

ÁP DỤNG HỆ SUY DIỄN MỜ DỰA TRÊN MẠNG THÍCH NGHI ĐỂ XỬ LÝ TRI THỨC TRONG DỰ BÁO THỊ TRƯỜNG CHỨNG KHOÁN

Trương Hải Bằng

Trung tâm phát triển Công nghệ Thông tin – Đại học Quốc gia Tp. Hồ Chí Minh
(Bài nhận ngày 12 tháng 11 năm 2001, hoàn chỉnh sửa chữa ngày 06 tháng 12 năm 2001)

TÓM TẮT: Bài báo giới thiệu một ứng dụng của công cụ máy học trong điều khiển tài chính. Sử dụng các hệ thống minh trong dự báo tài chính có ý nghĩa thực tiễn rất lớn, đặc biệt trong lĩnh vực thị trường chứng khoán. Mô hình sử dụng là Hệ suy diễn mờ dựa trên mạng thích nghi. Kết quả thực nghiệm trên dữ liệu của thị trường chứng khoán New York (NYSE) rất đáng tin cậy, tuy nhiên thời gian huấn luyện mô hình phụ thuộc vào cấu hình của hệ thống máy tính.

1. Giới thiệu

Các công nghệ máy học là một trong những cách tiếp cận mới trong lĩnh vực tài chính gọi là điều khiển học tài chính (*Financial Cybernetics-FC*). Kết quả nghiên cứu và thực nghiệm đã đưa ra những kết quả quan trọng. [3], [4], [8].

Đặc điểm chính của các kỹ thuật này trong tài chính là xử lý tri thức và vận dụng chúng vào công tác dự báo. Bài báo trình bày một cách tiếp cận dùng lợi thế của cả 2 phương pháp (mạng thích nghi và hệ suy diễn mờ) trong việc xây dựng mô hình dự đoán – gọi là hệ lai (Hybrid System).

Hệ suy diễn mờ dựa trên mạng thích nghi (Adaptive Network based Fuzzy Inference System) sử dụng luật học lai để xây dựng các hàm thành viên và sinh ra tập các luật mờ if then nhằm xấp xỉ một tập dữ liệu mong muốn.

2. Bài toán dự báo chứng khoán:

Trong lĩnh vực chứng khoán các nhà đầu tư cổ phiếu phải nắm vững các vấn đề bản chất của việc phân tích và dự báo giá cổ phiếu, với mục tiêu là đưa ra dự báo về giá cổ phiếu và xu hướng giá cổ phiếu trong tương lai, tìm khả năng sinh lời cao để đi đến quyết định đầu tư cổ phiếu có hiệu quả nhất.

Đầu tư chứng khoán với tính chất sinh lợi và rủi ro cao, nhà đầu tư thường sử dụng một lượng tiền khá lớn để kinh doanh chứng khoán do đó họ rất quan tâm đến diễn biến giá cổ phiếu. Nếu dự báo đúng sẽ mang lại thành công lớn, ngược lại sẽ bị thất bại có khi dẫn đến phá sản. Chính vì vậy phân tích cổ phiếu đã trở thành một ngành khoa học và có xu hướng ngày càng phát triển.

Bài báo áp dụng hệ suy diễn mờ dựa trên mạng thích nghi để xử lý tri thức trên dữ liệu lịch sử - dự báo dữ liệu tương lai. Kết quả thử nghiệm thành công trên dữ liệu của thị trường chứng khoán NYSE và chúng tôi đang tiếp tục triển khai mô hình trên thị trường chứng khoán Việt nam.

3. Cơ sở lý thuyết:

3.1. Luật mờ if-then và hệ thống suy diễn mờ:

3.1.1. Luật mờ if then takagi sugeno's

Các luật được sử dụng trong hệ ANFIS có dạng:

Rule 1: If x is A_1 and y is B_1 , then $f_1 = p_1x + q_1y + r_1$, (1)

Rule 2: If x is A_2 and y is B_2 , then $f_2 = p_2x + q_2y + r_2$, (2)

Trong đó:

x và y các biến input.

