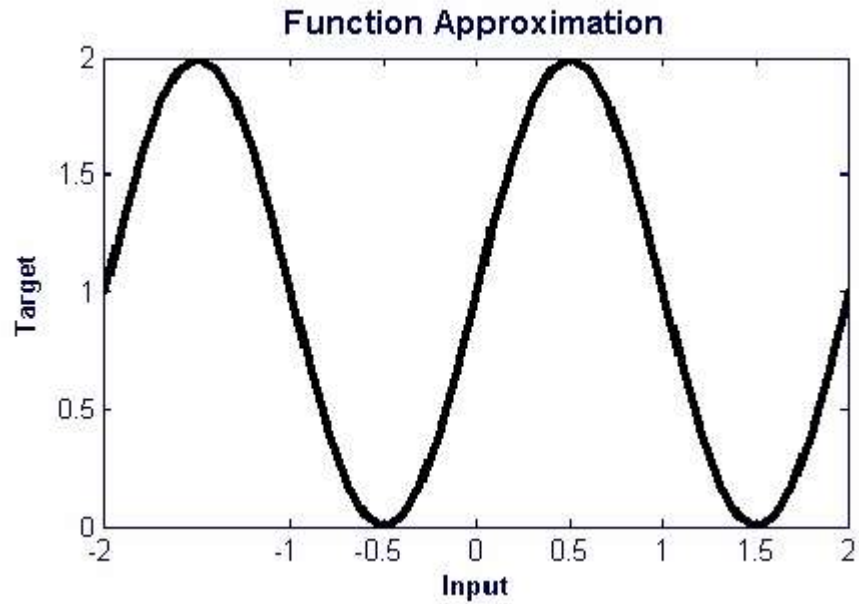
Lớp ẩn sử dụng hàm Sigmoid, lớp ra dùng hàm tuyến tính. Số nơron trong lớp ẩn là 3, kết quả xấp xỉ của mạng trong các trường hợp $i=1,2,4,8$ như trong các hình dưới đây.



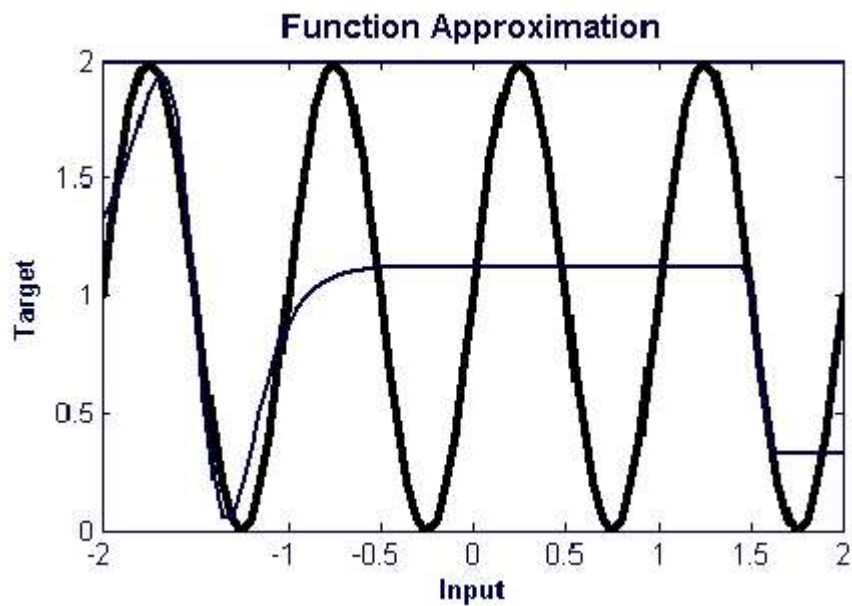
a, Trường hợp $i=1$



b, Trường hợp $i=2$



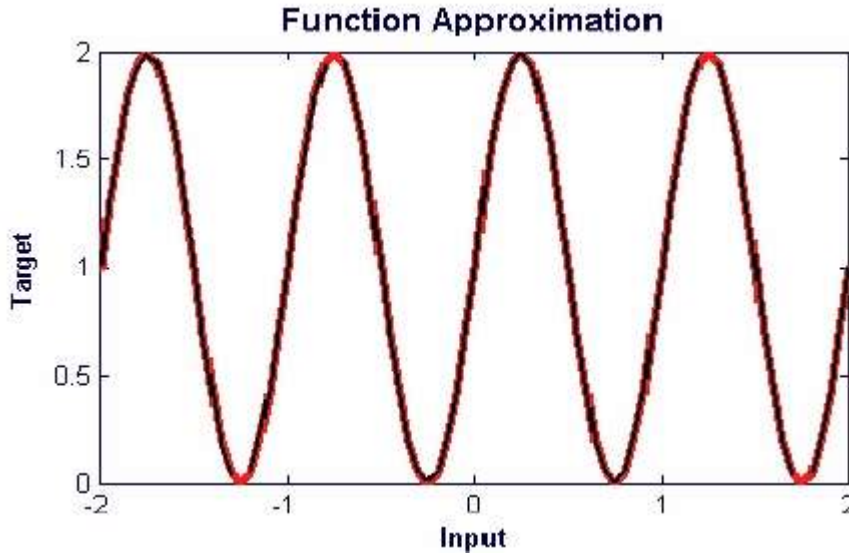
c, Trường hợp $i=4n$



d, Trường hợp $i=8o$

Xấp xỉ hàm $f(x) = 1 + \sin\left(\frac{i\pi}{4}x\right)$ với $-2 \leq x \leq 2$

Khi ta tăng số nơon trong lớp ẩn lên thì khả năng xấp xỉ hàm số của mạng sẽ tốt hơn. Chẳng hạn, xét trường hợp sử dụng 8 nơon trong lớp ẩn và $i=8$ ta có được kết quả sau:



Trường hợp $i=8$ và số nơron lớp ẩn là 8 p

Xấp xỉ hàm $f(x) = 1 + \sin\left(\frac{i\pi}{4}x\right)$ với $-2 \leq x \leq 2$ khi tăng số nơron

Như vậy nếu ta muốn xấp xỉ một hàm số mà có số điểm cần xấp xỉ là lớn thì ta sẽ cần số nơron lớn hơn trong lớp ẩn.

