

HỆ ANFIS CHO BÀI TOÁN DỰ BÁO DỮ LIỆU CHUỖI THỜI GIAN VÀ ỨNG DỤNG

Nguyễn Thị Thu Hiền¹, Lê Hữu Hà², Trần Thị Vân Anh³, Văn Thế Thành⁴, Quán Thành Thơ⁵

¹ Khoa Công nghệ thông tin, Đại học Công nghệ Tp.HCM

² Khoa Công nghệ thông tin, Học viện Kỹ thuật Quân sự

³ Khoa Công nghệ sinh học và Kỹ thuật môi trường, Đại học Công nghiệp Thực phẩm Tp.HCM

⁴ Trung tâm Công nghệ thông tin, Đại học Công nghiệp Thực phẩm Tp.HCM

⁵ Khoa Khoa học và Kỹ thuật Máy tính, Trường Đại học Bách khoa Tp.HCM

hienthuthi.nguyen@gmail.com, halh@cntp.edu.vn, anhhtt@cntp.edu.vn, vanthethanh@gmail.com

TÓM TẮT— Khai thác dữ liệu quá khứ để dự báo dữ liệu tương lai với độ chính xác cao là vấn đề thử thách. Hệ suy luận ANFIS (Adaptive Neuro-fuzzy Inference System) là một trong các kỹ thuật dùng để dự báo dữ liệu chuỗi thời gian. Vì vậy, bài báo đề xuất mô hình dự báo đa trị dựa trên hệ suy luận ANFIS. Từ đó, chúng tôi xây dựng thuật toán huấn luyện và thuật toán dự báo dựa trên mô hình đã đề xuất. Để minh chứng tính hiệu quả của mô hình đề xuất, chúng tôi xây dựng hai ứng dụng thực nghiệm: (1) Dự báo sản lượng sữa trên cơ sở dữ liệu của công ty Vinamilk; (2) Dự báo giá cổ phiếu trên cơ sở dữ liệu Cophieu68.

Từ khóa— Chuỗi thời gian, Dự báo, ANFIS.

I. GIỚI THIỆU

Theo thống kê của Tufte [16], hầu hết nguồn cung cấp dữ liệu của thế giới là ở dạng chuỗi thời gian, có khoảng 75% dữ liệu hình ảnh trên các tờ báo ở dạng chuỗi thời gian và kích thước của dữ liệu chuỗi thời gian tăng theo cấp số nhân. Bài toán dự báo dữ liệu chuỗi thời gian được ứng dụng trong nhiều lĩnh vực như dự báo giá cổ phiếu, dự báo thời tiết, dự báo sản lượng sản xuất,... [15, 19, 29]. Kết quả thu được từ mô hình dự báo có thể trở thành nguồn dữ liệu tư vấn và có thể áp dụng cho các loại dữ liệu khác nhau, từ đó đóng góp một công cụ tư vấn hữu ích cho cộng đồng.

Có nhiều mô hình dự báo đã được công bố trong thời gian gần đây như sử dụng cấu trúc cây TAN (Tree Augmented Naïve Bayes) [15], mạng neural ANN [3, 14, 21], mô hình SVM [6, 28],... Trong các mô hình dự báo, hệ ANFIS phù hợp cho các bài toán có dữ liệu đầu vào phức tạp và dự báo được đồng thời nhiều kết quả khác nhau, ví dụ như dự báo giá cổ phiếu dựa trên Hamacher T-Norm nhiều đầu vào và ANFIS [29], dự báo giá đóng (close) [25], mô hình ANFIS áp dụng trên di động [2].

Trong bài báo này, chúng tôi đề xuất mô hình dự báo đồng thời nhiều kết quả dựa trên hệ ANFIS và hệ suy luận mờ Takagi-Sugeno. Từ đó, bài báo trình bày các thuật toán huấn luyện và kiểm thử tương ứng. Dựa trên cơ sở lý thuyết đã có, chúng tôi thực nghiệm trên các dữ liệu thực tế từ công ty Vinamilk và Cophieu68 [1].

Đóng góp của bài báo gồm: (1) Đề xuất mô hình dự báo đồng thời nhiều kết quả dựa trên hệ ANFIS và phương pháp suy luận mờ Takagi-Sugeno; (2) Đưa ra thuật toán huấn luyện và thuật toán kiểm thử trên cơ sở hệ ANFIS đã đề xuất; (3) Thiết kế mô hình ứng dụng suy luận mờ; (4) Xây dựng và đánh giá thực nghiệm dựa trên các dữ liệu thực tế.

