

SỬ DỤNG MẠNG NƠON ĐA LỚP TRUYỀN THẮNG VÀ MẠNG TRUY HỒI DỰ BÁO TỔNG LƯỢNG BỨC XẠ NGÀY CHO MỘT SỐ TRẠM Ở ĐỒNG BẰNG PHÍA BẮC VIỆT NAM

CN. Hoàng Phúc Lâm, PGS. TS. Nguyễn Hương Điền
 CN. Công Thanh, ThS. Hoàng Thanh Vân
 Trường Đại học Khoa học Tự nhiên, ĐHQGHN

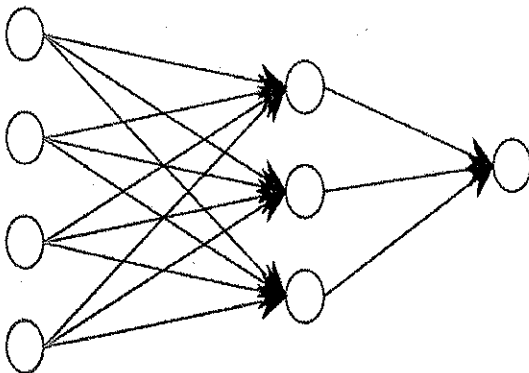
Bài báo trình bày việc sử dụng mạng nơon đa lớp truyền thẳng và mạng truy hồi để dự báo tổng lượng bức xạ ngày cho 4 trạm đặc trưng cho khu vực đồng bằng phía bắc Việt Nam, đó là: Láng (Hà Nội), Phủ Liễn (Hải Phòng), Yên Định (Thanh Hoá) và Vinh (Nghệ An). Bộ số liệu quan trắc của 4 trạm kể trên trong khoảng thời gian từ tháng 1 năm 1981 đến tháng 12 năm 1990 được chia làm 3 phần dùng để xây dựng, kiểm chứng chéo và đánh giá chất lượng của mạng.

Chất lượng của mạng nơon được so sánh với bộ số liệu tái phân tích của NCEP đã được nội suy về trạm. Tất cả các mạng nơon đều cho chất lượng dự báo với độ chính xác tăng. Các sai số RMSE, MSE và ME được cải thiện ít nhất 25%, còn hệ số tương quan r tăng tới 45% so với số liệu tái phân tích. Điều đó cho thấy mạng nơon đã nắm bắt được khá tốt các đặc trưng của bộ số liệu. Mạng nơon sẽ là một công cụ hữu hiệu dùng cho các bài toán thống kê đầu ra mô hình.

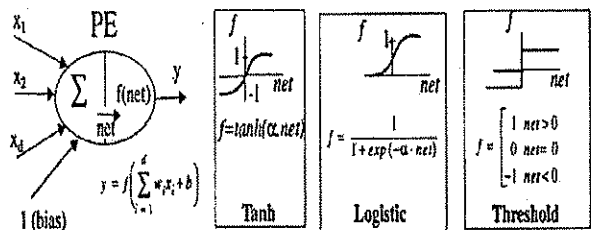
1. Lý thuyết về mạng nhận thức đa lớp và mạng truy hồi

a. Mạng nhận thức đa lớp

Mạng nhận thức đa lớp là một sự sắp xếp theo lớp của các phần tử hoạt động như trong hình 1.



Hình 1. Mạng nhận thức đa lớp với một lớp ẩn



Hình 2. Phân tử hoạt động và các toán tử phi tuyến thông dụng

Lớp nhận các giá trị đầu vào được gọi là lớp nhập, lớp tạo ra các giá trị đầu ra gọi là lớp xuất. Các lớp không có liên hệ trực tiếp với tập số liệu nhập và xuất gọi là các lớp ẩn. Một mạng chỉ bao gồm lớp nhập và lớp xuất được gọi là mạng nhận thức. Mỗi kết nối giữa các phần tử hoạt động được định lượng bởi giá trị w_i , gọi là trọng số, và được điều chỉnh trong suốt quá trình luyện mạng.