A_i and B_i là các tập mờ tương ứng trong các hàm thành viên.

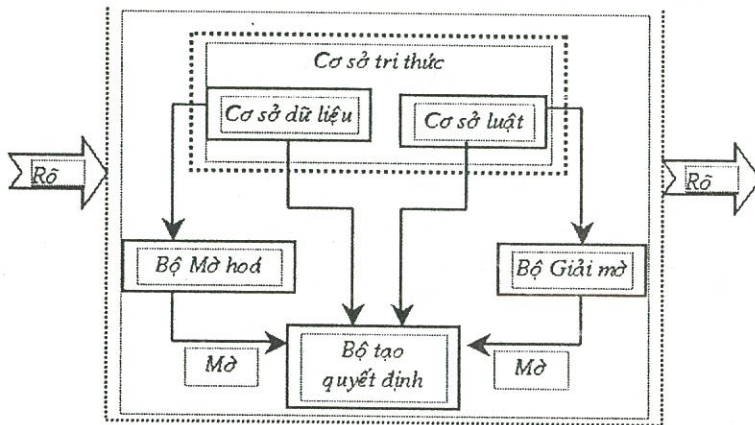
f_j là biến output.

p_j, q_j, r_j là các tham số kết luận.

3.1.2. Hệ suy diễn mờ:

Những hệ suy diễn mờ còn được gọi là những hệ thống dựa trên luật mờ, điều khiển mờ.

Một hệ suy luận mờ bao gồm năm phần cơ bản:

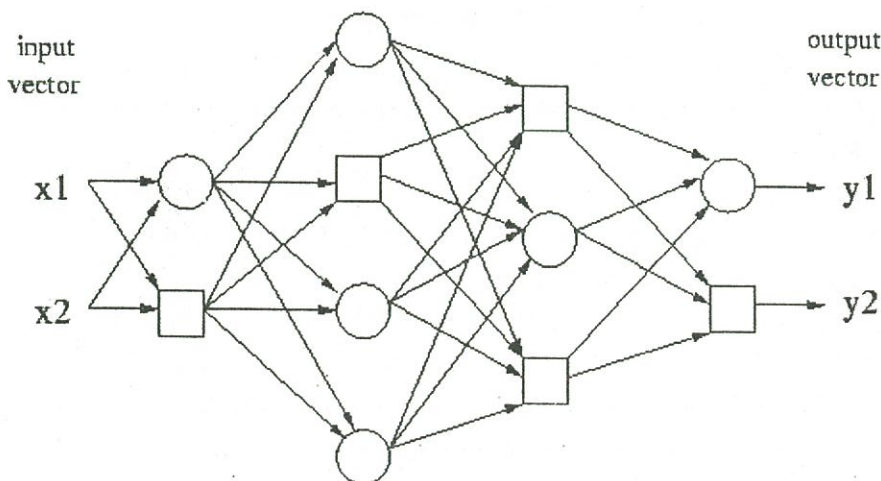


Hình 1. Hệ suy diễn mờ

- **Cơ sở luật:** gồm những luật mờ if-then.
- **Cơ sở dữ liệu:** xác định các hàm thành viên cho các tập mờ được sử dụng trong những luật mờ.
- **Bộ tạo quyết định:** thực hiện những tác vụ suy luận trên những luật mờ.
- **Bộ mờ hóa:** chuyển những giá trị đầu vào rõ sang mức độ mờ của những biến ngữ nghĩa.
- **Bộ giải mờ:** chuyển những kết quả mờ của sự suy luận sang những kết quả rõ.