Phần còn lại của bài báo được tổ chức như sau: Phần 2. Mô tả các công trình liên quan; Phần 3. Xây dựng mô hình dự báo và các thuật toán liên quan; Phần 4. Xây dựng ứng dụng thực nghiệm và đánh giá kết quả; Phần 5. Kết luận và hướng phát triển.

II. CÁC CÔNG TRÌNH LIÊN QUAN

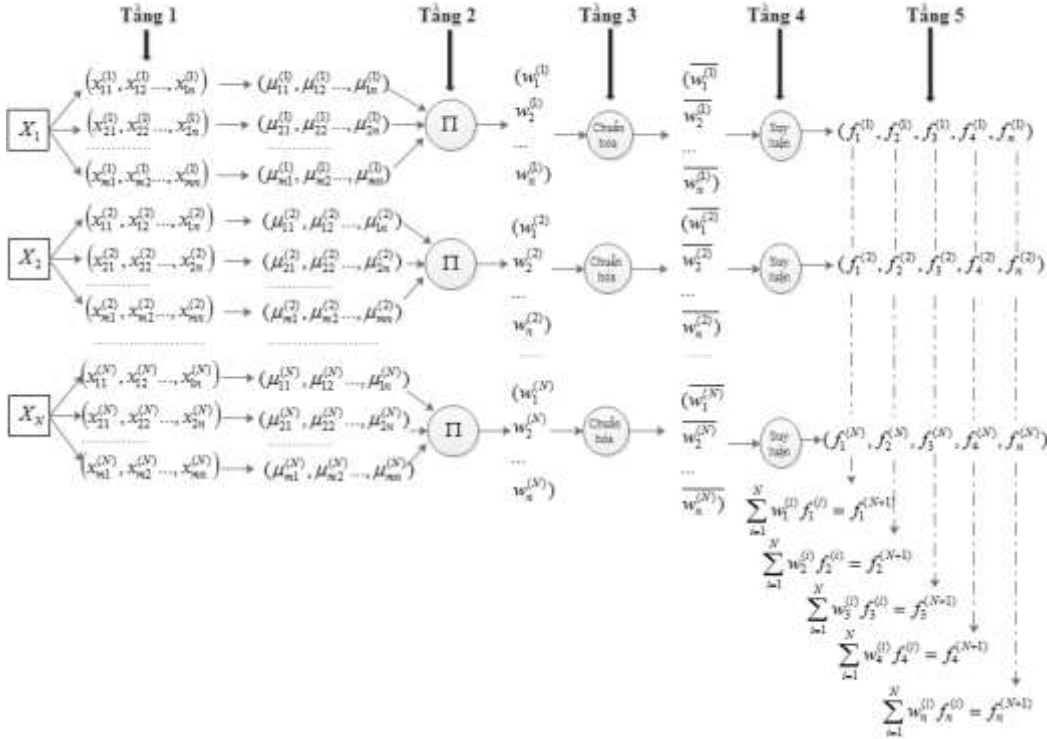
Có nhiều bài toán dự báo dữ liệu chuỗi thời gian dựa trên mô hình mạng neural như: dự báo nguồn tài nguyên nước [5], dự báo lượng điện cần tải cho thành phố [22], dự báo sự biến đổi của năng lượng mặt trời [24], dự báo mức độ ô nhiễm môi trường [18],... Ngoài ra, một số bài toán áp dụng hệ thống suy luận mờ FIS (Fuzzy Inference System) để đưa ra các dự báo như: dự báo chỉ số chứng khoán [26], dự báo tỷ giá thông qua kỹ thuật tính toán DNA [10], dự báo sự biến động của giá cả thị trường cổ phiếu [12], hệ hỗ trợ ra quyết định chiến lược kinh doanh [8],...

Mô hình dự báo neuro-fuzzy ANFIS dựa trên sự kết hợp giữa mô hình mạng neural và hệ suy luận mờ FIS cũng được đề cập trong nhiều công trình nghiên cứu như phân loại phương tiện và lưu lượng giao thông trong các ứng dụng vận chuyên [17], dự báo sự hỏng hóc của máy móc [27], xác định xác suất xảy ra dịch sốt xuất huyết ở Bangladesh [4], kỹ thuật mạng neural kết hợp với suy luận mờ để dự báo giá cổ phiếu, nhiệt độ, bán hàng, sản xuất cây trồng [23], xây dựng mô hình F-CONFIS (Fully Connected Neuro-Fuzzy Inference System) để dự báo nhiều dạng dữ liệu đầu ra [7], xây dựng kiến trúc mạng neuro-fuzzy và thuật toán suy luận mờ để dự báo dữ liệu tương lai [11], hệ thống suy luận kết hợp giữa logic mờ và mạng neural, mô hình phân lớp dựa trên mạng neuro-fuzzy [13],...

