

ÁP DỤNG THUẬT TOÁN GIẢM SAI SỐ VÀ THUẬT TOÁN DI TRUYỀN CHO HỆ MỜ SAM XẤP XỈ VỚI TẬP MỜ DẠNG HÌNH THANG

Lê Ngọc Thanh

Trường Đại học Kinh tế Tp.HCM

(Bài nhận ngày 22 tháng 04 năm 2006, hoàn chỉnh sửa chữa ngày 14 tháng 09 năm 2006)

TÓM TẮT: Như hầu hết các mô hình biểu diễn tri thức, việc học điều chỉnh tham số và học tối ưu kiến trúc có ý nghĩa quyết định đến chất lượng và hiệu quả của hệ mờ SAM xấp xỉ. Bài báo này trình bày một số nghiên cứu ứng dụng các thuật toán giảm sai số và thuật toán di truyền cho quá trình học điều chỉnh tham số và tối ưu của hệ mờ SAM. Đồng thời triển khai kết quả nghiên cứu với hệ mờ SAM xấp xỉ sử dụng tập mờ hình thang và áp dụng trong dự báo chuỗi thời gian.

Từ khóa: Lý thuyết tập mờ, thuật toán học giảm sai số, thuật toán di truyền.

1. GIỚI THIỆU

Xấp xỉ phi tuyến là vấn đề được quan tâm trong lĩnh vực điều khiển học. Bài toán có thể được biểu như sau:

Cho

X là tập compact trên không gian R_n .

Y là tập compact trên không gian R_p .

$y = f(x) : X \rightarrow Y$ liên tục, giới hạn trên X .

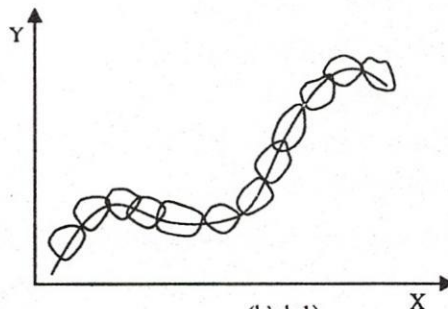
$\varepsilon > 0$ cho trước.

Yêu cầu

Xây dựng hệ thống $F(x) : X \rightarrow Y : \forall x \in X, |f(x) - F(x)| \leq \varepsilon$

Mô hình toán của f có thể chưa xác định, và được chỉ ra bằng tập các mẫu dữ liệu thống kê $\{xy, xy \in XY\}$.

Có nhiều cách tiếp cận khác nhau cho vấn đề xây dựng mô hình $F(x)$. Hệ mờ SAM [3][7] (Standard Additive Model) với hệ thống các luật mờ hoạt động phối hợp là một cách tiếp cận cho vấn đề này. Bằng việc sử dụng khối mờ hình thành từ luật mờ phủ lên một đoạn biến thiên trong không gian tích XY , các luật mờ của SAM cho phép lấp đầy toàn bộ đường biểu diễn của quan hệ $y = f(x)$. Cơ chế kích hoạt song song và kết hợp theo qui tắc cộng-tích trên các luật mờ của SAM đảm bảo ứng với mỗi giá trị đầu vào là một giá trị đầu ra xác định, cho phép xấp xỉ với kết quả của f . Tuy nhiên, chất lượng xấp xỉ của SAM tùy thuộc vào số luật mờ và cách phân bố của chúng trong không gian tích XY (hình 1).



(hình 1)

Hệ SAM có càng nhiều luật mờ thì chất lượng xấp xỉ càng cao. Nhưng nếu số luật mờ tăng mà phân bố không hợp lý thì sẽ làm giảm chất lượng xấp xỉ và tốc độ vận hành của SAM. Nghiên cứu các thuật toán học điều chỉnh tham số và học tối ưu cho hệ mờ SAM là nhằm giải quyết vấn đề này. Hiện nay đã có một số nghiên cứu cho vấn đề nêu trên, nổi bật

là [5] với việc sử dụng hệ SAM dựa trên tập mờ hình thang. Tuy nhiên [5] khởi tạo hàm hình thang cho tập mờ sau khi đã xác định hàm hình chuông (bellsharp) của nó. Do đó thuật toán mất nhiều thời gian. Hơn nữa, nếu số liệu học không liên tục thì mô hình tạo ra không đáp ứng được yêu cầu xấp xỉ liên tục. [5] tối ưu hệ luật mờ bằng cách tinh giản số tập mờ tương tự nhau nên chất lượng xấp xỉ của hệ mờ giảm sau khi tinh chỉnh. [1] chi tiết thuật toán học tham số trên các dạng hàm mờ nhưng phần tối ưu hệ luật chủ yếu dựa vào các đánh giá trên luật như: tính trùng lặp, tầm quan trọng để loại bỏ luật. Các chỉ số này cần một khối lượng tính toán có độ phức tạp rất lớn, đôi khi gần bằng phương pháp vét cạn. Trong nội dung tiếp theo, bài báo sẽ đề cập một số vấn đề về hệ mờ SAM, triển khai thuật toán giảm sai số cho quá trình học điều chỉnh tham số và thuật toán di truyền cho quá trình học tối ưu của SAM. Cuối cùng là ứng dụng các kết quả nghiên cứu cho phần mềm dự báo trong kinh tế.