Các phần tử hoạt động trong mạng nhận thức đa lớp bao gồm một toán tử tổng và một

toán tử phi tuyến dạng S (hình 2). Các toán tử phi tuyến dạng S thông dụng nhất bao gồm: hàm logic, hàm tang hypebol và đôi khi là hàm giới hạn.

Sức mạnh của mạng nhận thức đa lớp thể hiện ở chỗ, nó là một ánh xạ gần đúng tổng quát (ánh xạ phổ dụng - thực hiện một ánh xạ bất kỳ giữa lớp nhập và lớp xuất) khi nó được cấu tạo với ít nhất hai lớp ẩn và số phần tử hoạt động thích hợp trong từng lớp ẩn đó [Haykin, 1994]. Thậm chí, mạng nhận thức đa lớp với một lớp ẩn cũng có thể xấp xỉ các ánh xạ liên tục giữa lớp nhập và lớp xuất. Điều này có nghĩa là chúng ta hiếm khi phải chọn một mạng nhận thức đa lớp với nhiều hơn hai lớp ẩn. Điều này thực sự có ý nghĩa trong thực tế, do nhu cầu về thời gian tính toán, khối lượng tính toán... nên một mạng nhỏ, đơn giản... sẽ là rất hữu ích. Tuy nhiên, khó khăn đối với mạng nhận thức đa lớp lại là việc chọn số lớp (0, 1 hoặc 2), số phần tử hoạt động trong mỗi lớp nếu muốn đạt được một kết quả tốt nhất.

Rất nhiều bài toán kỹ thuật có thể được hình dung theo cách chuyển đổi thông tin từ không gian nhập sang thông tin trong không gian xuất. Ví dụ, chia số liệu thành các lớp, các pha khác nhau; tìm trạng thái của lớp xuất khi biết trạng thái của các phần tử nhập, [Bishop, 1995]; tương tự, việc xác định một hệ thống chưa biết cũng có thể coi là một xấp xỉ hàm (ánh xạ) từ lớp nhập đến lớp xuất [Kung, 1993]. Mạng nhận thức đa lớp rất thích hợp cho các ứng dụng dạng này.

Sai số bình phương trung bình (MSE) được sử dụng làm căn cứ điều chỉnh các trọng số của mạng:

$$E = \frac{1}{P} \sum_{p=1}^P \frac{1}{2} \sum_{n=1}^N (d_n - y_n)^2 \quad (1)$$

Trong đó: P là số lần luyện (kiểm tra), N là số mẫu luyện (kiểm tra), d_n là kết xuất của mạng và y_n là kết xuất mong muốn (kết xuất

đích). Sai số trung bình là một hàm của các trọng số (do kết xuất d của mạng chính là hàm của các trọng số) và đạo hàm của sai số trung bình theo toàn bộ các trọng số của mạng sẽ cho ta biết hướng và độ dốc của mặt phẳng sai số E, từ đó các trọng số sẽ được điều chỉnh theo "hướng dốc nhất".