3.2 Hệ suy diễn mờ dựa trên mạng thích nghi

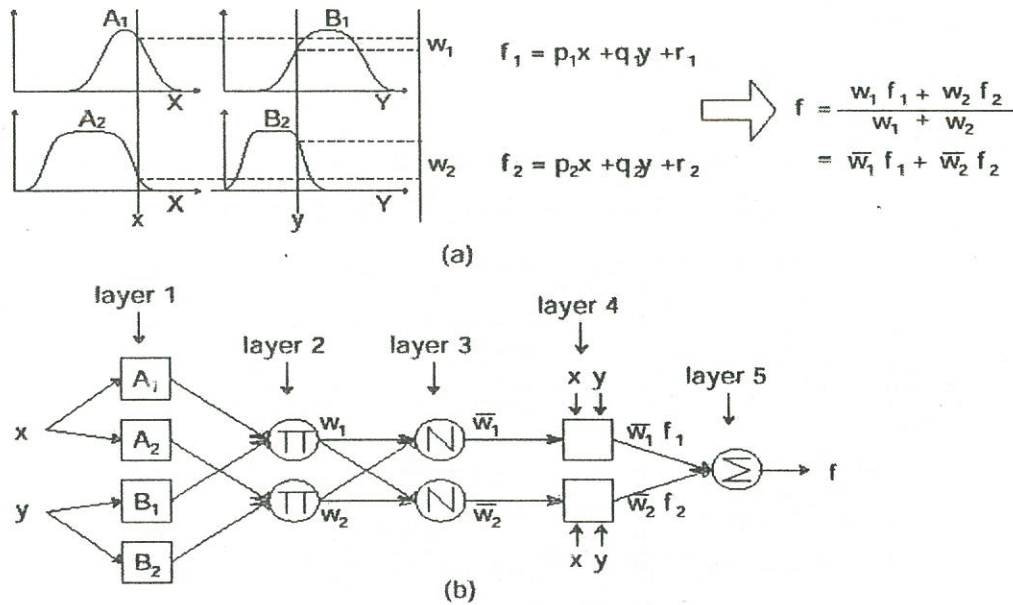
1. **Mạng thích nghi:** cấu trúc và luật học cơ bản



Hình 2. Mạng thích nghi

Mạng thích nghi là một mạng tiến nhiều lớp mà mỗi nút thực hiện một chức năng cụ thể và mỗi nút có một tập các tham số (nút này được gọi là nút chức năng). Bản chất của mỗi nút chức năng khác nhau, và việc chọn nút chức năng phụ thuộc vào hàm tổng hợp vào ra mà mạng thích nghi này thực hiện. Đường liên kết trong mạng thích nghi chỉ xác định hướng truyền của tín hiệu giữa các nút, không có trọng số trên các đường liên kết.

3.3. Kiến trúc hệ suy diễn mờ dựa trên mạng thích nghi



Hình 3. Kiến trúc hệ suy diễn mờ dựa trên mạng thích nghi

Tầng 1: mỗi nút i trong tầng này là nút vuông với hàm: $O_i^1 = \mu_{A_i}(x)$

Với x là input của node i , và A_i là biến ngôn ngữ (nhỏ, lớn,...) của hàm này. Nói cách khác O_i^1 là hàm thành viên của A_i và nó xác định mức độ x thuộc về A_i . Thông thường ta chọn $\mu_{A_i}(x)$ là hàm chuông với giá trị cực đại là 1 và giá trị cực tiểu là 0

Hay hàm Gaussian:

$$\mu_{A_i}(x) = e^{-\left(\frac{x-c_i}{a_i}\right)^2}, \tag{3}$$

Với $[a_i, b_i, c_i]$ (hay a_i, c_i) trong trường hợp hàm Gaussian) là tập tham số. Khi những giá trị của tham số thay đổi, hàm chuông sẽ thay đổi theo tạo nên hàm thành viên của nhóm ngữ

$$\mu_{A_i}(x) = \frac{\text{nghĩa } A_i}{1 + \left[\left(\frac{x-c_i}{a_i}\right)^2\right]^{b_i}} \tag{4}$$

Các tham số trong tầng này được gọi là tham số giả thiết. (thuộc phần giả thiết)

Tầng 2: Mỗi nút trong tầng này là nút hình tròn (ký hiệu Π) giá trị đưa ra sẽ là tích của các tín hiệu đầu vào.

$$W_i = \mu_{A_i}(x) \times \mu_{B_i}(y), \quad i=1,2.$$

Mỗi giá trị đầu ra của nút là giá trị đầu vào của luật. (thật ra toán tử T – norm thực hiện phép AND tổng quát có thể được sử dụng ở đây)

Tầng 3: Mỗi tầng trong lớp này là nút hình tròn (ký hiệu N_i) Nút thứ i tính tỷ số của đầu vào luật của luật thứ i với tổng giá trị đầu vào luật: $\bar{w}_i = \frac{w_i}{w_1 + w_2}, \quad i = 1, 2.$ (5)

Để tiện, ta gọi các giá trị đầu ra của lớp này là giá trị đầu vào luật được chuẩn hóa.