Sự hội tụ

Trong phần trên ta đã thấy các trường hợp mạng nơron không trả lại kết quả chính xác mặc dù thuật toán lan truyền ngược đã thực hiện tối thiểu hoá trung bình bình phương lỗi. Điều đó là do khả năng của mạng bị giới hạn bởi số nơron trong lớp ẩn. Nhưng cũng có trường hợp thuật toán lan truyền ngược không cho ta các tham số có thể dẫn đến kết quả chính xác nhưng mạng vẫn có thể xấp xỉ được hàm số. Điều này xảy ra là do trạng thái khởi đầu của mạng, sau khi huấn luyện, mạng có thể rơi vào điểm cực tiểu toàn cục hoặc rơi vào điểm cực tiểu địa phương.

Cần chú ý rằng trong thuật toán LMS, điểm cực trị toàn cục là luôn tồn tại bởi lẽ hàm trung bình bình phương lỗi của thuật toán LMS là hàm bậc hai, do là hàm bậc hai nên đạo hàm bậc hai của hàm lỗi sẽ là hằng số, do vậy mà độ cong của hàm theo một hướng cho trước là không thay đổi. Trong khi đó thuật toán lan truyền ngược áp dụng cho các mạng nhiều lớp sử dụng các hàm kích hoạt phi tuyến sẽ có nhiều điểm cực trị địa phương và độ cong của hàm lỗi có thể không cố định theo một hướng cho trước.

Sự tổng quát hoá

Trong phần lớn các trường hợp, mạng nơron truyền thẳng nhiều lớp được huấn luyện bởi một số cố định các mẫu xác định sự hoạt động đúng của mạng:

$$\{(p_1, t_1), (p_2, t_2), \dots, (p_q, t_q)\}$$

ở đây p_i là các đầu vào, t_i là đầu ra mong muốn tương ứng. Tập ví dụ huấn luyện này thông thường là thể hiện của số lớn nhất các lớp ví dụ huấn luyện có thể. Một điều rất quan trọng là mạng nơron có khả năng tổng quát hoá từ những cái nó đã học. Mặc dù dữ liệu có nhiều thì mạng vẫn có khả năng hoạt động tốt.

Để một mạng có khả năng tổng quát hoá tốt, nó cần có số tham số ít hơn số dữ liệu có trong tập ví dụ huấn luyện. Trong các mạng nơron, cũng như các bài toán mô hình hoá, ta thường mong muốn sử dụng một mạng đơn giản nhất mà có thể cho kết quả tốt trên tập ví dụ huấn luyện.

Một cách khác đó là dừng luyện mạng trước khi mạng xảy ra tình trạng quá khớp. Kỹ thuật này liên quan đến việc chia tập dữ liệu thu được thành ba tập : tập ví dụ huấn luyện để tính toán gradient và cập nhật các trọng số của mạng, tập kiểm định được dùng để kiểm tra điều kiện dừng của mạng và tập kiểm tra được sử dụng để so sánh khả năng tổng quát hoá của mạng đối với các bộ tham số của mạng sau các lần huấn luyện.

3.2.3 Một số biến thể của thuật toán lan truyền ngược

Ta đã biết một số đặc điểm của thuật toán lan truyền ngược sử dụng kỹ thuật giảm theo hướng. Mạng sử dụng thuật toán này tồn tại nhược điểm: rơi vào điểm cực tiểu địa phương đối với mạng nơron truyền thẳng nhiều lớp sử dụng các hàm kích hoạt phi tuyến. Hơn nữa, khi thực hiện luyện mạng bằng cách đưa từng ví dụ huấn luyện vào, sau đó thực hiện cập nhật tham số, sẽ làm ảnh hưởng đến quá trình học các ví dụ khác. Do đó, một phương pháp để tăng tốc độ hội tụ là sử dụng phương pháp học theo lô (batch training), nghĩa là tất cả các ví dụ huấn luyện được đưa vào mạng, sau đó mới thực hiện cập nhật các tham số. Bây giờ ta sẽ xem xét một số biến thể của thuật toán lan truyền ngược sử dụng phương pháp học theo lô nhằm vượt qua các nhược điểm này.

Sử dụng tham số bước đà

Đây là một phương pháp heuristic dựa trên quan sát kết quả luyện mạng nhằm làm tăng tốc độ hội tụ của thuật toán lan truyền ngược dựa trên kỹ thuật giảm nhanh nhất. Thuật toán lan truyền ngược cập nhật các tham số của mạng bằng cách cộng thêm vào một lượng thay đổi là:

$$\Delta W^m(k) = -\alpha s^m (a^{m-1})^T$$

$$\Delta b^m(k) = -\alpha s^m$$

Khi áp dụng thuật toán lan truyền ngược có sử dụng bước đà, phương trình trên thay đổi như sau:

$$\Delta W^m(k) = \mu \Delta W^m(k-1) - (1-\mu) \alpha s^m (a^{m-1})^T$$

$$\Delta b^m(k) = \mu \Delta b^m(k-1) - (1-\mu) \alpha s^m$$

Người ta đã chứng tỏ rằng khi sử dụng tham số bước đà thì hệ số học có thể lớn hơn rất nhiều so với thuật toán lan truyền ngược nguyên thủy không sử dụng tham số bước đà trong khi vẫn giữ được độ tin cậy của thuật toán. Một điểm nữa là khi sử dụng tham số bước đà thì sự hội tụ của thuật toán sẽ được tăng tốc nếu như thuật toán đang đi theo một hướng bền vững (chỉ đi xuống trong một khoảng dài).