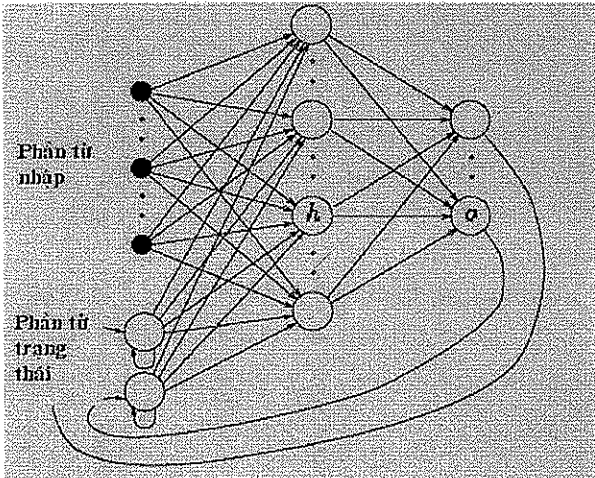
b. Mạng truy hồi

Ý tưởng cho sự ra đời của mạng truy hồi xuất phát từ mạng lan truyền tiến, với câu hỏi được đặt ra: Mạng nhận thức đa lớp là mạng phổ biến nhất nhưng lại chỉ là mạng lan truyền tiến, chỉ có khả năng xử lý các thông tin tức thời. Chuyện gì sẽ xảy ra nếu ta thêm vào một vòng lặp, ví dụ ta liên kết một nút ẩn với chính nó thông qua trọng số liên kết, hoặc liên kết các nút ẩn, và thậm chí là liên kết tất cả các phần tử của mạng với nhau? Mặc dù, như chúng ta đã biết, khả năng tính toán của mạng như vậy sẽ không tăng, chúng ta có thể giảm độ phức tạp, kích thước của mạng... khi giải cùng một bài toán (vì khi này, thông tin giữa các thành phần của mạng được trao đổi qua lại, tăng khả năng tổng quát của mạng).

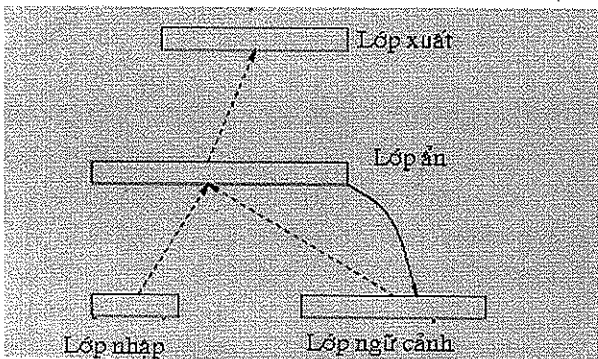
Vấn đề cần được xem xét liệu mạng truy hồi có thể tiếp tục truyền các giá trị tác động đến vô cùng, hay ít nhất là đến khi đạt được một điểm ổn định? Câu trả lời là có. Với mạng truy hồi, ta có thể đưa ra một ngưỡng để mạng ngừng quá trình lan truyền và điều chỉnh trọng số. Hai ví dụ cổ điển về mạng truy hồi là mạng Jordan và mạng Elman. Trong mạng Jordan (Jordan, 1986), các giá trị tác động của lớp xuất được đưa trở lại lớp nhập thông qua một tập hợp các nơron phụ (hình 3) được gọi là nút trạng thái (state units). Các liên kết giữa lớp xuất và các nơron phụ có trọng số không đổi là +1; việc luyện chỉ có ở các liên kết giữa lớp nhập và lớp ẩn và giữa các lớp ẩn và lớp xuất. Do đó, tất cả các quy tắc học của mạng nhận thức đa lớp đều có thể sử dụng với loại mạng này.

Mạng Elman (Elman, 1990), một tập hợp

nơon ngữ cảnh thay thế cho các nơon phụ trong mạng Jordan, và nó đưa thông tin từ lớp ẩn trở lại lớp nhập (hình 4). Do đó, về nguyên tắc cũng giống như mạng Jordan, trừ việc nó đưa thông tin trở lại từ nút ẩn chứ không phải từ nút xuất và trong tập hợp nơon ngữ cảnh không liên kết lẫn nhau.



Hình 3. Mạng Jordan. Giá trị tác động của lớp xuất được đưa trở lại lớp nhập tới một tập hợp các nơon phụ được gọi là nút trạng thái (state units)



Hình 4. Mạng Elman, thông tin từ lớp ẩn được đưa trở lại lớp nhập thông qua tập hợp nơon ngữ cảnh