Tầng 4: mỗi nút i trong tầng này là nút hình vuông với

$$O_i^4(x) = \bar{w}_i f_i = \bar{w}_i(p_i x + q_i y + r_i) \tag{6}$$

Với W_i là giá trị ra của lớp 3 và $[p_i, q_i, r_i]$ là tập các tham số trong lớp này, được gọi là tham số kết luận

Tầng 5: mỗi nút trong tầng này là nút hình tròn (ký hiệu Σ), tính tổng của tất cả các giá trị vào:

$$O_1^5(x) = \text{overall output} = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad (7)$$

3.4. Luật học lai

Mặc dù ta có thể sử dụng phương pháp gradient để xác định tập tham số trong mạng thích nghi, phương pháp này nói chung chậm và thường rơi vào cực tiểu cục bộ. Ở đây đề nghị một luật học hỗn học kết hợp phương pháp gradient và phương pháp ước lượng bình phương tối thiểu (LSE) để xác định tập tham số.

Để đơn giản, giả sử rằng mạng thích nghi chỉ có một đầu ra: $\text{output} = F(\bar{I}, S)$, (8)

Với \bar{I} : tập đầu vào.

S: tập tham số của mạng.

Nếu tồn tại một hàm H sao cho hàm tích HoF là tuyến tính theo một số phần tử của S , thì những phần tử này được xác định bởi phương pháp bình phương tối thiểu.

Giả sử S có thể phân rã thành 2 tập S_1 và S_2 :

$$S = S_1 \oplus S_2 \quad (\text{kí hiệu } \oplus : \text{ là tổng trực tiếp}) \quad (9)$$

và giả sử HoF là tuyến tính với các phần tử trong tập S_2 , ta có:

$$H(\text{output}) = H(F(\bar{I}, S)), \quad (10)$$

Các hàm $H(\cdot)$ và $F(\cdot, \cdot)$ là hàm đồng nhất và hàm chức năng của hệ thống.

Từ cấu trúc ANFIS loại 3 (hình 3) ở trên, có thể thấy rằng cho trước giá trị tham số giả thiết, giá trị ra cuối cùng có thể biểu diễn là tổ hợp tuyến tính của những tham số kết luận. Chính xác hơn, giá trị ra f của hình 3.b được viết như sau:

$$f = \bar{w}_1 f_1 + \bar{w}_2 f_2 = (\bar{w}_1 x) p_1 + (\bar{w}_1 y) q_1 + (\bar{w}_1) r_1 + (\bar{w}_2 x) p_2 + (\bar{w}_2 y) q_2 + (\bar{w}_2) r_2, \quad (11)$$

Là tổ hợp tuyến tính của các tham số kết luận ($p_1, q_1, r_1, p_2, q_2, r_2$). Ta có:

S = tập tất cả các tham số.

S_1 = tập các tham số giả thiết.

S_2 = tập các tham số kết luận.

Vậy giải thuật học hỗn hợp ở trên được áp dụng vào mô hình ANFIS một cách trực tiếp. Ở bước đi tới của giải thuật học, tín hiệu được truyền tới tầng 4 và các tham số kết luận được xác định bằng LSE ở bước đi lui, tỷ lệ lỗi được lan truyền ngược và tham số giả thiết được cập nhật bằng phương pháp giảm gradient. Bảng sau tổng hợp các tác vụ trong mỗi bước:

Tổng hợp thủ tục học lai của ANFIS.