Sử dụng hệ số học biến đổi

Trong thực tế, các hàm hiệu năng có dạng biểu diễn hình học là không đồng đều, có lúc có dạng phẳng (hàm không thay đổi giá trị hoặc thay đổi rất ít) hoặc có dạng phễu (giá trị của hàm thay đổi rất nhanh khi thay đổi tham số đầu vào). Nếu ta chỉ sử dụng hệ số học cố định thì có thể sẽ tốn thời gian tại các vùng phẳng. Vì vậy tư tưởng của thuật toán lan truyền ngược sử dụng hệ số học biến đổi là khi gặp vùng phẳng thì tăng hệ số học lên và ngược lại khi gặp vùng dạng phễu thì giảm hệ số học đi. Người ta đã đưa ra rất nhiều phương pháp để thực hiện điều trên, ở đây chỉ nêu ra một cách biến đổi hệ số học dựa trên hiệu năng của mạng.

- **Bước 1:** Nếu bình phương lỗi trên toàn bộ tập ví dụ huấn luyện tăng một số phần trăm cho trước γ (thông thường là từ 1% đến 5%) sau một lần cập nhật trọng số, thì bỏ qua việc cập nhật này, hệ số học được nhân với một số hạng ξ nào đó (với $0 < \xi < 1$) và tham số bước đà (nếu có sử dụng) được đặt bằng 0.

- **Bước 2:** Nếu bình phương lỗi giảm sau một lần cập nhật trọng số, thì cập nhật đó là chấp nhận được và hệ số học được nhân với một số hạng nào đó lớn hơn 1, nếu tham số bước đã được đặt bằng 0 thì đặt lại giá trị lúc đầu.

- **Bước 3:** Nếu bình phương lỗi tăng một lượng nhỏ hơn γ , thì cập nhật trọng số là chấp nhận được, nhưng hệ số học không thay đổi và nếu tham số bước đã đặt bằng 0 thì đặt lại giá trị lúc đầu.

Các thuật toán heuristic luôn cho ta sự hội tụ nhanh hơn trong một số bài toán, tuy nhiên chúng có hai nhược điểm chính sau đây:

Thứ nhất, việc sửa đổi thuật toán lan truyền ngược cần có thêm một số tham số, trong khi trong thuật toán lan truyền ngược nguyên thủy chỉ yêu cầu có một tham số đó là hệ số học. Một số thuật toán sửa đổi cần đến năm hoặc sáu tham số, trong khi hiệu năng của thuật toán khá nhạy cảm đối với những thay đổi của các tham số này. Hơn nữa việc chọn lựa các tham số lại độc lập với bài toán đặt ra.

Thứ hai, các thuật toán sửa đổi có thể không hội tụ trong một số bài toán mà thuật toán lan truyền ngược nguyên thủy có thể hội tụ được.

Người ta đã thấy rằng cả hai nhược điểm nêu trên thường xảy ra khi sử dụng các thuật toán sửa đổi phức tạp hơn (yêu cầu nhiều tham số hơn).

3.3 Chương trình dự báo đỉnh và đáy biểu đồ phụ tải

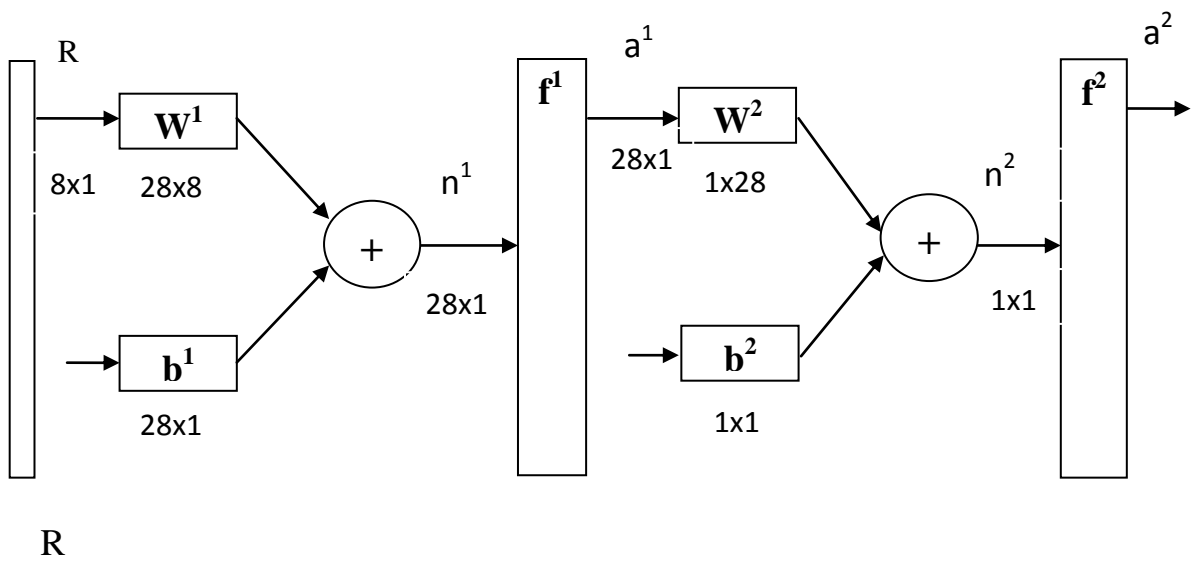
Mạng nơron lan truyền ngược sai số (Back - Propagation) sẽ được dùng để giải quyết bài toán thứ hai - dự báo đỉnh và đáy biểu đồ phụ tải của dự báo phụ tải.

