Tạp chí Tin học và Điều khiển học, T. 14, S. 1, (1998), (59- 64)

MẠNG NƠ RON NHIỀU LỚP VỚI THUẬT TOÁN HỌC TRÊN CƠ SỞ LỌC SIÊU KALMAN MỞ RỘNG

VŨ NHƯ LÂN - VŨ CHẤN HƯNG - ĐẶNG THÀNH PHU

Abstract. New learning algorithm is suggested for computing the weights of the multilayer neural networks.

1. ĐẶT BÀI TOÀN

Mạng nơ ron nhân tạo được thiết kế theo nguyên tắc tương tự nơ ron sinh học và có khả năng giải quyết hàng loạt bài toàn như nhận dạng ảnh, mô hình hóa, điều khiển các đối tượng phi tuyến...

Quá trình học của mạng nơ ron nhiều lớp ẩn có thể được thực hiện bằng nhiều thuật toán khác nhau như các thuật toán lan truyền ngược (back-propagation algorithm) các thuật toán di truyền (genetic algorithm) hoặc các thuật toán tìm kiếm ngẫu nhiên...

Qúa trình học là quá trình xác định các trọng số liên kết tối ưu của mạng nơ ron. Trong điều kiện có giám sát (supervisor), tiêu chuẩn học có dạng:

$$\frac{1}{m}\sum_{k=1}^{m} Q(\varepsilon(W,k)) \tag{1.1}$$

trong đó:

W - véctơ trọng số mạng

k- thời điểm xuất tín hiệu tại lớp ra của mạng

Xét mạng nơ ron M lớp sau đây:

$$Y^{(M)} = f^{(M)}(f^{(M-1)}(\dots f^{(1)}(U)\dots)) \tag{1.2}$$

Trong đó

$U = [u^1, u^2, \dots, u^{N(0)}]$ - v - véc tơ vào (1.2a)

$U^{(0)} = [U^{(1,1)}, U^{(1,2)}, \dots, U^{(1,N(0))}]^T$ - véc tơ vào tại lớp đầu tiên (1.2b)

$U^{(M)} = [Y^{(M,1)}, Y^{(M,2)}, \dots Y^{(M,N(M))}]^T$ véc tơ ra tại lớp M (1.2c)

$Y^{(r)} = [Y^{(r,1)}, Y^{(r,2)}, \dots, U^{(r,N(r))}]^T$ - véc tơ ra tại lớp ẩn r (1.2d)

$r = \overline{1,(M-1)}$

Với; N(M) - số nơ ron tại lớp ra M

N(r) - số nơ ron tại lớp ẩn r

N(0) - số nơ ron tại lớp vào

Véctơ trọng số liên kết toàn mạng nơ ron:

$$W = [W^{(1)}, W^{(2)}, \dots, W^{(r)} \dots W^{(M)}]^T \tag{1.2e}$$

Trong đó:

$W^{(r)}$ - vec tơ trọng số liên kết của lớp r

ở đây:

$$W^{(r)} = [W^{(r,1)}, W^{(r,2)}, \dots, W^{(r,i)}, \dots W^{(r,N(M))}]^T \tag{1.2f}$$

với

$$W^{(r,i)} = [W^r_{i1}, W^r_{i2}, \dots, W^r_{iN(r-1)}]^T \tag{1.2g}$$

các thành phần W^r_{ij}, $j = \overline{1, N(r-1)}$ - là trọng số liên kết của nơ ron thứ j, $j = \overline{1, N(r-1)}$ của lớp (r-1) đến nơ ron thứ i của lớp r.

Sơ đồ hình 1 dưới đây mô tả mạng nơ ron M lớp dạng (1.2).

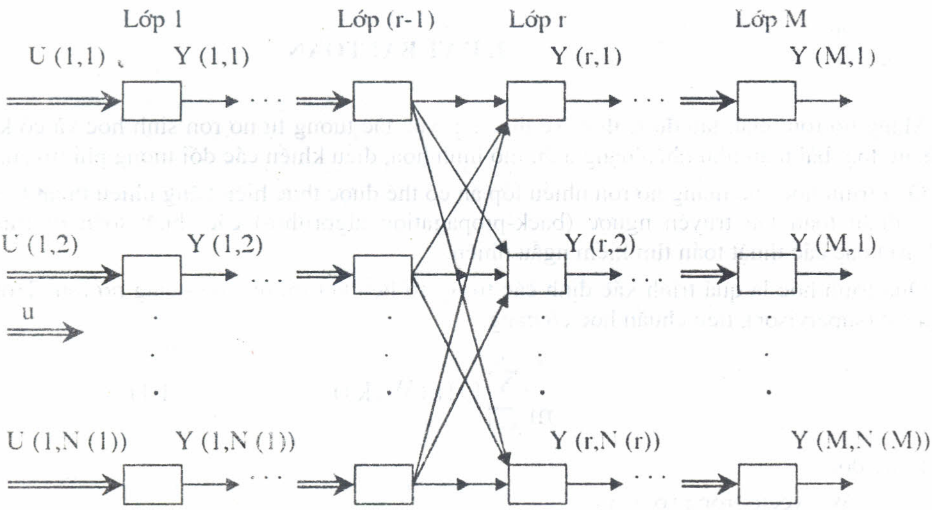
Tiêu chuẩn học Q phụ thuộc vào vec tơ sai số ε(W,m):