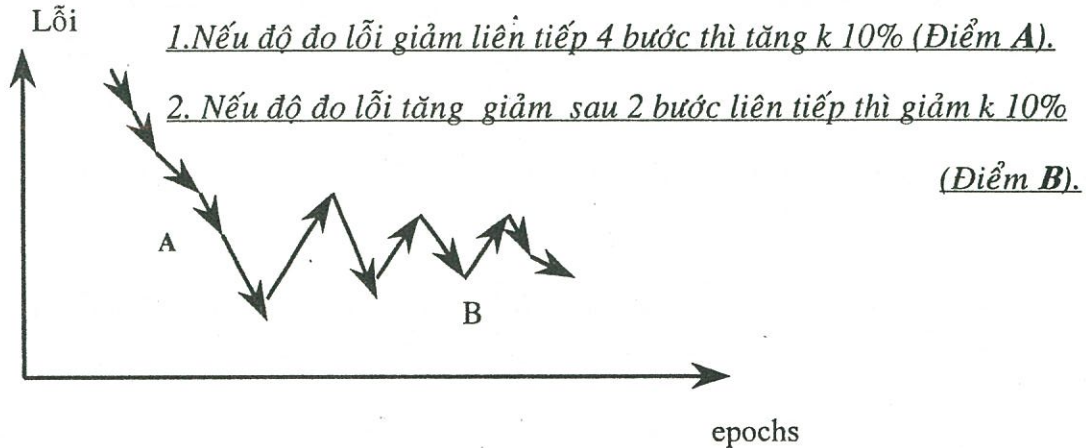
	PHA ĐI TỚI	PHA ĐI LUI
THAM SỐ GIẢ THIẾT	Cố định	Giảm gradient
THAM SỐ KẾT LUẬN	LSE	Cố định
TÍN HIỆU	Giá trị ra	Tốc độ lỗi

3.5. Xác định tham số học thích nghi

Sau khi số hàm thành viên được gán cố định cho mỗi input, các giá trị khởi tạo của các tham số giả thiết qua hàm thành viên trong miền giá trị của các biến input. Với một giá trị cho trước x của một input trong miền tính toán chúng ta luôn tìm được một biến ngôn ngữ A sao cho: $\mu_A(x) \geq \epsilon$

Theo cách này hệ suy diễn mờ có thể tạo ra sự biến đổi mịn và vừa đủ sự chồng lấp giữa các biến ngôn ngữ.

Tham số học k ảnh hưởng đến tốc độ hội tụ. Nếu k nhỏ có thể xấp xỉ sát với đường gradient nhưng hội tụ chậm và tốn thời gian. Ngược lại nếu k lớn thì sự hội tụ sẽ nhanh nhưng giải thuật sẽ dao động xung quanh điểm tối ưu. Bằng sự quan sát chúng ta cập nhật k theo hai Heuristic sau:



Hình 4. Xác định tham số học thích nghi

4. Kết quả thử nghiệm

4.1. Xây dựng hệ thống:

- **Input:** Mô hình ANFIS được sử dụng trong chương trình gồm có 5 Input, số hàm thành viên cho mỗi input là:
 - Open Price: 3 hàm thành viên
 - Close Price: 3 hàm thành viên
 - High Price : 3 hàm thành viên
 - Low Price : 3 hàm thành viên
 - Volume : 5 hàm thành viên

- **Output:** một Node output là giá mở cửa của ngày kế tiếp.

Mô hình ANFIS được sử dụng trong hệ thống thử nghiệm gồm 405 luật. Số hàm thành viên cho 4 input đầu là 3, số hàm thành viên cho input thứ 5 (Volume) là 5. Sự biến đổi giá chứng khoán là lên, xuống và phẳng lặng cho nên số hàm thành viên được chọn cho dữ liệu INPUT này là 3. trong khi đó số phiên giao dịch (Volume) có nhiều mức thay đổi nên số hàm thành viên được chọn là 5.

- Số luật được tính theo công thức :

$$Rule_n = \prod_{i=1}^{In_n} M f_i \quad (13)$$

Trong đó In là số node input, Mf_i là số hàm thành viên cho mỗi input.