3.3.1 Cấu trúc mạng nơron dự báo đỉnh biểu đồ phụ tải

Cấu trúc mạng gồm có 2 lớp: 1 lớp ẩn và một lớp ra, lớp ẩn gồm 28 nơron, lớp ra gồm 1 nơron, các thông số khác của mạng

Mạng gồm 8 đầu vào:

- Nhiệt độ cao nhất của ngày cần dự báo
- Nhiệt độ cao nhất của ngày trước ngày cần dự báo
- Nhiệt độ cao nhất của ba ngày có cùng kiểu ngày trước đó
- Phụ tải cao nhất của ba ngày có cùng kiểu ngày trước đó



Hình 3.2: Kiến trúc mạng q

3.3.2 Dữ liệu dự báo phụ tải

Các thông số phụ tải mẫu dùng để huấn luyện mạng nơron là P_{\max} , P_{\min} và nhiệt độ T_{\max} , T_{\min} tương ứng cùng ngày trong vài tuần trước. Tốc độ huấn luyện được chọn giá trị ban đầu trong khoảng 0.3 đến 0.5, hệ số quan tính được chọn giá trị ban đầu là 0.5

Ta sẽ chọn số liệu phụ tải P_{\max} và nhiệt độ T_{\max} trong ngày của các ngày làm việc tháng 6/2003 làm các ví dụ để học và dự báo. Bảng 5.1 là tập hợp các dữ liệu phụ tải cao điểm, phụ tải thấp điểm, nhiệt độ cao và nhiệt độ thấp trong ngày của tháng 6/2003.

DuLieuPhuTai	
🔑	Ngày
	PhuTai_Max
	NhietDo_Max
	PhuTai_Min
	NhietDo_Min

Hình 3.3: Biểu đồ cơ sở dữ liệu r

Dữ liệu dự báo đỉnh và đáy BDPT

Thứ	Ngày	DL dự báo đỉnh		DL dự báo đáy	
		Phụ tải cao điểm (MW)	Nhiệt độ cao nhất ($^{\circ}\text{C}$)	Phụ tải thấp điểm(MW)	Nhiệt độ thấp($^{\circ}\text{C}$)
CN	1/6/2003	5718.4	35	3334.2	26
2	2/6/2003	6258.2	36	3257.5	27
3	3/6/2003	6257.5	35	3532.3	27
4	4/6/2003	6189.6	37	3654.3	27
5	5/6/2003	6358.8	35	3460.5	26
6	6/6/2003	6364.4	36	3618.4	27
7	7/6/2003	6210.7	37	3905	28
CN	8/6/2003	6020.5	37	3967	27
2	9/6/2003	6588.5	36	3997.5	26
3	10/6/2003	6581.5	36	4316.5	26
4	11/6/2003	6454.9	35	4124.5	26
5	12/6/2003	6450.2	32	3837.5	25
6	13/6/2003	6587.1	33	3858.5	25
7	14/6/2003	6245.5	35	3658.5	26
CN	15/6/2003	5896.5	34	3842	26
2	16/6/2003	6412.9	35	3760.5	25
3	17/6/2003	6431.4	35	3890	25
4	18/6/2003	6506.5	35	3969	26
5	19/6/2003	6547.8	35	4149.5	27
6	20/6/2003	6592.2	33	4063.5	25
7	21/6/2003	6237.3	33	3850	26
CN	22/6/2003	5743.5	34	3618	26

2	23/6/2003	6396.7	33	3452	26
3	24/6/2003	6404.8	32	3829.5	25
4	25/6/2003	6391.6	35	3814	25
5	26/6/2003	6483.1	35	3956.5	26
6	27/6/2003	6284.8	35	3995.5	27
7	28/6/2003	5958.3	36	3781	27
CN	29/6/2003	5551.3	36	3634	26
2	30/6/2003	6050.8	35	3289	25

Có các số liệu về nhiệt độ cao nhất T_{\max} và phụ tải cao nhất P_{\max} tương ứng trong ngày của các ngày làm việc trong tháng, ta có thể xây dựng tập hợp ví dụ huấn luyện dùng để huấn luyện và kiểm tra khả năng dự báo của mạng. Một ví dụ huấn luyện cụ thể trong tập ví dụ huấn luyện nói trên được xây dựng từ tám giá trị vào:

- T_{\max} của ngày cần dự báo
- T_{\max} của ngày trước ngày cần dự báo
- T_{\max} của ba ngày có cùng kiểu ngày trước đó
- P_{\max} của ba ngày có cùng kiểu ngày trước đó

và một giá trị đầu ra : P_{\max} của ngày cần dự báo

Chẳng hạn, ngày cần dự báo là thứ 6 - 20/6/2003

Giá trị vào của mạng								Giá trị ra của mạng
T_{\max} (20/6)	T_{\max} (19/6)	T_{\max} (19/6)	T_{\max} (18/6)	T_{\max} (17/6)	P_{\max} (19/6)	P_{\max} (18/6)	P_{\max} (17/6)	P_{\max} (20/6)
33	35	35	35	35	6547.8	6506.5	6431.4	6592.2

Số liệu đầu vào và ra của các ngày khác được xây dựng theo phương pháp tương tự. Bảng 3.2 là tập ví dụ huấn luyện dùng cho quá trình học và dự báo của

mạng vào tháng 6/2003, được xây dựng trên cơ sở dữ liệu phụ tải của các ngày làm việc bình thường trong tháng.

Như vậy có dữ liệu về nhiệt độ cao nhất trong ngày và phụ tải cao nhất tương ứng của ngày đó ta có thể xây dựng tập ví dụ huấn luyện cho bất kỳ giai đoạn nào trong năm.

Các dữ liệu đầu vào và ra của mạng phải được chuyển đổi sao cho dữ liệu nằm trong khoảng $[0,1]$.

Đối với bài toán DBPT định ở Trung tâm Điều độ Quốc gia các giá trị phụ tải P_{\max} được chia cho 9000, nhiệt độ cao nhất trong ngày vào mùa hè được chia cho 60.