2. HỆ MỜ SAM

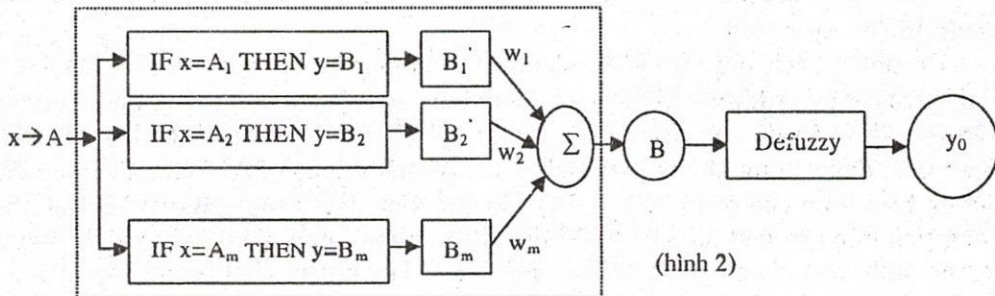
Hệ mờ SAM là hệ thống m luật mờ hoạt động song song và kết hợp theo mô hình cộng-tích. Gọi R_j là luật mờ thứ j của hệ SAM. Ta có:

$$R_j: \text{IF } X = A_j \text{ THEN } Y = B_j, \quad j = \overline{1, m}$$

A_j : Tập mờ đầu vào. A_j có thể là tập hợp n tập mờ thành phần $A_{ji}, i = \overline{1, n}$, tương ứng với n chiều của không gian vào của luật.

B_j : Tập mờ đầu ra. B_j có thể có nhiều tập mờ thành phần. Để đơn giản cho việc tính toán và cài đặt, chúng tôi giả sử B_j có một thành phần. Như vậy số chiều của không gian ra của luật là 1.

Kiến trúc và nguyên tắc hoạt động của mô hình hệ mờ SAM được biểu diễn như (hình 2)



- x : Giá trị vào, $x \in R_n$.
- y_0 : Giá trị ra của hệ thống, $y_0 \in R$.
- A : Giá trị mờ hóa của x .
- B_j : Tập mờ kết quả cho bởi luật R_j .
- w_j : Trọng số luật R_j trong hệ luật.
- Σ : Qui tắc cộng trong mô hình kết hợp cộng-tích của hệ mờ SAM
- B : Tập mờ kết quả của toàn hệ thống.

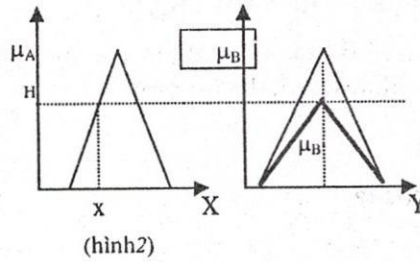
Với mỗi giá trị vào x , x được mờ hóa và kích hoạt tất cả các luật.

$$\text{Đặt } a_j(x) \text{ là mức kích hoạt luật } R_j \text{ của } x: a_j(x) = \prod_{i=1}^n a_j^i(x_i) \tag{1}$$

Với $a_j(x)$ là giá trị hàm liên thuộc mờ của tập mờ thành phần A_j trong A_j .

Giá trị tập mờ đầu ra của $R_j : B_j' = a_j(x).B_j$

Ví dụ, xét luật R_j . (Hình 3) sử dụng đường đơn biểu diễn tập mờ về trái A_j và về phải B_j . Giá trị x kích hoạt về trái luật $H = \mu_A(x)$. Tập mờ kết quả B' có $\mu_{B'}$ được xác định bằng đường nét đậm.



Tập mờ đầu ra B của hệ thống:

$$B = \sum_{j=1}^m w_j . B_j' = \sum_{j=1}^m w_j . a_j(x) . B_j \quad (3)$$

Đặt :

- b : Hàm liên thuộc mờ của tập mờ B.
- b_j : Hàm liên thuộc mờ của tập mờ B_j , $j = \overline{1, m}$.
- b_j' : Hàm liên thuộc mờ của tập mờ B_j' , $j = \overline{1, m}$.
- V_j : Kích thước khối mờ B_j .
- c_j : Trọng tâm khối mờ B_j .

Ta có:

$$V_j = \int_R b_j(y) dy \quad (4)$$

$$c_j = \frac{\int_R y . b_j(y) dy}{\int_R b_j(y) dy} \quad (5)$$

$$b(y) = \sum_{j=1}^m w_j . b_j'(y) = \sum_{j=1}^m w_j . a_j(x) . b_j(y) \quad (6)$$

Sử dụng phương pháp khử mờ trọng tâm [4] đối với tập mờ B để xác định giá trị ra y_0 :

$$\begin{aligned} y_0 = \text{Centroid}(B) &= \frac{\int_R y . b(y) dy}{\int_R b(y) dy} = \frac{\int_R y . \sum_{j=1}^m w_j . b_j'(y) dy}{\int_R \sum_{j=1}^m w_j . b_j'(y) dy} = \frac{\int_R y . \sum_{j=1}^m w_j . a_j(x) . b_j(y) dy}{\int_R \sum_{j=1}^m w_j . a_j(x) . b_j(y) dy} \\ &= \frac{\sum_{j=1}^m w_j . a_j(x) . \int_R y . b_j(y) dy}{\sum_{j=1}^m w_j . a_j(x) . V_j} \cdot \frac{\int_R y . b_j(y) dy}{\int_R b_j(y) dy} \\ &= \frac{\sum_{j=1}^m w_j . a_j(x) . \int_R b_j(y) dy}{\sum_{j=1}^m w_j . a_j(x) . V_j} \end{aligned} \quad (7)$$

Mô hình hệ mờ SAM: $y = F(x) : R_n \rightarrow R$

$$\text{Với : } F(x) = \frac{\sum_{j=1}^m w_j \cdot a_j(x) \cdot V_j \cdot c_j}{\sum_{i=1}^m w_i \cdot a_i(x) \cdot V_i} \quad (8)$$

3. ĐIỀU CHỈNH THAM SỐ HỆ MỜ SAM [3][7]