- Tổng số Luật được tính là: $SumRule = 3*3*3*3*5=405$
- Tổng số node được tính: $SumNode=In+3*Rule+1$
- Lỗi của hệ thống được xác định là loại lỗi APE (Avarage Percentage Error)

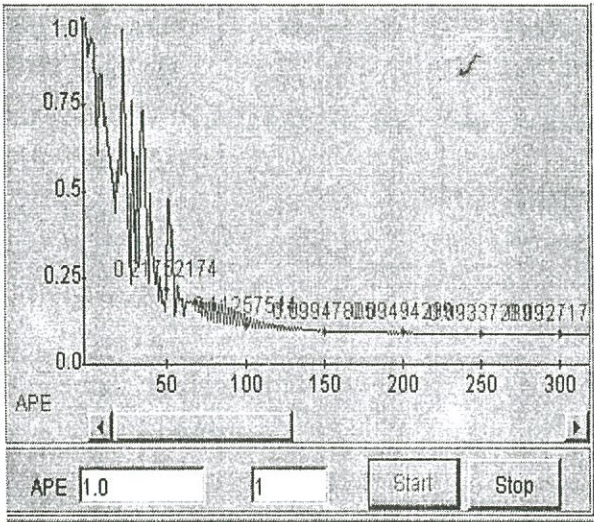
Được xây dựng theo công thức:

$$APE = \text{average percentage error} = \frac{1}{P} \sum_{i=1}^P \frac{|T(i) - O(i)|}{|T(i)|} * 100\% \quad (14)$$

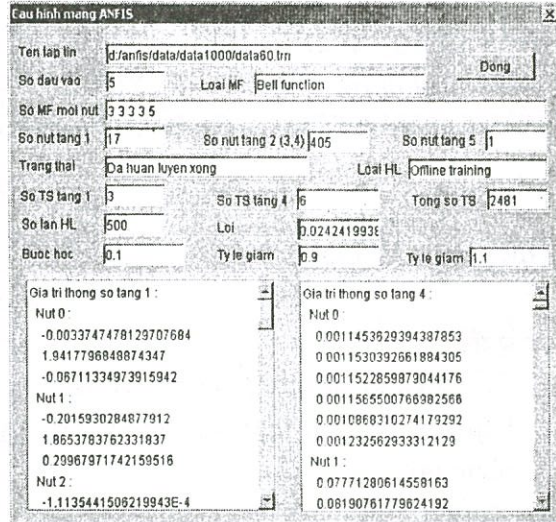
Trong đó T(i) và O(i) là output thực và output mong muốn ở tầng thứ I
P là số cặp.

- Chương trình được cài đặt bằng JBuilder 4.0 với số đầu vào 3-3-3-3-5 và tập huấn luyện 160 mẫu trong 1000 lần huấn luyện (Epochs) thời gian thực hiện trên máy PIII-500Mhz là 91 giờ.
- Chương trình cài đặt được thử nghiệm trên nhiều tập dữ liệu khác nhau, các kết quả lỗi nhỏ hơn 0.1%.
- Tập dữ liệu được lấy từ thị trường chứng khoán NYSE (New York Market Stock Exchange) tại website: Historical Quotes: <http://www.chart.yahoo.com/t>
 - Dữ liệu huấn luyện : năm 1997-1998
 - Dữ liệu thử nghiệm : năm 2000-2001

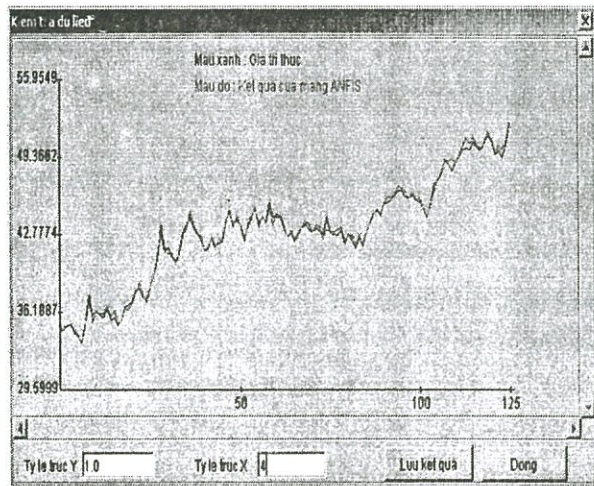
Kết quả thử nghiệm được trình bày theo các hình sau