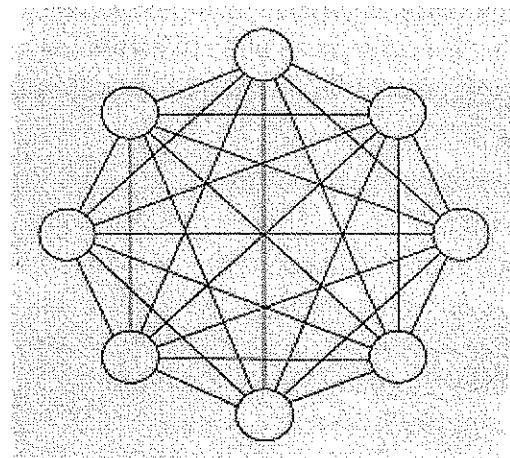
Mạng Hopfield là một trong những loại mạng truy hồi xuất hiện sớm nhất và nổi tiếng nhất. Năm 1982, Hopfield đưa ra một lý thuyết toán học hoàn chỉnh về mạng truy hồi; sau đó được phát triển thành mô hình "Ising spin" (Amit, Gutfreund và Sompolinsky, 1986)

Mạng Hopfield bao gồm tập hợp N nơon liên kết lẫn nhau (hình 5) và chúng cập nhật giá trị tác động một cách độc lập. Mọi nơon đều vừa là nơon nhập và xuất. Các giá trị tác động dưới dạng mã nhị phân (1 và 0), điều này khiến mạng Hopfield tương tự như một mạch điện gồm nhiều công tắc được điều khiển bằng các ký tự nhị phân 0 (tắt) và 1 (bật). Trạng thái của hệ thống được cho bởi các giá trị tác động $y = y_k$. Tập hợp giá trị nhập $s_k(t+1)$ của nơon thứ k tại vòng t+1 được lấy tổng có trọng số:

$$s_k(t+1) = \sum y_j(t)w_{jk} + \theta_k \quad (2)$$

Một hàm giới hạn đơn giản được sử dụng cho tập hợp nhập để thu được giá trị tác động mới $y_i(t+1)$ ở thời điểm t+1:

$$y_k(t+1) = \begin{cases} +1 & \text{Nếu } s_k(t+1) > U_k \\ -1 & \text{Nếu } s_k(t+1) < U_k \\ y_k(t) & \text{Trường hợp còn lại} \end{cases} \quad (3)$$



Hình 5. Mạng tự liên kết. Tất cả các nơon đồng thời là nơon nhập và xuất.

2. Ứng dụng dự báo tổng xạ ngày sử dụng mạng nhận thức đa lớp và mạng truy hồi

- Số liệu đầu vào bao gồm số liệu quan trắc bức xạ ngày tại 4 trạm tiêu biểu cho khu vực đồng bằng phía bắc bao gồm: Láng (Hà Nội), Phủ Liễn (Hải Phòng), Yên Định (Thanh Hoá), Vinh (Nghệ An) trong vòng 10 năm (từ năm 1981 đến 1990).

- Số liệu tái phân tích của NCEP tại nút lưới gần nhất với từng trạm của các yếu tố: Tổng

xạ dự báo tại bề mặt (Qsf), tổng xạ tại đỉnh khí quyển (Qtoa); lượng mây (CL), nhiệt độ bề mặt (Tsf), độ ẩm tương đối bề mặt (sHsf), hai thành phần gió tại mực 10m (U10m, V10m), tổng lượng nước khả giáng trong khí quyển (PW) trung bình ngày, được lấy cùng thời điểm với giá trị quan trắc và được xem là giá trị dự báo cho trạm đó.

- Riêng trạm Yên Định (Thanh Hoá) do số liệu những tháng đầu năm 1981 không có liên tục nên chuỗi thực tế sử dụng là từ 1/9/1981 đến 31/12/1990

- Chuỗi số liệu được lấy trung bình trượt với bước trượt là 3 (ngày) để lọc bỏ sai số ngẫu nhiên.

- 70% chuỗi số liệu được sử dụng để xây dựng mạng, 20% cho kiểm chứng chéo và 10% được dùng làm số liệu kiểm tra độc lập chất lượng của mạng.