$$\varepsilon(W, k) = Y^{(M)}(k) - Y^*(k) \qquad (1.3)$$

Trong đó:

$Y^{(M)}(k)$ - vec tơ đầu ra tại thời điểm k, $m = \overline{1, k}$

$Y^*(k)$ - vec tơ giám sát tại thời điểm k



Hình 1. Sơ đồ mạng nơ ron M lớp dạng (1.2)

Thông thường tiêu chuẩn học Q có dạng bình phương:

$$Q(\varepsilon(W, k)) = \varepsilon^T(W, k)\, S\, \varepsilon(W, k) \qquad (1.4)$$

với S là ma trận xác định dương.

Từ những năm 90, nhiều thuật học đã sử dụng lọc Kakman mở rộng tối ưu hóa quá trình học [2,3,4]. Ưu việt của lọc Kalman mở rộng thể hiện ở tốc độ hội tụ và độ chính xác của ước lượng trọng số liên kết.

Lọc Kalman mở rộng cũng có thể được phát triển trên cơ sở lọc siêu Kalman [1] nhằm nâng cao chất lượng bài toán lọc với trạng thái mở rộng. Từ đó có thể nâng cao chất lượng đánh giá trọng số liên kết mạng nơ ron.

2- THUẬT TOÁN HỌC VỚI LỌC KALMAN MỞ RỘNG.

Mạng nơ ron nhiều lớp phản hồi xây dựng từ (1.2) có thể biểu diễn dưới dạng sau:

$$Y(k+1) = f(Y(l), W(k), U(k)) \qquad (2.1)$$

Lưu ý rằng:

$Y = [Y^{(M)}, Y^{(r)}]$ (2.1a)

$f = f^{(M)}(f^{(M-1)}(...(f^{(1)}(\ )...))$ (2.1b)

Trọng số W có thể viết dưới dạng:

$$W(k+1) = W(k) \qquad (2.2)$$

Khi đó Y(k) là vec tơ kích hoạt cỡ N của tất cả các nơ ron tại thời điểm k.

W(k) - vec tơ các trọng số liên kết của toàn mạng trên cơ sở (1.2e) ÷ (1.2g) thu được:

$$W(k) = [w^1_{11}, w^1_{12}, \dots w^1_{1N(0)}, W^1_{21}, W^1_{22}, \dots, W^1_{2N(0)}, \dots W^1_{N(1)1}, W^1_{N(1)2}, \dots W^1_{N(1)N(0)}$$

.

.

,

$$W^r_{11}, W^r_{12}, \dots, W^r_{1N(r-1)}, \dots, W^r_{N(r)1}, W^r_{N(r)2}, \dots, W^r_{N(r)N(r-1)}$$

.

.

.

$$W^M_{11}, W^M_{12}, \dots, W^M_{1N(M)}, \dots W^M_{N(M)1}, W^M_{N(M)2}, \dots W^1_{N(M)N(M-1)}] \quad (2.2a)$$

ở đây vec tơ W(k) có cỡ N(w):

$$N(w) = \sum_{j=1}^{M} N(j)N(j-1) \quad (2.2b)$$

Tổng số đầu ra của tất cả các lớp $N(w) = \sum_{j=1}^{M} N(r)$ là (2.2c)

U(k) - vec tơ vào tại thời điểm k dạng (1.2a)

Như vậy mạng nơ ron phản hồi (2.1) có N(0) đầu vào, N(w) trọng số liên kết, N(Y) đầu ra của các lớp trong đó có N(M) đầu ra của mạng.