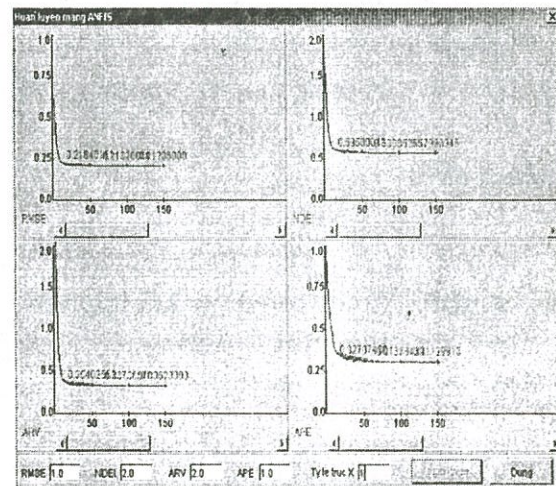
Hình 5. Lỗi huấn luyện RMSE sau 300 epochs



Hình 6. Cấu hình mạng sau khi huấn luyện



Hình 7. Kết quả thử nghiệm trên 126 mẫu



Hình 8. Màn hình đang huấn luyện

4.2. Kết quả các hàm thành viên được xây dựng sau khi huấn luyện:

Bao gồm 17 hàm thành viên biểu diễn 17 tập mờ (A_0, A_1, \dots, A_{16})

Dạng của hàm chuông với các tham số a, b, c : $Bell(a, b, c)$

$$A_0 = Bell(-0.0033747478129707684, 1.9417796848874347, -0.06711334973915942)$$

$$A_{15} = Bell(-6.992510287330894E-4, 1.991876282521868, 0.7754947984009342)$$

$$A_{16} = Bell(-4.0703969035741555E-5, 1.995542963498856, 1.0589708324585527)$$

4.3. Kết quả các luật mờ if-then thuộc dạng takagi và sugeno's được sinh ra sau khi huấn luyện:

Dạng luật mờ: IF X IS A_1 VÀ Y IS B_1 , THEN $F_1 = P_1X + Q_1Y + \gamma_1$

Với số đầu vào 3-3-3-3-5 mạng huấn luyện sinh ra gồm 405 luật

$$L1: IF X_0 IS A_0 AND X_1 IS A_3 AND X_2 IS A_6 AND X_3 IS A_9 AND X_4 IS A_{12} THEN F_0 = 0.0011453629394387853 * x_0 + 0.0011530392661884305 * x_1 + 0.0011522859879044176 * x_2 + 0.0011565500766982566 * x_3 + 0.0010868310274179292 * x_4 + 0.001232562933312129$$

$$L2: IF X_0 IS A_0 AND X_1 IS A_3 AND X_2 IS A_6 AND X_3 IS A_9 AND X_4 IS A_{13} THEN F_1 = 0.07771280614558163 * x_0 + 0.08190761779624192 * x_1 + 0.0813912960096475 * x_2 + 0.08359672117337053 * x_3 + 0.04561045489150418 * x_4 + 0.11947056624482104$$

5. Đánh giá hệ thống

5.1. Ưu điểm.

- Bằng cách sử dụng giải thuật học lai, kiến trúc ANFIS có thể tinh chỉnh đưa ra các luật mờ If then thay vì chúng được cung cấp từ các chuyên gia trong các hệ thống phức tạp.
- Nếu không có sẵn tri thức của các chuyên gia, chúng ta có thể khởi tạo các hàm thành viên bằng trực giác hợp lý và bắt đầu quá trình học để sinh ra tập các luật mờ if then nhằm xấp xỉ một tập dữ liệu mong muốn.
- Sử dụng giải thuật học lai làm cho hệ thống hội tụ nhanh nhưng vẫn đảm bảo độ chính xác mong muốn.
- Vấn đề xác định số node ẩn không yêu cầu đặt ra.