Theo tài liệu [9], A. S. George tiếp cận mô hình ANFIS để dự báo sản lượng sữa của hai nông trại. Theo tài liệu [20], M. D. Murphy tiếp cận bài toán dự báo sản lượng sữa và so sánh kết quả bằng nhiều phương pháp khác nhau. Theo tài liệu [29], F. Zhang tiếp cận hệ ANFIS để dự báo giá cổ phiếu. Các công trình trên cho thấy mô hình ANFIS phù hợp với bài toán dự báo dữ liệu chuỗi thời gian. Trong bài báo này, chúng tôi cải tiến mô hình ANFIS để thực hiện dự báo đồng thời nhiều kết quả cho hai bài toán thuộc hai lĩnh vực khác nhau nhằm minh chứng tính hiệu quả của mô hình đề xuất.

III. XÂY DỰNG MÔ HÌNH DỰ BÁO VÀ THUẬT TOÁN

A. Mô hình dự báo



Hình 1. Mô hình hệ ANFIS dự báo giá cổ phiếu và sản lượng sữa

Theo Hình 1, cho N bộ dữ liệu X_1, X_2, \dots, X_N . Mỗi bộ dữ liệu X_i là một véc-tơ mô tả giá cổ phiếu tại ngày thứ i của m mã cổ phiếu với n trạng thái tương ứng $x_{11}^{(i)}, x_{12}^{(i)}, \dots, x_{1n}^{(i)}, x_{m1}^{(i)}, x_{m2}^{(i)}, \dots, x_{mn}^{(i)}$. Vì vậy, tại tầng thứ nhất của mô hình gồm N bộ dữ liệu mô tả giá cổ phiếu theo chuỗi thời gian liên tục. Đầu ra của tầng thứ nhất là các giá trị mờ ứng với giá trị rõ trên từng bộ dữ liệu giá cổ phiếu. Các giá trị mờ ứng với chuỗi dữ liệu giá cổ phiếu trong N ngày liên tục được tích hợp và trở thành đầu vào của tầng thứ hai. Kết quả đầu ra của tầng thứ hai là một véc-tơ $W = (w_1^{(1)}, w_2^{(1)}, \dots, w_n^{(1)}, w_1^{(2)}, w_2^{(2)}, \dots, w_n^{(2)}, w_1^{(N)}, w_2^{(N)}, \dots, w_n^{(N)})$ sao cho mỗi $w_j^{(N)}$ ($j=1..n$) là giá trị tích hợp từ N thành phần mờ ứng với n trạng thái của chuỗi dữ liệu cổ phiếu. Tại tầng thứ 3, thực hiện chuẩn hóa véc-tơ W để tạo thành $\bar{W} = (\bar{w}_1^{(1)}, \bar{w}_2^{(1)}, \dots, \bar{w}_n^{(1)}, \bar{w}_1^{(2)}, \bar{w}_2^{(2)}, \dots, \bar{w}_n^{(2)}, \bar{w}_1^{(N)}, \bar{w}_2^{(N)}, \dots, \bar{w}_n^{(N)})$ với $w_j^{(N)} \in [0,1]$. Dựa trên phương pháp suy luận mờ Takagi-Sugeno, tại tầng thứ tư thực hiện quá trình suy luận để tìm ra các giá trị của ma trận dự báo, $f_n^{(N)} = r_{N0} + r_{N1} \times Y_n^{(1)} + r_{N2} \times Y_n^{(2)} + r_{N3} \times Y_n^{(3)} + \dots + r_{Nn} \times Y_n^{(N)}$ với $Y_n^{(N)} = (x_{1n}^{(N)}, x_{2n}^{(N)}, \dots, x_{mn}^{(N)})$ tương ứng để làm đầu vào cho tầng thứ năm nhằm đưa ra giá trị dự báo.

Nếu quá trình suy luận chưa hội tụ ứng với một sai số ε cho trước thì ta thực hiện quá trình huấn luyện lại các hệ số suy luận $R_{m \times (m+1)} = (r_{ij})$, $i=1, \dots, m$ và $j=0, \dots, m$ sao cho hàm lỗi đạt giá trị cực tiểu.