Cấu hình mạng thần kinh nhân tạo (TKNT) được chọn dựa vào thuật toán Gen (Genetic Optimization). Thuật toán được hiểu đơn giản là một mạng được luyện nhiều lần với với các bộ tham số khác nhau của mạng (số nút ẩn, bước học...); với mỗi bộ tham số như vậy, mạng được luyện cho đến khi xuất hiện dấu hiệu quá khớp hoặc đạt được một sai số mong muốn. Cụ thể với bài toán dự báo tổng xạ, các tham số bao gồm: số nút ẩn của mỗi lớp ẩn được thay đổi từ 1 đến 25, tốc độ học dao động trong khoảng từ 0.1 đến 1; và với mỗi bộ tham số, mạng sẽ được luyện tối đa là 10 000 bước hoặc khi sai số bình phương trung bình trên bộ số liệu kiểm chứng chéo tăng)

Tập hợp các tham số cho sai số trên bộ số liệu kiểm chứng chéo nhỏ nhất sẽ được lưu lại và cố định trong quá trình kiểm tra với số liệu độc lập. Cấu hình của mạng được cho trong các bảng dưới đây:

Bảng 1. Số nút ẩn của mạng TKNT nhận thức đa lớp

	Láng	Phủ Liễn	Yên Định	Vinh
Lớp nhập	8	8	8	8
Lớp ẩn 1	8	5	10	8
Lớp ẩn 2	5	5	5	5
Lớp xuất	1	1	1	1

Bảng 2. Số nút ẩn của mạng truy hồi

Đơn vị	Láng	Phủ Liễn	Yên Định	Vinh
Lớp nhập	8	8	8	8
Lớp ẩn 1	20	20	23	22
Lớp ẩn 2	19	17	15	24
Lớp xuất	1	1	1	1

3. Đánh giá kết quả, nhận xét và kết luận

a. Các chỉ số đánh giá

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (f_i - O_i)^2} \tag{4}$$

- Sai số toàn phương trung bình.

$$MAE = \frac{1}{N} \sum_{i=1}^N |f_i - O_i| \tag{5}$$

- Sai số tuyệt đối trung bình:

$$ME = \frac{1}{N} \sum_{i=1}^N (f_i - O_i) \tag{6}$$

- Sai số trung bình:

$$r = \frac{\sum(f - \bar{f})(O - \bar{O})}{\sqrt{\sum(f - \bar{f})^2} \sqrt{\sum(O - \bar{O})^2}} \quad (7)$$

- Hệ số tương quan.

Trong đó, f là giá trị dự báo, O là giá trị quan trắc, N là tổng số mẫu đánh giá. Bên cạnh đó, sai số tương đối (ký hiệu là $E\%$) được sử dụng để so sánh giữa các trạm với nhau. Sai số tương đối được định nghĩa:

$$E\% = \frac{E}{\frac{1}{N} \sum_{i=1}^N O_i} \quad (8)$$

Và để kiểm tra hiệu quả của mạng, đánh giá sự cải thiện của dự báo giữa sử dụng mạng và không sử dụng mạng, chỉ số kỹ năng được sử dụng. Để so sánh dự báo A với dự báo B

dựa trên một chỉ số đánh giá nào đó (ME, RMSE, r , ... - ký hiệu chung là S), ta sử dụng chỉ số kỹ năng:

$$Skill(A, B) = \frac{S_A - S_B}{S_p - S_B} \quad (9)$$

Trong đó S_p là chỉ số đánh giá lý tưởng, ví dụ với ME, RMSE là 0, với r là 1. Chỉ số Skill (A,B) càng dương thì mô hình A càng tốt hơn mô hình B và ngược lại. Trong bài báo này, các mô hình mạng được đánh giá kỹ năng so với tập số liệu tái phân tích của NCEP đã được nội suy về trạm.

b. Kết quả đánh giá

Kết quả đánh giá được cho cụ thể như sau:

Bảng 3. Đánh giá chất lượng dự báo của mô hình NCEP (Đơn vị: ME, MAE, RMSE: MJ/m²/ngày; hệ số tương quan r không có đơn vị)