Giai đoạn học cấu trúc đã kiến tạo hệ luật mờ của mô hình SAM $y = F(x)$ xấp xỉ hàm phi tuyến $y = f(x)$. Sai số mô hình xác định như sau:

$$E(x) = \max\{[(y-F(x))^2]/2\} \quad (9)$$

Điều kiện chấp nhận được là:

$$E \leq \varepsilon, \quad \varepsilon > 0 \text{ cho trước} \quad (10)$$

Khi kiến trúc hệ mờ SAM được tạo, (10) có thể chưa thỏa. Học điều chỉnh tham số là để giúp SAM giải quyết vấn đề này. Bài toán học điều chỉnh tham số được phát biểu như sau:

Cho :

- Mô hình SAM với kiến trúc và các thông tin liên quan như (hình 2)

- $D = \{x_{yt} \in XY, t = \overline{1, ntd}\}$ tập hợp dữ liệu phản ánh quan hệ cần xấp xỉ $y = f(x) : X \rightarrow Y$.

Với ntd: Số mẫu dữ liệu trong D.

Yêu cầu: Điều chỉnh tham số của các $w_j, A_{ji}, B_j, j = \overline{1, m}, i = \overline{1, p}$; của hệ SAM để (10) được thỏa mãn.

Áp dụng thuật toán cho quy trình học giảm sai số (gradient descent) [3], cho tham số ξ của hệ mờ SAM, ta có mô hình học điều chỉnh: $\xi(t+1) = \xi(t) - \mu_t \frac{\partial E}{\partial \xi} \quad (11)$

$$\mu_t \text{ là hệ số học, xác định như sau [3]: } \mu_t = \mu_{\max} \left[1 - \frac{t}{1.1 * ntd} \right]$$

với μ_{\max} : Giá trị μ lớn nhất, được chọn trước.

Triển khai (11) cho các tham số của SAM tại luật R_j :

▪ Các tham số liên quan tập mờ về trái A_j :

$$\text{- Tập mờ } A_{ji} : \frac{\partial E}{\partial a_{ij}} = \frac{\partial E}{\partial F} \cdot \frac{\partial F}{\partial a_j} \cdot \frac{\partial a_j}{\partial a_{ji}} \quad (12)$$

▪ Các tham số liên quan tập mờ về phải B_j :

$$\text{- Diện tích: } \frac{\partial E}{\partial V_j} = \frac{\partial E}{\partial F} \cdot \frac{\partial F}{\partial V_j} \quad (13)$$

$$\text{- Trọng tâm: } \frac{\partial E}{\partial c_j} = \frac{\partial E}{\partial F} \cdot \frac{\partial F}{\partial c_j} \quad (14)$$

$$\text{▪ Trọng số luật: } \frac{\partial E}{\partial w_j} = \frac{\partial E}{\partial F} \cdot \frac{\partial F}{\partial w_j} \quad (15)$$

Với : $E(x) = \frac{1}{2} [f(x) - F(x)]^2$, suy ra: $\frac{\partial E}{\partial F} = -(f(x) - F(x)) = -\varepsilon(x)$. Kết hợp với (8), ta được:

$$\frac{\partial F}{\partial c_j} = \frac{w_j \cdot a_j(x) \cdot V_j \left(\sum_{i=1}^m w_i \cdot a_i(x) \cdot V_i \right)}{\left(\sum_{i=1}^m w_i \cdot a_i(x) \cdot V_i \right)^2} = \frac{w_j \cdot a_j(x) \cdot V_j}{\left(\sum_{i=1}^m w_i \cdot a_i(x) \cdot V_i \right)} = p_j(x) \quad (16)$$

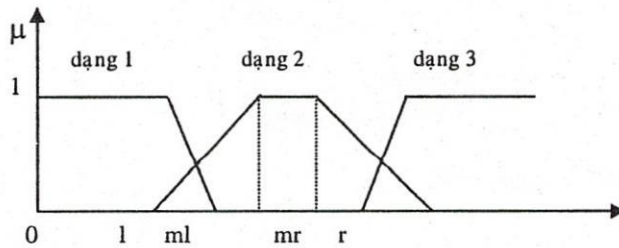
$$\frac{\partial F}{\partial V_j} = \frac{w_j \cdot a_j(x) \cdot c_j \left(\sum_{i=1}^m w_i \cdot a_i(x) \cdot V_i \right) - w_j \cdot a_j \left(\sum_{i=1}^m w_i \cdot a_i(x) \cdot V_i \cdot c_j \right)}{\left(\sum_{i=1}^m w_i \cdot a_i(x) \cdot V_i \right)^2} = \frac{w_j \cdot a_j(x) \cdot c_j - w_j \cdot a_j(x) \cdot F(x)}{\sum_{i=1}^m w_i \cdot a_i(x) \cdot V_i} = [c_j - F(x)] \frac{w_j \cdot a_j(x)}{\sum_{i=1}^m w_i \cdot a_i(x) \cdot V_i} \quad (17)$$

$$\frac{\partial F}{\partial w_j} = \frac{a_j(x) \cdot V_j \cdot c_j \left(\sum_{i=1}^m w_i \cdot a_i(x) \cdot V_i \right) - a_j(x) \cdot V_j \left(\sum_{i=1}^m w_i \cdot a_i(x) \cdot V_i \cdot c_j \right)}{\left(\sum_{i=1}^m w_i \cdot a_i(x) \cdot V_i \right)^2} = [c_j - F(x)] \frac{a_j(x) \cdot V_j}{\left(\sum_{i=1}^m w_i \cdot a_i(x) \cdot V_i \right)} \quad (18)$$

$$\frac{\partial F}{\partial a_j} = \frac{w_j \cdot V_j \cdot c_j \left(\sum_{i=1}^m w_i \cdot a_i(x) \cdot V_i \right) - w_j \cdot V_j \left(\sum_{i=1}^m w_i \cdot a_i(x) \cdot V_i \cdot c_j \right)}{\left(\sum_{i=1}^m w_i \cdot a_i(x) \cdot V_i \right)^2} = \frac{w_j \cdot V_j \left[c_j - \frac{\left(\sum_{i=1}^m w_i \cdot a_i(x) \cdot V_i \cdot c_j \right)}{\left(\sum_{i=1}^m w_i \cdot a_i(x) \cdot V_i \right)} \right]}{\left(\sum_{i=1}^m w_i \cdot a_i(x) \cdot V_i \right)} \quad (19)$$

$$\frac{\partial a_j}{\partial a_{ji}} = \frac{\partial \left(\prod_{k=1}^p a_{jk} \right)}{\partial a_{ji}} = \prod_{\substack{k=1 \\ k \neq i}}^p a_{jk} \quad (20)$$