Để sử dụng lọc Kalman mở rộng cần phải xây dựng trạng thái mở rộng như sau:

$$X(k) = \begin{bmatrix} Y(k) \\ W(k) \end{bmatrix} \quad (2.3)$$

vec tơ X có cỡ N(Y) + N(M)

Do N(M) < N(Y) nên vec tơ quan sát Z tại lớp ra với N(Y) nơ ron có thể biểu diễn theo trạng thái mới (2.3) tại thời điểm k với nhiều quan sát $V_z(k)$.

$$Z(k) = [Y^{(M,1)}(k), Y^{(M,2)}(k), \dots, Y^{(M,N(M))}(k), 0, 0, \dots, 0] + V_z(k)$$

hoặc:

$$Z(k) = H(k)X(k) + V_z(k) \quad (2.4)$$

trong đó:

$$H(k) = [I, 0, 0]^T \quad (2.4a)$$

$$V_z(k) \sim N(O, R(k)) \quad (2.4b)$$

Ma trận 0 thứ nhất trong biểu thức (2.4a) biểu diễn đặc trưng quan sát đầu ra tại các lớp ẩn $Y^{(r,N(r))}(k)$, $r = \overline{1, M-1}$

Ma trận 0 thứ hai trong biểu thức (2.4a) biểu diễn đặc trưng quan sát đối với các trọng số liên kết mạng.

Như vậy phương trình trạng thái mạng nơ ron phản hồi có dạng:

$$X(k+1) = \bar{f}(x(k), u(k)) \quad (2.5)$$

trong đó

$$\bar{f}^T = \begin{bmatrix} f \\ I \end{bmatrix} \quad (2.5a)$$

sau khi tuyến tính hóa hàm $\bar{f}$ theo trạng thái X trên cơ sở phân rã theo chuỗi Taylor lân cận ước lượng trạng thái x^ ta có:

$$X(k+1) = A(k)X(k) + U(k) \qquad (2.6)$$

trong đó:

$$A(k) = \left.\frac{\partial \bar{f}}{\partial X(k)}\right|_{X(K)=X^\wedge(k/k)} \qquad (2.6a)$$

$$U(k) = \bar{f}(x^\wedge(k/k), u(k) - \left.\frac{\partial \bar{f}}{\partial X(k)}\right|_{X(K)=X^\wedge(k/k)} X^\wedge(k/k) \qquad (2.6b)$$

phương trình quan sát (2.4) tại thời điểm (k+1):

$$Z(k+1) = H(k+1)X(k+1) + V_z(k+1) \qquad (2.7)$$

thuật toán học dựa trên lọc Kalman mở rộng với hệ (2.6) và quan sát (2.7) có dạng sau:

$$X^\wedge(k+1/k+1) = X^\wedge(k+1/k) + K(k)[Z(k+1) - H(k+1)X^\wedge(k+1/k)] \qquad (2.8)$$

$$X^\wedge(k+1/k) = \bar{f}(X^\wedge(k/k), u(k)) = A(k)X^\wedge(k/k) + U(k) \qquad (2.9)$$

với X(0) = X^(0/0)

$$K(k+1) = P(k+1)H^T(k+1)[H(k+1)P(k+1)H^T(k+1) + R(k+1)]^{-1} \qquad (2.10)$$

$$P(k+1/k) = A(k)P(k/k)A^T(k) + Q(k) \qquad (2.11)$$

với điều kiện P(0/0) = Po

Nếu lọc Kalman mở rộng phân ly, có thể chọn:

$$Q(k) = [x^\wedge(k/k) - x^\wedge(k/k-1)][X^\wedge(k/k) - X^\wedge(k/k-1)]^T + Q(k-1) \qquad (2.11a)$$

với Q(0) = 0

$$P(k+1/k+1) = P(k+1/k) + K(k+1)H(k+1)P(k+1/k) \qquad (1.12)$$

Lưu ý rằng:

$$P(k/k) = \begin{bmatrix} P_1(k/k) & P_2(k/k) \\ P_3(k/k) & P_4(k/k) \end{bmatrix} \qquad (2.13)$$

trong đó:

$$P_1(k/k) = E\{[Y(k) - Y^\wedge(k/k)][Y(k) - Y^\wedge(k/k)]^T\} \qquad (2.13a)$$

$$P_4(k/k) = E\{[W(k) - W^\wedge(k/k)][W(k) - W^\wedge(k/k)]\} \qquad (2.13b)$$

$$P_2(k/k) = P^T_3(k/k) = E\{[Y(k) - Y^\wedge(k/k)][W(k) - W^\wedge(k/k)]^T\} \qquad (2.13c)$$

$P_1(k/k)$ - diễn tả sự tương quan giữa các đầu ra của các lớp.

$P_4(k/k)$ - diễn tả sự tương quan giữa các trọng số liên kết.

$P_2(k/k)$ hoặc $P_3(k/k)$ diễn tả tương quan chéo giữa trọng số liên kết và đầu ra ẩn của các lớp.