Đơn vị	Láng	Phủ Liễn	Yên Định	Vinh
ME (ME%)	-0,25 (-2,70%)	2,45 (37,08%)	0,31 (4,23%)	-0,59 (-5,84%)
MAE(MAE%)	3,94 (42,98%)	3,52 (53,11%)	3,94 (45,48%)	5,13 (50,84%)
RMSE(RMSE%)	4,91 (53,57%)	4,19 (63,37%)	4,84 (55,78%)	6,53 (64,76%)
r	0,29	0,37	0,31	0,39

Qua bảng 3 ta thấy rõ ràng rằng số liệu tổng xạ bề mặt của mô hình toàn cầu NCEP với bước lưới 2,5x2,5 được nội suy về trạm cho chất lượng số liệu không tốt, thường dự báo hụt ở hai trạm Láng và Vinh và dự báo vượt ở hai trạm Phủ Liễn và Yên Định. Sai số tuyệt đối của cả 4 trạm đều trên 3,5 MJ/m²/ngày (tương ứng 40 - 50% giá trị trung bình). Sai số quân phương đặc biệt lớn, đều trên 50% giá trị trung bình khí hậu, trong đó hai trạm sai số

trên 63% là Phủ Liễn (63,37%) và Vinh (64,76%); hệ số tương quan rất nhỏ, không quá 0,4. Qua đây có thể nhận xét, không thể sử dụng trực tiếp số liệu của mô hình khí hậu NCEP trong thực tế nghiên cứu và ứng dụng. Do đó, các mô hình dự báo cho từng trạm riêng biệt, với các loại mạng TKNT khác nhau được sử dụng làm công cụ thống kê sau mô hình.

Bảng 4. Đánh giá chất lượng mạng nhận thức đa lớp

Đơn vị	Láng	Phủ Liễn	Yên Định	Vinh
ME (ME%)	1,18 (12,89%)	1,03 (15,40%)	0,88 (10,16%)	-1,82 (-18,01%)
MAE(MAE%)	2,63 (28,64%)	2,33 (35,0%)	2,47 (28,47%)	2,97 (29,44%)
RMSE(RMSE%)	3,29 (35,86%)	3,28 (49,17%)	3,17 (36,54%)	4,10 (40,61%)
r	0,80	0,66	0,80	0,88

Bảng 5. Chỉ số kỹ năng của mạng nhận thức đa lớp so với dự báo của mô hình NCEP

Đơn vị	Láng	Phủ Liễn	Yên Định	Vinh
RMSE	0,33	0,22	0,34	0,38
R	0,72	0,45	0,71	0,77

Bảng 6. Đánh giá chất lượng mạng truy hồi

Đơn vị	Láng	Phủ Liễn	Yên Định	Vinh
ME (ME%)	1,13 (12,28%)	1,20 (17,97%)	-0,05 (-0,63%)	-0,80 (-7,5%)
MAE(MAE%)	2,59 (28,22%)	2,44 (36,56%)	2,27 (26,21%)	2,90 (27,9%)
RMSE(RMSE%)	3,27 (35,69%)	3,16 (47,34%)	2,91 (33,65%)	4,04 (39,4%)
r	0,80	0,68	0,82	0,86

Bảng 7. Chỉ số kỹ năng của mạng truy hồi so với dự báo của mô hình NCEP

Đơn vị	Láng	Phủ Liễn	Yên Định	Vinh
RMSE	0,33	0,25	0,40	0,37
r	0,72	0,49	0,74	0,80

Nhìn chung, chất lượng dự báo của mạng nhận thức đa lớp là thấp hơn so với mạng truy hồi. Điều này là do mức độ phức tạp của mạng truy hồi cao hơn so với mạng nhận thức đa lớp và phù hợp hơn với bài toán phức tạp như dự báo tổng xạ. Tuy nhiên, sự khác biệt này là không đáng kể (khoảng vài phần trăm đơn vị). Nếu xét về mức độ kinh tế (thời gian luyện mạng) thì mạng nhận thức đa lớp hơn hẳn so với mạng truy hồi khi tất cả các tham số khác là như nhau.