Tập dữ liệu

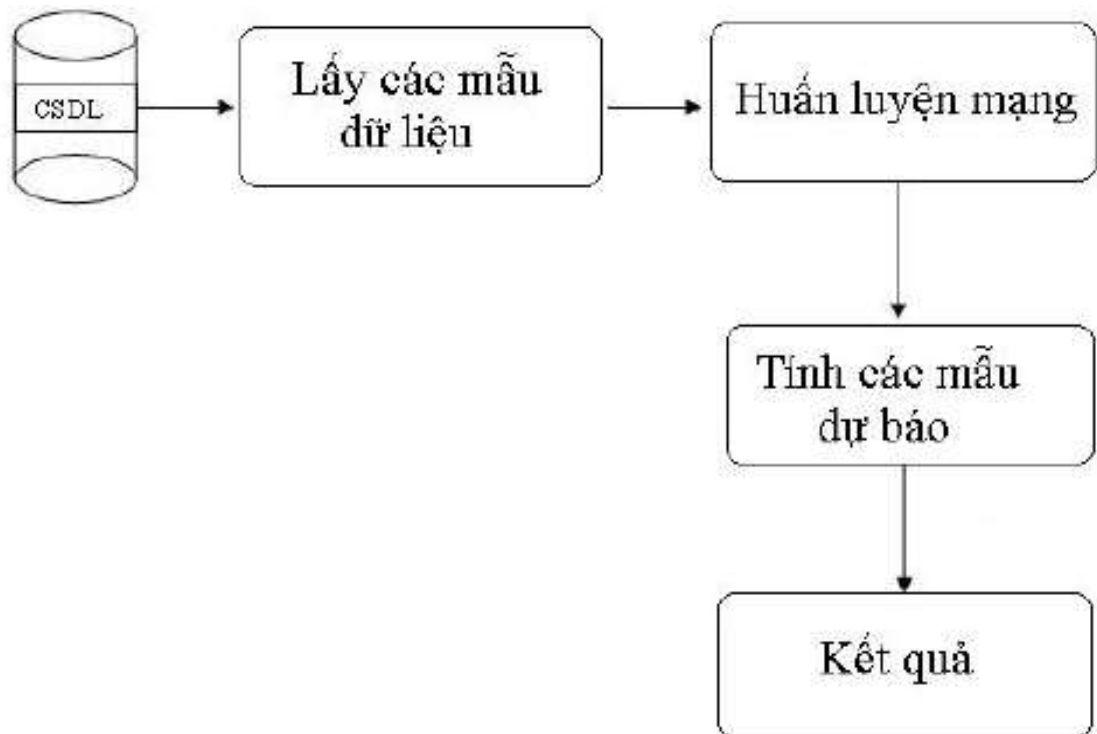
Giá trị vào của mạng									Giá trị ra của mạng
Ngày	Tmax ngày dự báo	Tmax ngày trước	Tmax ngày thứ 1 cùng kiểu ngày	Tmax ngày thứ 2 cùng kiểu ngày	Tmax ngày thứ 3 cùng kiểu ngày	Pmax ngày thứ 1 cùng kiểu ngày	Pmax ngày thứ 2 cùng kiểu ngày	Pmax ngày thứ 3 cùng kiểu ngày	Pmax ngày dự báo
9/6/2003	36	36	36	35	37	6364.4	6358.8	6189.6	6588.5
10/6/2003	36	36	36	36	35	6588.5	6364.4	6358.8	6581.5
11/6/2003	35	36	36	36	36	6581.5	6588.5	6364.4	6454.9
12/6/2003	32	35	35	36	36	6454.9	6581.5	6588.5	6450.2
13/6/2003	33	32	32	35	36	6450.2	6454.9	6581.5	6587.1
16/6/2003	35	34	34	33	32	6587.1	6450.2	6454.9	6412.9
17/6/2003	35	35	35	33	32	6412.9	6587.1	6450.2	6431.4
18/6/2003	35	35	35	35	33	6431.4	6412.9	6587.1	6506.5
19/6/2003	35	35	35	35	35	6506.5	6431.4	6412.9	6547.8
20/6/2003	33	33	33	35	35	6547.8	6506.5	6431.4	6592.2
23/6/2003	33	34	34	33	35	6592.2	6547.8	6506.5	6396.7
24/6/2003	32	33	33	33	35	6396.7	6592.2	6547.8	6404.8
25/6/2003	35	32	32	33	33	6404.8	6396.7	6592.2	6391.6
26/6/2003	35	35	35	32	33	6391.6	6404.8	6396.7	6483.1
27/6/2003	35	35	35	35	32	6483.1	6391.6	6404.8	6284.8
30/6/2003	35	36	36	35	35	6284.8	6483.1	6391.6	6050.8

Sau khi chuyển đổi ta có bảng giá trị sau:

Tập dữ liệu sau khi chuyển đổi

Giá trị vào của mạng									Giá trị ra của mạng
Ngày	Tmax ngày dự báo	Tmax ngày trước	Tmax ngày thứ 1 cùng kiểu ngày	Tmax ngày thứ 2 cùng kiểu ngày	Tmax ngày thứ 3 cùng kiểu ngày	Pmax ngày thứ 1 cùng kiểu ngày	Pmax ngày thứ 2 cùng kiểu ngày	Pmax ngày thứ 3 cùng kiểu ngày	Pmax ngày dự báo
9/6/2003	0.6	0.6	0.6	0.58	0.62	0.71	0.71	0.69	0.73
10/6/2003	0.6	0.6	0.6	0.6	0.58	0.73	0.71	0.71	0.73
11/6/2003	0.58	0.6	0.6	0.6	0.6	0.73	0.73	0.71	0.72
12/6/2003	0.53	0.58	0.58	0.6	0.6	0.72	0.73	0.73	0.72
13/6/2003	0.55	0.53	0.53	0.58	0.6	0.72	0.72	0.73	0.73
16/6/2003	0.58	0.57	0.57	0.55	0.53	0.73	0.72	0.72	0.71
17/6/2003	0.58	0.58	0.58	0.55	0.53	0.71	0.73	0.72	0.71
18/6/2003	0.58	0.58	0.58	0.58	0.55	0.72	0.71	0.73	0.72
19/6/2003	0.58	0.58	0.58	0.58	0.58	0.72	0.72	0.71	0.73
20/6/2003	0.55	0.55	0.55	0.58	0.58	0.73	0.72	0.72	0.73
23/6/2003	0.55	0.57	0.57	0.55	0.58	0.73	0.73	0.72	0.71
24/6/2003	0.53	0.55	0.55	0.55	0.58	0.71	0.73	0.73	0.71
25/6/2003	0.58	0.53	0.53	0.55	0.55	0.71	0.71	0.73	0.71
26/6/2003	0.58	0.58	0.58	0.53	0.55	0.71	0.71	0.71	0.72
27/6/2003	0.58	0.58	0.58	0.58	0.53	0.72	0.71	0.71	0.7
30/6/2003	0.58	0.6	0.6	0.58	0.58	0.7	0.72	0.71	0.67