Tương tự như trên, chúng tôi áp dụng mô hình trên để dự báo sản lượng sữa từ m bộ dữ liệu X_1, X_2, \dots, X_m . Mỗi bộ dữ liệu X_i là một véc-tơ mô tả sản lượng sữa của một chuỗi n ngày liên tục $x_{i1}, x_{i2}, \dots, x_{in}$. Do đó, tại tầng thứ nhất của mô hình này gồm m dữ liệu mô tả sản lượng sữa theo chuỗi thời gian liên tục. Đầu ra của tầng thứ nhất là các giá trị mờ tương ứng với từng giá trị rõ trên từng bộ dữ liệu sản lượng sữa. Các giá trị mờ này được tích hợp và đồng thời là đầu vào của tầng thứ hai trong mô hình. Do đó, tại tầng thứ hai của mô hình thực hiện quá trình tích hợp của các giá trị mờ tương ứng với chuỗi dữ liệu sản lượng sữa trong n ngày liên tục. Kết quả đầu ra của tầng thứ hai là một véc-tơ $W = (w_1, w_2, \dots, w_m)$, sao cho mỗi w_i là giá trị tích hợp từ n thành phần mờ của chuỗi sản lượng sữa. Tại tầng

thứ ba của mô hình thực hiện quá trình chuẩn hóa véctor $W = (w_1, w_2, \dots, w_m)$ để tạo thành véctor $\bar{W} = (\bar{w}_1, \bar{w}_2, \dots, \bar{w}_m)$, sao cho $\bar{w}_i \in [0,1]$. Dựa trên phương pháp suy luận mờ Takagi-Sugeno, tại tầng thứ tư sẽ thực hiện quá trình suy luận để tìm ra các giá trị của ma trận dự báo $f_i = r_{i0} + r_{i1} \times X_1 + r_{i2} \times X_2 + \dots + r_{im} \times X_m$ tương ứng để từ đó làm đầu vào cho tầng thứ năm nhằm tìm ra giá trị dự báo đầu ra sau cùng.

Đóng góp trong mô hình này là cùng một thời điểm có thể dự báo một ma trận kết quả một cách đồng thời. Điều này cải thiện đáng kể về mặt tốc độ so với hệ ANFIS truyền thống. Ngoài ra, nếu sử dụng dữ liệu dưới dạng phân đoạn ngang nghĩa là dự báo cho từng khoảng dữ liệu liên tục thì mô hình này hoàn toàn đáp ứng được cho bài toán dự báo với các giá trị đơn lẻ, tức là mô hình đã đề xuất hoàn toàn thỏa mãn hệ ANFIS truyền thống. Đây là một cải tiến làm cho hệ ANFIS có thể cơ động mềm dẻo và hoàn toàn áp dụng được dữ liệu dạng ma trận các số mờ và đồng thời áp dụng được cho các giá trị đơn lẻ, vô hướng.

B. Thuật toán

1. Thuật toán huấn luyện hệ ANFIS

Đầu vào: Tập ma trận dữ liệu huấn luyện $Training = \{M_1, \dots, M_n\}$, Tập véctor giá trị thực $Real = \{r_1, \dots, r_n\}$

Đầu ra: tập ma trận hệ số suy luận $R = \{R_1, R_2, \dots, R_n\}$

Phương pháp

Bước 1: Khởi tạo

Tạo tập ma trận mờ $M-fuzzy = \{Mf_1, Mf_2, \dots, Mf_n\}$ từ tập ma trận huấn luyện $Training$;

Tạo tập véctor trọng số $W = \{W_1, W_2, \dots, W_n\}$ từ tập ma trận mờ $M-fuzzy$;

Chuẩn hóa tập véctor trọng số W ;

Khởi tạo ngẫu nhiên ma trận hệ số suy luận mờ R ;

Tính sai số $Error$ ban đầu;

Bước 2: Huấn luyện (tức là cập nhật ma trận hệ số suy luận R)

Foreach (với mỗi bộ dữ liệu huấn luyện Mf_i) **do**

While ($Error > \epsilon$ AND số lần lặp < 1000) **do**

Tạo mới ma trận hệ số suy luận R dựa trên hàm lỗi;

Tính ma trận dự báo F ;

Chuẩn hóa ma trận dự báo;

Tính kết quả dự báo;

Tính sai số $Error$;

EndWhile

Lưu trữ ma trận hệ số R_i của bộ dữ liệu huấn luyện Mf_i ;

EndForeach

Kết thúc

2. Thuật toán kiểm thử

Đầu vào: Ma trận dữ liệu kiểm thử M , tập ma trận hệ số suy luận $R = \{R_1, R_2, \dots, R_n\}$

Đầu ra: Véctor dự báo $F = \{f_1, f_2, \dots, f_m\}$,

Véctor sai số $Error = \{e_1, e_2, \dots, e_m\}$ với m là số dòng của ma trận M

Phương pháp

Bước 1: Khởi tạo

Tạo ma trận hệ số mờ T từ ma trận M ;

Tính véctor trọng số W từ ma trận T ;