Những kết quả đạt được với mạng truy hồi và mạng nhận thức đa lớp đã giúp cải thiện được đáng kể chất lượng của bộ số liệu tái phân tích.

Sử dụng mạng TKNT đã cải thiện được các loại sai số từ 25 đến 40%, riêng hệ số tương quan tăng từ 45 đến 80%. Rõ ràng mạng đã nắm bắt được rất tốt xu thế của chuỗi số liệu và dự báo khá chính xác trường tổng xạ, một trong những trường khí tượng có sai số lớn nhất trong các mô hình khí hậu.

c. Kết luận

Kết quả mô phỏng bằng mô hình mạng nhận thức đa lớp và mạng truy hồi đối với bài toán dự báo tổng xạ sử dụng đầu ra của mô hình khí hậu toàn cầu của NCEP là rất khả quan. Với bộ số liệu trong 10 năm được chia làm 3 phần: 70% cho luyện mạng, 20% cho kiểm chứng chéo và 10% dùng làm số liệu kiểm tra độc lập, sai số của mạng nhận thức đa lớp có phần lớn hơn so với mạng truy hồi nhưng sự khác biệt là không đáng kể (khoảng vài phần trăm đơn vị).

Chất lượng dự báo của mô hình mạng TKNT đã cải thiện được so với dự báo của mô hình khí hậu toàn cầu của NCEP khoảng 25%, đặc biệt hệ số tương quan tăng 45%. Qua đó có thể khẳng định, mạng TKNT đã nắm bắt được tương đối đầy đủ xu thế của chuỗi số liệu cũng như quan hệ giữa các nhân tố dự báo với yếu tố tổng xạ. Để ứng dụng bài toán trong dự báo thực tế, cần bổ sung nguồn số liệu dự báo của các mô hình khí hậu khu vực hoặc toàn cầu và sử dụng mạng nơron như một công cụ thống kê đầu ra mô hình cho các trạm kể trên.

Tài liệu tham khảo

1. Nguyễn Hương Điền, Hoàng Phúc Lâm. Dự báo tổng lượng bức xạ ngày cho khu vực đồng bằng phía bắc Việt Nam bằng phương pháp sử dụng mạng thần kinh nhân tạo, Tạp chí khoa học - Đại học Quốc gia Hà nội, tập 22, tr1-10, tr 9-19, 2006.
2. Hồ Thị Minh Hà, Nguyễn Hương Điền. Thử nghiệm dự báo lượng mưa ngày bằng quy tắc dùng mạng thần kinh nhân tạo hiệu chỉnh sản phẩm mô hình số, Tạp chí khoa học - Đại học Quốc gia Hà nội, tập 22, số 1, tr1-10.2006.
3. Nguyễn Đình Thúc, Hoàng Đức Hải. Mạng nơron: Quy tắc và ứng dụng. NXB GD, Hà nội 2000.
4. Bin Li (2002), *Spatial interpolation of weather variables using artificial neural network*, Master of Science, University of Georgia, Greece.
5. David Silverman, John A. Dracup. *Artificial neural networks and long-range precipitation prediction in California*, Journal of applied meteorology, vol 39 (Jan 2000), pp57-66.
6. Gianluigi Rech, *Forecasting with artificial neural network models*, SSE/EFI working paper series in economics and finance, No491, Jan 2002.
7. Marcelo C.Medeiros, Timo Terasvirta, Gianluigi Rech, *Building neural network models for time series: A statistical approach*, SSE/EFI working paper series in economics and finance, No508, Sep 2002.
8. Patrick van der Smagt, Ben Krose, *An Introduction To Neural Network*, 1996, 8th Edition, University of Amsterdam
9. Ricardo A. Guarnieri et al "Solar radiation forecast using Artificial Neural Networks in South Brazil", *Proceedings of 8 ICSHMO*, April 24 – 28, 2006, p.1777 – 1785.