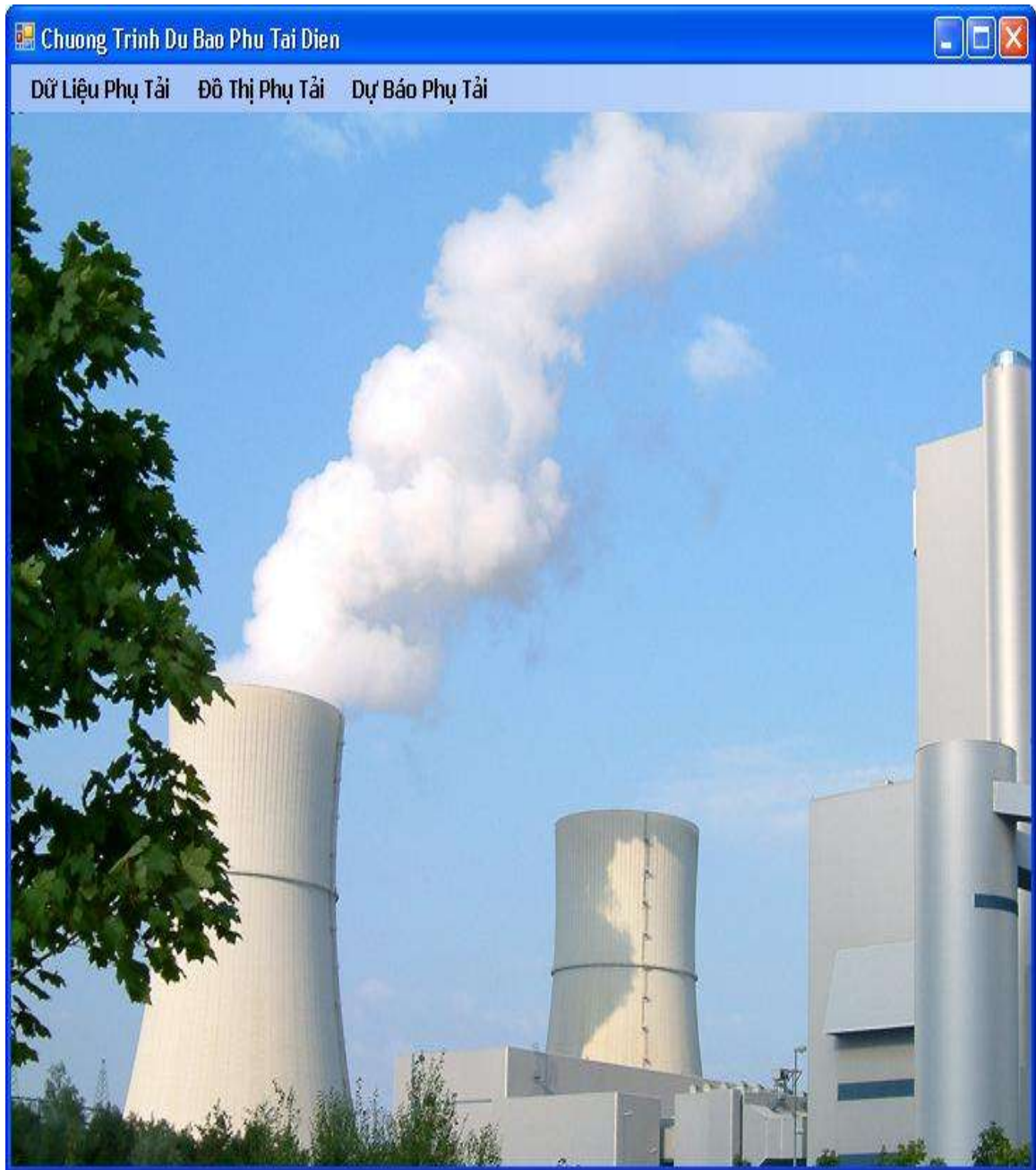
3.3.3 Sơ đồ khối của chương trình



Hình 3.4 : Sơ đồ khối của chương trình s

3.3.4 Hướng dẫn sử dụng chương trình

Giao diện chính



Hình 3.5 Giao diện chính t

Giao diện form dữ liệu

Dữ Liệu Phụ Tải

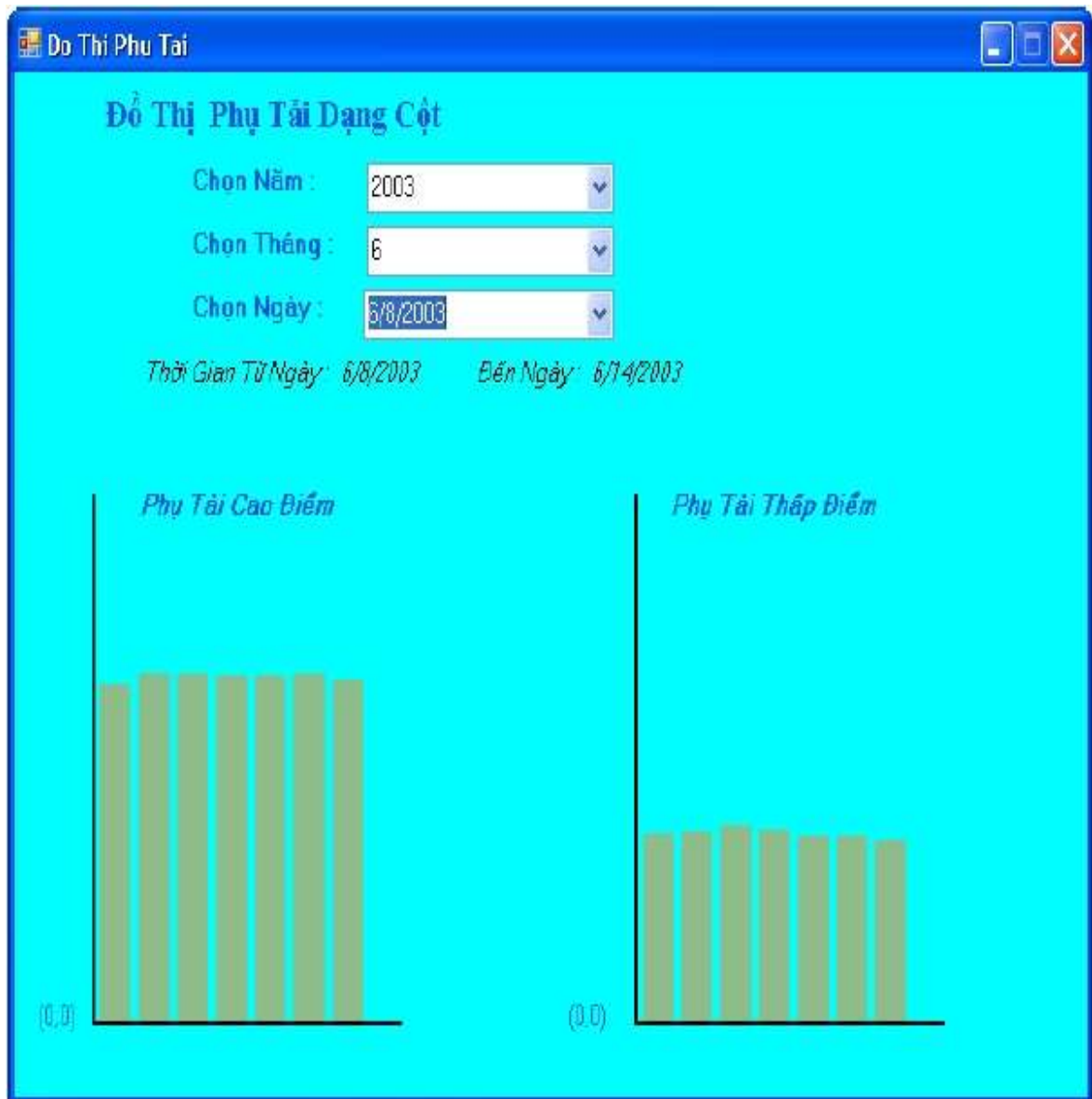
Ngày:
 Phụ Tải Max:
 Nhiệt Độ Max:
 Phụ Tải Min:
 Nhiệt Độ Min:

Ngày	PhuTai_Max	NhietDo_Max	NhietDo_Min	PhuTai_Min
3/2/2003	5106	25	15	3120
6/1/2003	5718.4	34	26	3334.2
6/2/2003	6258.2	33	25	3257.5
6/3/2003	6257.5	34	26	3532.3
6/4/2003	6189.6	34	26	3654.3
6/5/2003	6358.8	34	26	3460.5
6/6/2003	6364.4	35	27	3618.4
6/7/2003	6210.7	34	26	3905
6/8/2003	6020.5	34	26	3967
6/9/2003	6588.5	32	24	3997.5
6/10/2003	6581.5	33	25	4316.5
6/11/2003	6454.9	33	25	4124.5
6/12/2003	6450.2	32	25	3837.5
6/13/2003	6587.1	33	26	3858.5
6/14/2003	6245.5	35	27	3658.5

Hình 3.6 Giao diện form dữ liệu u

Gồm các chức năng : Thêm, Sửa, Xóa dữ liệu

Giao diện form đồ thị phụ tải



Hình 3.7 Giao diện form đồ thị phụ tải v

Chức năng mô phỏng biểu đồ dữ liệu dạng cột của 7 ngày trong tuần

Giao diện form dự báo phụ tải

The screenshot shows a window titled "Du Bao Phu Tai" with a light blue background. At the top, the title bar contains the window name and standard Windows window controls. Below the title bar, the text "Chọn Dữ Liệu Phụ Tải" is displayed in a bold, dark blue font. The form consists of several input fields: "Chọn Năm" with a dropdown menu showing "2003"; "Chọn Tháng" with a dropdown menu showing "6"; "Chọn Ngày Dự Báo" with a dropdown menu showing "6/23/2003"; "Số Vòng Huấn Luyện" with a text input field containing "1000"; and "Hệ Số Lỗi" with a text input field containing "0.0001". Below these fields are two buttons: "Dự Báo Phụ Tải Cao Điểm" and "Dự Báo Phụ Tải Thấp Điểm". At the bottom of the window is a large, empty white rectangular area.

Hình 3.8 Giao diện form dự báo phụ tải w

Chức năng :

Lấy các mẫu dữ liệu để huấn luyện

Chọn ngày cần dự báo

Chọn số vòng huấn luyện và hệ số lỗi khi huấn luyện

Dự báo các đỉnh và đáy của phụ tải

Giao diện form huấn luyện phụ tải

P Dự Báo	P Mong Muốn	Sai Số	% Sai Số
6344.2425599	6431.4	87.1574401	1.3551861
6342.6043948	6412.9	70.2956052	1.0961594
6342.7319267	6587.1	244.3680733	3.7097975
6342.6315969	6450.2	107.3684031	1.6645748
6344.6808628	6454.9	110.2191372	1.7075266
6345.7473187	6581.5	234.7526913	3.5669568
6346.7809577	6588.5	241.7190423	3.6688023
6347.1893347	6364.4	17.2106653	0.2704209
6347.4355231	6358.8	11.3644769	0.1787205
6349.0380391	6189.6	-159.4380391	-2.5759021
6346.9579753	6257.5	-89.4579753	-1.4296121
6347.6183101	6258.2	-89.4183101	-1.4288184

Hình 3.9 Giao diện form huấn luyện phụ tải x

Đưa ra các kết quả mẫu huấn luyện, điều chỉnh trọng số của tập huấn luyện sao cho mẫu huấn luyện có sai số và % sai số là nhỏ nhất.