Trong bài báo này, chúng tôi sử dụng đường biểu diễn dạng hình thang cho các tập mờ. Tập mờ A được biểu diễn với 4 tham số: A(l, ml, mr, r)(xem hình)



Với tập mờ A_{ji}, ta có: A_{ji} (l_{ji} , ml_{ji} , mr_{ji} , r_{ji})

Khai triển(12) trên các tham số A_{ji} :

$$\begin{aligned} \frac{\partial E}{\partial l_{ij}} &= \frac{\partial E}{\partial F} \cdot \frac{\partial F}{\partial a_j} \cdot \frac{\partial a_j}{\partial a_{ji}} \cdot \frac{\partial a_{ji}}{\partial l_{ij}} & \frac{\partial E}{\partial mr_{ij}} &= \frac{\partial E}{\partial F} \cdot \frac{\partial F}{\partial a_j} \cdot \frac{\partial a_j}{\partial a_{ji}} \cdot \frac{\partial a_{ji}}{\partial mr_{ij}} \\ \frac{\partial E}{\partial ml_{ij}} &= \frac{\partial E}{\partial F} \cdot \frac{\partial F}{\partial a_j} \cdot \frac{\partial a_j}{\partial a_{ji}} \cdot \frac{\partial a_{ji}}{\partial ml_{ij}} & \frac{\partial E}{\partial r_{ij}} &= \frac{\partial E}{\partial F} \cdot \frac{\partial F}{\partial a_j} \cdot \frac{\partial a_j}{\partial a_{ji}} \cdot \frac{\partial a_{ji}}{\partial r_{ij}} \end{aligned}$$

Gọi xyt là mẫu dữ liệu học ở thời điểm t. Ta có:

▪ $x_t \leq l_{ji}, x_t \geq r_{ji}, ml_{ji} \leq x_t \leq mr_{ji}$: vì $a_{ji} = 0, \frac{\partial a_{ji}}{\partial l_{ji}} = \frac{\partial a_{ji}}{\partial ml_{ji}} = \frac{\partial a_{ji}}{\partial mr_{ji}} = \frac{\partial a_{ji}}{\partial r_{ji}} = 0$

▪ $l_{ji} < x_t < ml_{ji}$: $a_{ji} = \frac{x_t - l_{ji}}{ml_{ji} - l_{ji}}$

$\frac{\partial a_{ji}}{\partial mr_{ji}} = \frac{\partial a_{ji}}{\partial r_{ji}} = 0$; $\frac{\partial a_{ji}}{\partial l_{ji}} = \frac{x_t - ml_{ji}}{(ml_{ji} - l_{ji})^2}$; $\frac{\partial a_{ji}}{\partial ml_{ji}} = \frac{l_{ji} - x_t}{(ml_{ji} - l_{ji})^2}$

▪ $mr_{ji} < x_t < r_{ji}$: $a_{ji} = \frac{r_{ji} - x_t}{r_{ji} - mr_{ji}}$

$\frac{\partial a_{ji}}{\partial ml_{ji}} = \frac{\partial a_{ji}}{\partial l_{ji}} = 0$; $\frac{\partial a_{ji}}{\partial r_{ji}} = \frac{x_t - mr_{ji}}{(r_{ji} - mr_{ji})^2}$; $\frac{\partial a_{ji}}{\partial mr_{ji}} = \frac{x_t - r_{ji}}{(r_{ji} - mr_{ji})^2}$

Thay các kết quả tính toán ở trên vào các công thức (12), (13), (14), (15). Áp dụng (11)

$c_j(t+1) = c_j(t) + \mu_t \cdot \varepsilon(x) \cdot p_j(x)$ (21)

$V_j(t+1) = V_j(t) + \mu_t \cdot \varepsilon(x) \cdot [c_j - F(x)] \frac{p_j(x)}{V_j}$ (22)

$w_j(t+1) = w_j(t) + \mu_t \cdot \varepsilon(x) \cdot [c_j - F(x)] \frac{p_j(x)}{w_j}$ (23)

$l_{ji} < x_t < ml_{ji}$: (24)

$l_{ji}(t+1) = l_{ji}(t) + \mu_t \cdot \varepsilon(x) \cdot [c_j - F(x)] \frac{p_j(x)}{a_j(x)} \prod_{\substack{k=1 \\ k \neq i}}^p a_{jk} \frac{x_t - ml_{ji}}{(ml_{ji} - l_{ji})^2}$

$ml_{ji}(t+1) = ml_{ji}(t) + \mu_t \cdot \varepsilon(x) \cdot [c_j - F(x)] \frac{p_j(x)}{a_j(x)} \prod_{\substack{k=1 \\ k \neq i}}^p a_{jk} \frac{l_{ji} - x_t}{(ml_{ji} - l_{ji})^2}$

$mr_{ji} < x_t < r_{ji}$: (25)