5.2. Hạn chế

- Hệ thống ANFIS chỉ có một output vì vậy chỉ được áp dụng trong các hệ thống xấp xỉ hoặc dự báo của các hàm phi tuyến chỉ có một output.
- Số hàm thành viên được kết hợp với mỗi input được gán từ thời điểm ban đầu, trong quá trình huấn luyện không điều chỉnh được. Vấn đề chọn hàm thành viên đóng vai trò quan trọng trong việc xây dựng hệ thống ANFIS.

6. Kết luận và hướng phát triển

Bài báo trình bày một ứng dụng của hệ thống lai trong quản trị tài chính, cụ thể là dự báo giá chứng khoán.

Hy vọng rằng mô hình này trong tương lai sẽ được phát triển thành ứng dụng trong lĩnh vực quản trị tài chính nói chung cũng như điều khiển tài chính nói riêng. Các hướng phát triển trong tương lai là:

- Sử dụng mô hình ANFIS phát triển các ứng dụng trong dự báo tài chính: Tỷ giá hối đoái, dự báo vỡ nợ doanh nghiệp...
- Nghiên cứu các mô hình khác trong hệ thống thông minh áp dụng vào lĩnh vực quản trị tài chính.
- Phát triển ứng dụng vào dự báo thị trường chứng khoán Việt nam.

APPLICATION ADAPTIVE-NETWORK-BASED FUZZY INFERENCE SYSTEM FOR PROCESSING KNOWLEDGE IN STOCK MARKET PREDICTIONS

Truong Hai Bang

Vietnam National University-HCMC

Center for Information Technology Development

(Received 12 November 2001, Revised 6 December 2001)

ABSTRACT: Paper deals with the application of machine learning tools in Financial Cybernetics. The use of intelligent systems in forecast financial application as e.g. stock market predictions is very actual topic. Model used *Adaptive Network Based Fuzzy Inference System*. Experiment results on New York Market Stock Exchange New York (NYSE) are promising and there are only computational limitations for this goal.

TÀI LIỆU THAM KHẢO

- [1] Roger Jang, "ANFIS: Adaptive Network Based Fuzzy Inference System", Departement of Electrical Engineering And Computer Science, University California. 1993
- [2] P.Sincak, M.Holecy, M.Ducaj "Computational Itelligence in Financial Cybernetics", Computational Itelligence in Financial Group,Laboratory of AI, Department of Cybernetics and AI, Faculty of Electrical Eengineering and Informatics Technical University of Kosice, Slovakia, 1998.
- [3] Jason E.Kutsurelis, "Forecasting Financial Markets Using neural network:an Annalysis of Methods and Accuracy", United States Navy, University California. Master Thesis .1998.
- [4] Adam Sing, "Application of Neural Network for Predicting Financial Market", University of Queensland. Master Thesis. 1997.
- [5] Chin-Teng Lin, C.S George Lee, "Neural Fuzzy system", Prentice Hall, P T R, 1996.
- [6] Joseph S. Zirilli, "Financial Prediction using Neural Networks", International Thomson computer Press. 1997.
- [7] Đức Trường Phạm and Liu Xing, "Neural networks for Identification Prediction Control", University of Wales Cardiff, UK. 1998
- [8] Tiến sỹ Lý Vinh Quang, " Chứng khoán và Phân tích Đầu tư chứng khoán", Nhà xuất bản thống kê. 1998.
- [9] TS Nguyễn Quang Thu, "Quản trị tài chính căn bản", Nhà xuất bản giáo dục –2000.
- [10] TS Nguyễn Thị Cành, Ths Nguyễn Văn Phúc, "Vận dụng các mô hình toán trong phân tích và dự báo kinh tế", Nhà xuất bản thống kê. 1999
- [11] Historical Quotes: <http://www.chart.yahoo.com/t>.
- [12] Thị trường chứng khoán việt nam : <http://www.stockmarket.vnn.vn>