3. THUẬT TOÁN HỌC VỚI LỌC SIÊU KALMAN MỞ RỘNG.

Sử dụng mô hình quan sát mới [1] tại mức lọc thứ hai:

$$Z_2(k+1) = P^{-1/2}(k+1/k+1)X(k+1) + V_2(k+1) \qquad (3.1)$$

với

$$V_2(k+1) \sim N(0, I) \qquad (3.1a)$$

Kết hợp mô hình động học (2.6) và (3.1) sau khi sử dụng định lý 1 [1] thu được:

$$\hat{X}_2(k+1/k+1) = \hat{X}_2(k+1/k) + \boldsymbol{K}_2(k+1)[Z_2(k+1) - P^{-1/2}(k+1/k+1)\hat{X}_2(k+1/k)] \quad \hat{X}_2(k+1/k) = \bar{f}(\hat{X}_2(k/k), U(k)) \qquad (3.3)$$

với $X(0) = \hat{X}(0/0) = \hat{X}_2(0/0)$

$$\boldsymbol{K}_2(k+1) = P_2(k+1/k)P^{-1/2}(k+1/k+1)[P^{-1/2}(k+1/k+1)\,P_2(k+1/k)P^{-1/2}(k+1/k+1)+I]^{-1} \qquad (3.4)$$

$$P_2(k+1/k) = A(k)P_2(k/k)A^T(k) \qquad (3.5)$$

với $Po = P(0/0) = P_2(0/0)$

$$P_2(k+1/k+1) = P_2(k+1/k) + \boldsymbol{K}_2(k+1)P^{-1/2}(k+1/k+1)P_2(k+1/k) \qquad (3.6)$$

Tiếp tục sử dụng mô hình quan sát mới tại các mức lọc 3,4,...,n, sẽ có:

$$Z_n(k+1) = P^{-1/2}_{n-1}(k+1/k+1)X(k+1) + V_n(k+1) \qquad (3.7)$$

với $V_n(k+1) \sim N(0, I)$ (3.7a)

Tương tự như trên thu được thuật học mức n:

$$\hat{X}_n(k+1/k+1) = \hat{X}_n(k+1/k) + \boldsymbol{K}_2(k+1)[Z_n(k+1) - P^{-1/2}_{n-1}(k+1/k+1)\hat{X}_n(k+1/k) \qquad (3.8)$$

$$\hat{X}_n(k+1/k) = \bar{f}(\hat{X}_n(k/k), U(k)) \qquad (3.9)$$

với $X(0) = \hat{X}(0/0) = \hat{X}_2(0/0) = \ldots = \hat{X}_n(0/0)$

$$\boldsymbol{K}_n(k+1) = P_n(k+1/k)P^{-1/2}_{n-1}(k+1/k+1)[P^{-1/2}_{n-1}(k+1/k+1)P_n(k+1/k)P^{-1/2}_{n-1}(k+1/k+1) + I]^{-1} \qquad (3.10)$$

$$P_n(k+1/k) = A(k)P_n(k/k)A^T(k) \qquad (3.11)$$

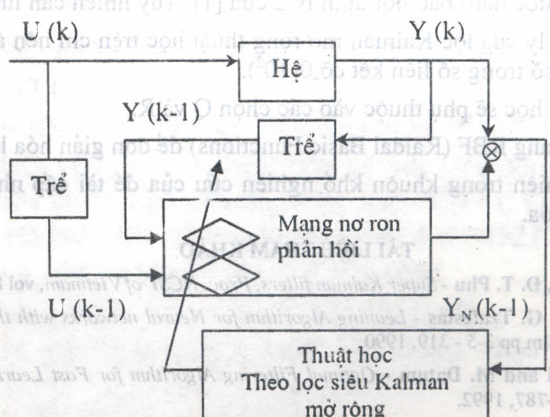
với $Po = P(0/0) = P_2(0/0) = \ldots = Pn(0/0)$

$$P_n(k+1/k+1) = P_n(k+1/k) + \boldsymbol{K}_n(k+1)P^{-1/2}_{n-1}(k+1/k+1)P_n(k+1/k) \qquad (3.12)$$

Tóm lại bằng cách trạng thái hóa đầu ra các lớp mạng nơ ron và trọng số liên kết, qua một số biến đổi, thu được mô hình thuận tiện cho việc áp dụng lọc Kalman mở rộng và sử dụng kết quả [1] để tính toán trọng số liên kết tối ưu của mạng phản hồi. Từ tính chất của định lý 2 trong [1] $P_1(k/k) > P_2(k/k) > \ldots > P_n(k/k) >$ suy ra tính ưu việt của việc sử dụng lọc siêu Kalman mở rộng cho quá trình học.