$r_{ji}(t+1) = r_{ji}(t) + \mu_t \cdot \varepsilon(x) \cdot [c_j - F(x)] \frac{p_j(x)}{a_j(x)} \prod_{\substack{k=1 \\ k \neq i}}^p a_{jk} \frac{x_t - mr_{ji}}{(r_{ji} - mr_{ji})^2}$

$mr_{ji}(t) = mr_{ji}(t) + \mu_t \cdot \varepsilon(x) \cdot [c_j - F(x)] \frac{p_j(x)}{a_j(x)} \prod_{\substack{k=1 \\ k \neq i}}^p a_{jk} \frac{x_t - r_{ji}}{(r_{ji} - mr_{ji})^2}$

Thuật toán điều chỉnh tham số hệ SAM

b1. Cho trước ε .

b2. $t = 0$.

b3. Đọc vào mẫu dữ liệu xyt, Tính $yt = F(xt)$ và $et = ydt - yt$.

b4. Điều chỉnh các tham số của hệ luật theo các công thức (21), (22), (23), (24), (25).

b5. $t = t+1$.

b6. Nếu $t \leq ntd$ thì đến b3.

b7. Tính E theo công thức (9).

b8. Nếu $E > \varepsilon$ thì đến b2.

b9. Dừng thuật toán.

4. TỐI ƯU HỆ LUẬT MỜ

Tối ưu hệ mờ SAM là loại bỏ các luật mờ ứng với các khối mờ không có ý nghĩa xấp xỉ nhằm giảm kích thước, tăng tốc độ hoạt động nhưng vẫn duy trì chất lượng xấp xỉ của SAM. Nếu dùng phương pháp vét cạn để học tối ưu hệ mờ SAM có m luật thì số tổ hợp luật phải xét là:

$$N = C_m^1 + C_m^2 + C_m^3 + \dots + C_{mm}^{m-1} + C_{mm}^m, \text{ với } c_m^k = \frac{m!}{k!(m-k)!}$$

Rõ ràng N khá lớn. Như thế phương pháp vét cạn là không khả thi. Bài báo này trình bày một giải pháp cải tiến cho vấn đề trên [2], đó là sử dụng thuật toán di truyền.

▪ **Nhiệm sắc thể**: Thông tin di truyền sử dụng trong bài toán này là tính lợi ích của các luật mờ trong hệ SAM. Thông tin đó được mã hóa bằng nhiệm sắc thể dạng vector có m thành phần tương ứng với m luật trong SAM. Thành phần thứ i bất kỳ chỉ nhận giá trị 0 hoặc 1. Giá trị 1 xác định luật Ri tồn tại trong hệ, giá trị 0 mang ý nghĩa ngược lại.

Ví dụ: [1 0 1 0 0 1 1] là một nhiệm sắc thể ứng với hệ SAM có 7 luật. Nhiệm sắc thể này chỉ xác nhận sự tồn tại của các luật R_1, R_3, R_6 và R_7 trong hệ SAM.

▪ **Thế hệ**: Tập hợp chứa các nhiệm sắc thể, có 10 phần tử. Ở thế hệ đầu tiên, các nhiệm sắc thể được khởi tạo ngẫu nhiên nhưng luôn luôn có một nhiệm sắc thể mà các gene đều mang giá trị 1 (26). Nhiệm sắc thể này đảm bảo quá trình tối ưu không sai lầm.

▪ **Hàm thích nghi (fitness)**: Trên cơ sở cân nhắc giữa số luật sử dụng và chất lượng xấp xỉ trong quá trình tối ưu, nguyên lý Kalman được [2] vận dụng để xây dựng hàm fitness:

$$\text{Fit}(m) = \ln(\bar{\sigma}_\varepsilon^2) + \frac{\log_{\text{ntd}}(m_1)}{\text{ntd}} \quad (27)$$

m_1 : Số phần tử bằng 1 trong nhiệm sắc thể.

$$\bar{\sigma}_\varepsilon^2 = \frac{1}{\text{ntd}} \sum_{j=1}^n (y_j^d - F(x_j))^2 \quad (28)$$

▪ **Lai tạo**: Thực hiện lai tạo tại điểm giữa của hai nhiệm sắc thể bố và mẹ. Xác suất lai tạo cá thể mới là $P_c = 0.5$. (29)

▪ **Đột biến**: Gene đột biến trên mỗi nhiệm sắc thể được chuyển giá trị từ 0 sang 1 hoặc ngược lại. Xác suất thực hiện đột biến gen là $P_m = 0.01$. (30)

▪ **Chọn lọc tự nhiên**: Do hàm thích nghi ở trên có tính hội tụ cao nên phương pháp chọn lọc tự nhiên được áp dụng trong thuật toán này là phương pháp xếp hạng (ranking).

Thuật toán tối ưu hệ SAM

b1. Khởi tạo thế hệ đầu tiên thỏa (26).

b2. Thực hiện lai tạo theo (29)

Đột biến theo (30)

b3. Dùng phương pháp xếp hạng cho chọn lọc tự nhiên dựa trên hàm fitness (27).

b4. Nếu điều kiện dừng chưa thỏa, đến b1

b5. Dùng nhiệm sắc thể trội để loại các luật không cần thiết trong hệ mờ SAM.

b6. Dừng thuật toán.

5. ỨNG DỤNG DỰ BÁO GIÁ CẢ HÀNG HÓA

Các kết quả nghiên cứu trong bài báo này được vận dụng vào phần mềm “Ứng dụng SAM cho dự báo chuỗi thời gian – Times-SAM” [6]. Nội dung nghiên cứu của bài báo cũng góp phần ứng dụng trong đề tài “PHÂN TÍCH DỰ BÁO KINH TẾ ỨNG DỤNG TRONG NGÀNH CÔNG NGHIỆP TẠI TP HCM”, Viện Kinh Tế TP.HCM năm 2003.