Giao diện form dự báo phụ tải

P Dự Báo	P Mong Muốn	Sai Số	% Sai Số
6439.0656599	6391.6	-47.4656599	-0.7426256

Hình 3.9 Giao diện form dự báo phụ tải y

Đưa ra kết quả của mẫu dự báo, sai số và % sai số

KẾT LUẬN

Dự báo phụ tải ngắn hạn sử dụng mạng nơron là phương pháp hiện nay đang được áp dụng mạnh trong rất nhiều lĩnh vực, đặc biệt là giải quyết các bài toán mà quan hệ đầu vào và đầu ra là phi tuyến với nhiều đầu vào và nhiều đầu ra, được các chuyên gia đánh giá là có độ chính xác cao so với các phương pháp trước đó.

Trong luận văn, mạng nơron nhân tạo nhiều lớp lan truyền ngược được nghiên cứu, ứng dụng để giải quyết bài toán dự báo phụ tải ngắn hạn.

Giai đoạn 1 là giai đoạn nhận dạng biểu đồ. Với dữ liệu của quá khứ lưu trữ được ta dễ dàng xác định được kiểu ngày của ngày cần dự báo.

Giai đoạn 2, mạng nơron truyền ngược sai số qua quá trình huấn luyện với tập mẫu là các giá trị phụ tải đỉnh (phụ tải đáy) và nhiệt độ cao nhất (nhiệt độ thấp nhất) của các ngày có cùng kiểu ngày với ngày dự báo, được sử dụng để dự báo phụ tải đỉnh và phụ tải đáy theo nhiệt độ môi trường.

Qua việc nghiên cứu và áp dụng mạng nơron vào dự báo phụ tải ngắn hạn có thể thấy việc dự báo đỉnh và đáy biểu đồ phụ tải đem lại độ chính xác hơn hẳn phương pháp hồi qui tuyến tính. Sai số dự báo đỉnh và đáy biểu đồ rất nhỏ vào khoảng từ 0-5%. Có véc tơ BĐPT, phụ tải đỉnh và phụ tải đáy của ngày cần dự báo. Một yếu tố quan trọng là tập ví dụ huấn luyện cần phải thu thập sao cho đạt độ chính xác cao. Nếu có các ví dụ huấn luyện vào những ngày có sự cố bất thường xảy ra, mạng nơron sẽ cho kết quả với sai số cao. Việc loại trừ sai số chủ quan do lấy thông tin phụ tải xảy ra trong ngày có sự cố, hoặc loại trừ việc ghi số liệu không chính xác sẽ làm tăng độ chính xác của quá trình dự báo.

Để đảm bảo độ chính xác của quá trình dự báo, cần thường xuyên kiểm tra kết quả dự báo để kịp thời huấn luyện lại mạng theo các số liệu mới, nếu thấy sai số dự báo tăng. Việc sử dụng mạng nơron vào quá trình dự báo phụ tải ngắn hạn theo phương pháp hai giai đoạn như trên cho phép dự báo phụ tải từng ngày khá chính xác, đơn giản và nhanh chóng. Hơn nữa phương pháp dự báo này còn giúp các kỹ sư mới vào nghề, qua khoảng thời gian nghiên cứu không lâu, có thể nhanh chóng nắm bắt và làm việc độc lập trong quá trình dự báo.

Ngoài những ưu điểm nêu trên, phương pháp dự báo này cũng có một số hạn chế như : quá trình huấn luyện mất nhiều thời gian và không đảm bảo chắc chắn là hội tụ. Mặc dù đã có những điều chỉnh để nhằm rút ngắn thời gian huấn luyện mạng và để có được kết quả tốt nhất, nhưng đây vẫn là vấn đề cần tiếp tục nghiên cứu. Việc xác định cấu trúc mạng, lựa chọn hệ số học chỉ có thể thông qua thực nghiệm. Để có một nghiên cứu toàn diện, mang lại hiệu quả thực sự cho quá trình dự báo phụ tải trong ngành Điện lực, trong tương lai cần phải đi sâu nghiên cứu hơn nữa.

TÀI LIỆU THAM KHẢO

Tài liệu tham khảo tiếng Việt

- [1.] Đinh Mạnh Tường(2002), *Trí tuệ nhân tạo*, NXB Khoa học và Kỹ thuật
- [2.] Nguyễn Hữu Tình, Lê Tấn Hùng, Phạm Ngọc Yến, Nguyễn Lan Hương(1999), *Cơ sở và ứng dụng Matlab*, NXB Khoa học và Kỹ thuật.
- [3.] Trần Đình Long VS GS TSKH(1998), *Quy hoạch phát triển Năng lượng và Điện lực*, NXB ĐHBK Hà Nội.

Tài liệu tham khảo tiếng Anh

- [4.] Dipti Srinivasan, A. C. Liew, John S., P. Chen(1991), “*Short term forecasting using neural network aproach*”, IEEE 91TH0374-9/91/0000-0012, pp. 12-16.
- [5.] Hagan M. T., Demuth H. B., Beale M(1996), *Neural Networks Design*, PWS Publishing Company, Boston, Ma.
- [6.] Kaastra, I., & Boyd, M(1996), – *Designing a neural network for forecasting financial and economic time series* - Neurocomputing 10, pp. 215-236.
- [7.] Oh S.H., Lee Yj(1995), *A modified error function to improve the error Back-Propagation algorithm for Multi-layer perceptrons*, ETRI Journal Vol 17, No 1.
- [8.] Ooyen A. V., Nienhuis B(1992), *Improving the Convergence of the Back-Propagation algorithm*, Neural networks, Vol. 5, pp.465-471.
- [9.] T. masters(1993), *Practical neural network Recipes in C++*, Academic Press, Inc.
- [10.] Tom M. Mitchell(1997), *Machine learning*, The McGraw-Hill Companies, Inc.
- [11.] <http://www.codeproject.com/>
- [12.] Artificial neural network From Wikipedia, the free encyclopedia.