4- NHẬN DẠNG HỆ THỐNG THEO INPUT - OUTPUT.

Các kết quả ở phần 2 và 3 có thể sử dụng bài toán nhận dạng theo input - output. Hình 2 dưới đây mô tả sơ đồ nhận dạng



Hình 2. Sơ đồ nhận dạng trọng số liên kết thuật học với lọc siêu Kalman mở rộng.

Cho cặp $(U(k), Y(k))$, k=1,2, ... với n - cố mức lọc cần thiết.

Bước 1 - Xây dựng mạng nơ ron truy hồi với vec tơ trọng số liên kết W_N, có số lớp M và số phần tử nơ ron tại lớp r là N(r), r = $\overline{1, M}$

$$Y_N(k+1) = f(Y_N(k), W_N(k), U(k))$$

Bước 2- Xây dựng mô hình trạng thái mạng:

$$X(k+1) = \bar{f}\,(X(k),, U(k))$$

$$X = \begin{bmatrix} Y_N \\ W_N \end{bmatrix}$$

$$\bar{f}^T = \begin{bmatrix} f \\ I \end{bmatrix}$$

Xây dựng mô hình quan sát:

$$Z(k+1) = H(k+1)X(k+1) + V_Z(k+1)$$

$$H(k+1) = [\mathbf{I}, 0, 0]$$

$$V_z(k+1) \sim N(0, R(k+1))$$

Bước 3 - Cập nhật thông số liên kết theo lọc Kalman mở rộng, thu được cặp thống kê (X^(k/k), P(k/k)

Bước 4- Xây dựng mô hình quan sát

$$Z_i(k+1) = P^{-1/2}{}_{i-1}(k+1/k+1)X(k+1) + V_i(k+1)$$

$$V_i(k+1) \sim N(0, \mathbf{I})$$

Cập nhật trọng số liên kết theo lọc Kalman mở rộng [1], thu được:

$$(X\hat{\ }i(k/k), P_i(k/k)$$

Bước 5: i = i + 1

Nếu i < n trở lại bước 4

Còn không dừng thuật học - thu được cặp:

$$(X\hat{\ }_n(k/k), P_n(k/k))$$

& 5 - KẾT LUẬN

Bài toán đề cập đến vấn đề sử dụng lọc siêu Kalman mở rộng [1] cho quá trình học của mạng nơ ron phản hồi. Thuật học này cho phép nâng cao chất lượng học của mạng nơ ron so với các thuật [2,3,4].

Ngoài ra một thuật toán nhận dạng được đề xuất dựa trên cơ sở các kết quả thu được từ thuật học trên. Ưu việt của thuật toán được đảm bảo bởi định lý 2 của [1]. Tuy nhiên cần lưu ý các điểm sau đây:

1) - Do tính chất phân ly của lọc Kalman mở rộng thuật học trên chỉ nên áp dụng cho mạng nơ ron có số lớp từ 2 đến 3 với tổng số trọng số liên kết cỡ $0(10^2)$.

2)- Hiệu quả của thuật học sẽ phụ thuộc vào các chọn Q và R.

3) - Hàm f nên chọn dạng RBF (Raidal Basic Functions) để đơn giản hóa lọc Kalman mở rộng.

Công trình này thực hiện trong khuôn khổ nghiên cứu của đề tài cấp nhà nước KHCN - 04-09 t huộc chương trình Tự động hóa.

TÀI LIỆU THAM KHẢO

[1] - **V.N Lan, V. C. Hưng, Đ. T. Phu** - *Super Kalman filters, Proc. NCST of Vietmam*, vol 8, No1 (1996) (35-42)

[2] - **K. Watanabe and S. G. Tzafestas** - *Leaning Algorithm for Neural networks with the Kalman Filter. J. Intell and Robitic Syst.* Vol 3m NextChainLevelm pp 3-5 - 319, 1990.

[3] - **S. Shah, F. Palmieri and M. Datum** - *Optimal Filtering Algorithm for Fast Learning in Feed Forward Neural Networks - Neural Networks* 5: 779-787, 1992.

[4] - **Williams. R. J** - *Training Recurrent Networks Using the Extended Kalman Fillter - Proc. Int Joint Conf. Neural Networks.* Vol IV, 241-246, Baltimore. 1992.

[5] - **Singhal S. and L. Wu** - *Training Multilayer percepreons with the Extended Kalman Algorithm - Advances in Neural Informations Processing Systems.* pp 133-140, san Mateo, CA, 1989.