Phần mềm [6] được thiết cho việc xây dựng mô hình SAM dự báo giá cả của một số hàng hóa như: vàng, dollar, ciment, gạo. Mô hình SAM dự báo giá vàng được trình bày dưới đây là một trong những kết quả của phần mềm.

Qua việc phối hợp nghiên cứu với Viện Kinh Tế thành phố Hồ Chí Minh, chúng tôi nhận được bộ dữ liệu với 2327 mẫu về giá vàng mỗi ngày trong các năm 93-97 và 99-2000 tại thành phố Hồ Chí Minh. Với giả định chuỗi giá vàng/ngày là chuỗi số biến thiên theo thời gian, các chuyên gia ở Viện đã sử dụng phương pháp Arima (trong kinh tế lượng) để xây dựng mô hình dự báo giá vàng. Mô hình mà các chuyên gia đưa ra như sau:

$$D(t) = 0.139 + 0.056[D(t-1) - D(t-2)] \pm 0.262 + D(t-1)$$

$D(t)$, $D(t-1)$, $D(t-2)$ lần lượt là giá vàng tại thời điểm khảo sát, trước thời điểm khảo sát 1 và 2 ngày.

Nhằm xây dựng mô hình máy học thay thế mô hình kinh tế lượng nói trên, chúng tôi sử dụng hệ SAM có hai đầu vào tương ứng với $D(t-2)$, $D(t-1)$ và một đầu ra tương ứng với $D(t)$. Giả sử kết quả học cấu trúc (một tiện ích của phần mềm) đã xây dựng hệ mờ SAM có 51 luật mờ (bảng 2) trong không gian vào-ra: $\text{vang}(t-2) \times \text{vang}(t-1) \times \text{vang}(t)$.

Sau đây là kết quả áp dụng các thuật toán học điều chỉnh tham số và học tối ưu đối với mô hình SAM dự báo giá vàng. Bộ dữ liệu thử nghiệm được chọn ngẫu nhiên từ dữ liệu thống kê. Cơ sở của việc lựa chọn là mô hình dự báo do các chuyên gia Viện Kinh Tế tp.Hồ Chí Minh thực hiện. Bộ dữ liệu thử nghiệm như sau:

Bảng 1

Stt	Vàng(t-2)	Vàng(t-1)	Vàng(t)	Stt	Vàng(t-2)	Vàng(t-1)	Vàng(t)
1	482	481	480	11	481	483	484
2	481	480	481	12	483	484	488
3	480	481	482	13	484	488	485
4	481	482	484	14	488	485	487
5	482	484	486	15	485	487	488
6	484	486	484	16	487	488	487
7	486	484	484	17	488	487	486
8	484	484	481	18	487	486	494
9	484	481	481	19	486	494	492
10	481	481	483	20	494	492	491

Chọn chiều dài dự báo là 30 ngày, lấy thời điểm cuối của bộ dữ liệu thử nghiệm làm thời điểm hiện tại. Các kết quả thực nghiệm với ứng dụng [6] như sau :

Chọn ngân *Data Source* để đọc dữ liệu học và dữ liệu thử nghiệm vào chương trình.

▪ **Chất lượng của mô hình SAM sau giai đoạn học cấu trúc :**

Dùng mô hình SAM sau học cấu trúc để dự báo. Đường biểu diễn của (đồ thị 1) phản ánh kết quả dự báo, (bảng 3) là các chỉ số so sánh giữa SAM và mô hình kinh tế lượng:

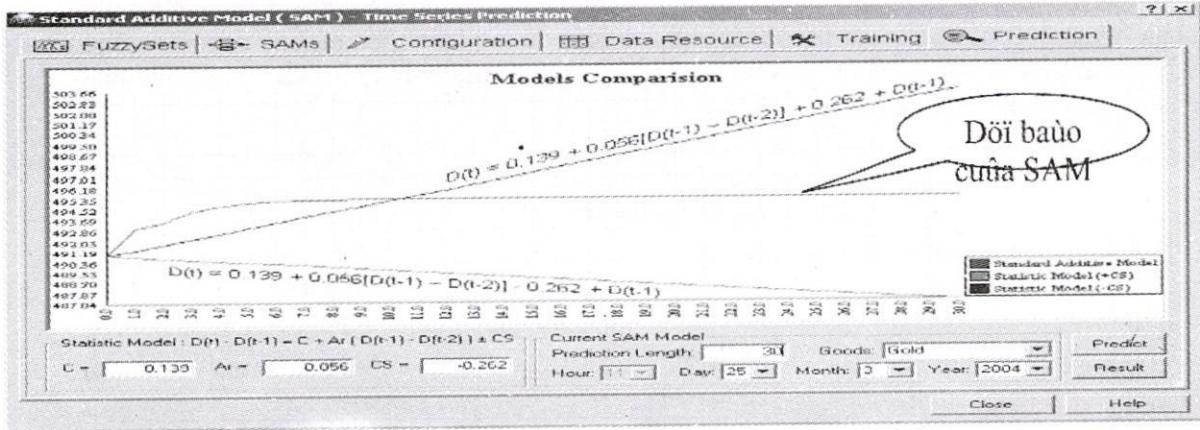
Bảng 2

No	Trọng số	Vế trái	Vế phải
1	0.0353	$[\text{vang}(t-2) \text{FZ-8}][\text{vang}(t-1) \text{FZ-8}]$	$[\text{vang}(t) \text{FZ-9}]$
2	0.0822	$[\text{vang}(t-2) \text{FZ-9}][\text{vang}(t-1) \text{FZ-9}]$	$[\text{vang}(t) \text{FZ-9}]$
3	0.0241	$[\text{vang}(t-2) \text{FZ-9}][\text{vang}(t-1) \text{FZ-10}]$	$[\text{vang}(t) \text{FZ-10}]$
4	0.1351	$[\text{vang}(t-2) \text{FZ-10}][\text{vang}(t-1) \text{FZ-10}]$	$[\text{vang}(t) \text{FZ-10}]$
5	0.0499	$[\text{vang}(t-2) \text{FZ-11}][\text{vang}(t-1) \text{FZ-10}]$	$[\text{vang}(t) \text{FZ-11}]$
...			
47	0.0013	$[\text{vang}(t-2) \text{FZ-8}][\text{vang}(t-1) \text{FZ-8}]$	$[\text{vang}(t) \text{FZ-10}]$
48	0.0004	$[\text{vang}(t-2) \text{FZ-8}][\text{vang}(t-1) \text{FZ-10}]$	$[\text{vang}(t) \text{FZ-12}]$

49	0.0004	[vang(t-2) FZ-11][vang(t-1) FZ-11]	[vang(t) FZ-11]
50	0.0004	[vang(t-2) FZ-11][vang(t-1) FZ-11]	[vang(t) FZ-9]
51	0.0004	[vang(t-2) FZ-11][vang(t-1) FZ-9]	[vang(t) FZ-9]

Mỗi tập mờ gồm [tên cơ sở | tên tập mờ]. FZ_xxx : Tên tập mờ tạo tự động bởi phần mềm

Đồ thị 1



Bảng 3

Các chỉ số thống kê	SAM	Kinh tế lượng	
		+0.262	-0.262
Giá trị lớn nhất	494.19	494.32	494.85
Giá trị nhỏ nhất	480.10	479.82	480.35
Giá trị trung bình	484.30	484.80	485.33
Sai số trung bình	2.10	1.96	1.91
Sai số phần trăm trung bình	0.43	0.40	0.39
Sai số T bình bình phương	7.77	6.81	6.47

Bảng 4

Các chỉ số thống kê	SAM	Kinh tế lượng	
		+0.262	-0.262
Giá trị lớn nhất	492.96	494.32	494.85
Giá trị nhỏ nhất	481.15	479.82	480.35
Giá trị trung bình	485.26	484.80	485.33
Sai số trung bình	1.78	1.96	1.91
Sai số phần trăm trung bình	0.37	0.40	0.39
Sai số T bình bình phương	5.96	6.81	6.47

Bảng 3 cho thấy chất lượng dự báo của SAM kém hơn so với mô hình kinh tế lượng mặc dù trong giai đoạn dài hạn, theo các chuyên gia viện kinh tế tp.HCM, dự báo này có thể tốt hơn do đường biểu diễn dự báo đi vào vùng giới hạn bởi hai đường biểu diễn của mô hình kinh tế lượng.

❖ **Áp dụng học tham số cho mô hình** : Sử dụng ngăn Configuration của phần mềm ứng dụng để ấn định chế độ học tham số của mô hình SAM.

Chọn chế độ học tham số (parameters learning) với các ấn định:

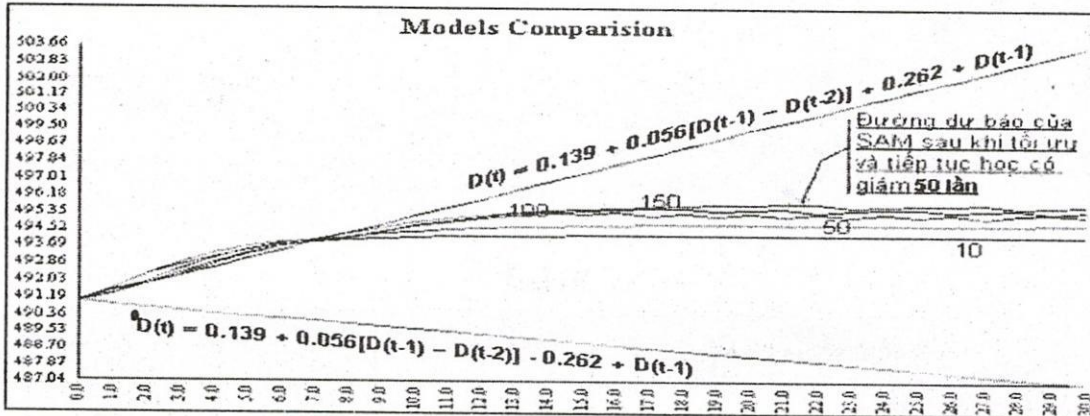
- **All fuzzy sets on the left of fuzzy Rule** : Chỉnh các tập mờ về trái của hệ luật
- **The Fuzzy Sets output of fuzzy Rule** : Chỉnh tập mờ về phải của hệ luật.

- **Weight of Rule :** Chính trọng số của các luật mờ.
- **Iteration :** Chọn số lần lặp trên bộ dữ liệu học.

Áp dụng với số lần học là 10. Sử dụng mô hình SAM kết quả cho dự báo. (bảng 4) cung cấp các chỉ số so sánh giữa mô hình SAM và mô hình kinh tế lượng. Rõ ràng mô hình SAM nhận được có chất lượng tốt hơn (sai số thấp hơn theo mọi chỉ số đánh giá) so với mô hình kinh tế lượng và cả mô hình SAM nhận được sau giai đoạn học cấu trúc (bảng 3).

Tiếp tục huấn luyện SAM với 50, 100, 150 lần, đồng thời thực hiện dự báo sau mỗi đợt huấn luyện. Kết quả quan sát (đồ thị 2) cho thấy các đường biểu diễn dự báo của SAM ứng với các đợt huấn luyện (đánh số bằng số lần lặp) có xu hướng đi vào giữa vùng cánh quạt của mô hình kinh tế lượng.

Đồ thị 2



❖ **Giảm kích thước SAM bằng học tối ưu hệ luật**

Trong ngăn Configuration, chọn cơ chế học tối ưu hệ luật (Fuzzy System Optimization).

- **Number of Generation :** Ấn định số thế hệ lặp. Chọn bằng 60
- **Auto-Delete redundant rules :** Chọn để thực hiện cơ chế tự động hủy bỏ luật thừa. Nếu không chọn mục này, các luật thừa sẽ được gán trọng số bằng 0.

Sau quá trình học, hệ SAM chỉ còn 42 luật.

Các luật bị hủy bỏ có số thứ tự lần lượt là 2,4,7,8,22,33,37,40,50 (xem bảng 2).

Do đặc điểm về quan hệ giữa số luật và chất lượng xấp xỉ được thể hiện trong hàm fitness (27), quá trình học tối ưu có thể gây tác động nhỏ đến độ chính xác của mô hình SAM. Khi đó, quá học tham số sẽ giúp khắc phục sự cố nói trên.

Sau quá trình học tối ưu mô hình SAM, chúng tôi cho hệ SAM học có điều chỉnh tham số với số lần lặp là 50. Tiến hành dự báo với mô hình SAM vừa tìm được, ta có đường biểu diễn kết quả dự báo của mô hình này như trong (đồ thị 2). Các chỉ số so sánh cuối cùng như sau

Bảng 5

Các chỉ số thống kê	SAM	+CS	-CS
Giá trị lớn nhất	492.96	494.32	494.85
Giá trị nhỏ nhất	481.15	479.82	480.35
Giá trị trung bình	485.26	484.80	485.33
Sai số trung bình	1.96	1.96	1.91
Sai số phần trăm tr. bình	0.40	0.40	0.39
Sai số t. bình bình phương	6.58	6.81	6.47

Tuy các chỉ số so sánh tương đương nhau nhưng đường dự báo của SAM được đánh giá là tốt hơn, và đáng điệu của nó phù hợp với qui luật khách quan của giá vàng: “*nếu không có sự thay đổi về chính sách và khủng hoảng kinh tế thì khi giá vàng tăng sẽ chỉ đến một mức nhất định, sau đó nó dao động trong một giới hạn xác định*”.

6. KẾT LUẬN

Xấp xỉ hàm phi tuyến là vấn đề của điều khiển học được nhiều nhà nghiên cứu thuộc lĩnh vực thông minh nhân tạo quan tâm. Bên cạnh mạng neuron, hệ mờ SAM là một công cụ cho phép xây dựng các mô hình máy học cho vấn đề nói trên. Mặc dù có những hạn chế nhất định về giới hạn giá trị của các đại lượng trong mô hình, nhưng hệ SAM có những ưu điểm là tính gần gũi với người dùng không chuyên về tin học, khả năng diễn đạt các khái niệm ngôn ngữ tự nhiên, kiến trúc hệ luật thuận tiện cho quá trình giải thích, và nhất là tốc độ hội tụ của các thuật toán cho quá trình học rất cao.

Với mong muốn nghiên cứu và ứng dụng công nghệ thông minh nhân tạo vào lĩnh vực kinh tế, trong thời gian qua, tác giả bài báo này đã tập trung xây dựng một số mô hình kinh tế học, đồng thời phát triển và hoàn thiện các công cụ máy học phù hợp cho việc cài đặt mô hình. Hệ mờ SAM và những nghiên cứu trình bày trong bài báo này là một trong những vấn đề của định hướng nghiên cứu nói trên.

APPLICATION OF GRADIENT DESCENT AND GENETIC ALGORITHMS TO THE TRAPEZOIDAL FUZZY SET BASED APPROXIMATE SAM

Le Ngoc Thanh

University of Economic in HoChiMinh city

ABSTRACT: *Parameter learning and Optimal learning, as well to all other knowledge presentative models, have an essential sense to the quality and effectivity in approximation of SAM. This paper presents not only an applied research on gradient descent algorithm and genetic algorithm to the parameter learning process and the optimal learning process of trapezoidal fuzzy set based SAM system, but also, an application of the research to the time-series prediction.*

Keywords: Fuzzy set theory, Gradient Descent Learning algorithm, Genetic Algorithm.

TÀI LIỆU THAM KHẢO

- [1]. Rudolf Kruse - Adolf Grauel, *Data Analysis with Neuro-Fuzzy Methods*, (2000).
- [2]. http://neuron.tuke.sk/~hudecm/PDF_PAPERS/DataalanyisWithNeuroFuzzySystem s.pdf
- [3]. Liang Wang and John Yen, *Extracing fuzzy rules for system modeling using a hibrid of genetic algorithms and Kalman filter*, Fuzzy Sets and System Vol.101 (1999).
- [4]. Bart Kosko, *Fuzzy Engineering*, Prentice Hall International Inc. (1997)
- [5]. Chin-Teng Lin and C.S. George Lee, *Neural Fuzzy System*, Prentice Hall, (1996).
- [6]. M.Y. Chen, D.A. Linkens, *Rule-base self-generation and simplification for data-driven fuzzy models*, Fuzzy Sets and Systems 142 243–265., (2004).

- [7]. Lê Ngọc Thanh, *Phần mềm Ứng dụng mô hình SAM cho dự báo chuỗi thời gian*, Đề tài phân tích dự báo kinh tế ứng dụng trong ngành công nghiệp tại tp hcm, Sở KH-CN Tp.HCM (2003).
- [8]. Lê Ngọc Thanh, *Ứng dụng logic mờ cho Mô hình thực tế ảo Thị trường các Doanh nghiệp cạnh tranh và Phần mềm Trò Chơi Kinh Doanh*, Luận văn Thạc sĩ Công Nghệ Thông Tin - Trường Đại Học Khoa Học Tự Nhiên, Đại Học Quốc Gia Thành phố Hồ Chí Minh (2000).