Chuẩn hóa véctor trọng số W ;

Bước 2: Dự báo

Tính ma trận suy luận F ;

Chuẩn hóa ma trận suy luận F ;

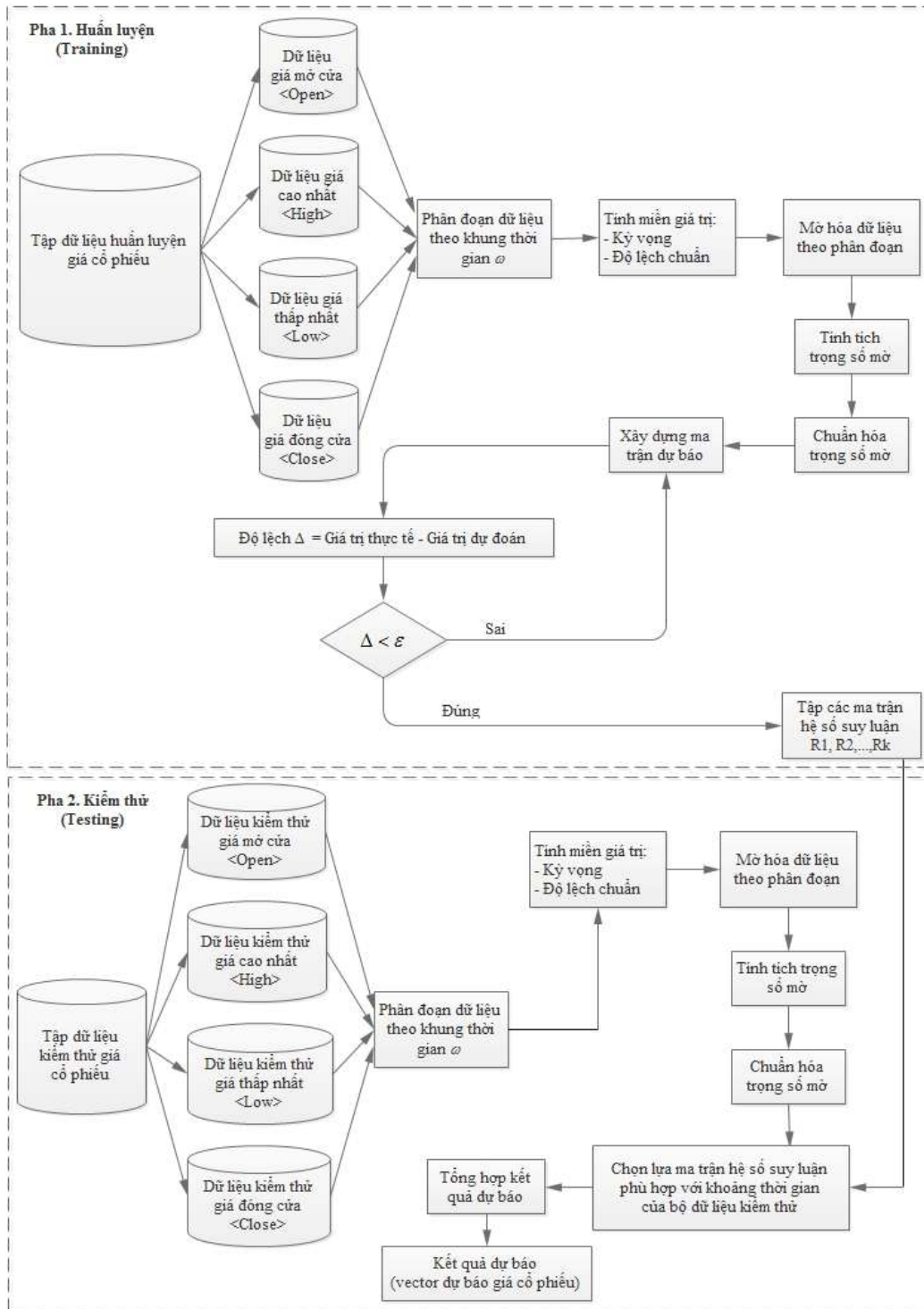
Thực hiện tính kết quả dự báo;

Tính sai số dự báo;

Kết thúc

IV. XÂY DỰNG ỨNG DỤNG THỰC NGHIỆM

A. Mô hình thực nghiệm



Hình 2. Mô hình ứng dụng hệ ANFIS

Quá trình dự báo gồm hai pha: (1) pha huấn luyện dữ liệu nhằm tìm ra tập các ma trận hệ số suy luận, chuẩn bị dữ liệu để tạo thành ứng dụng; (2) pha kiểm thử nhằm kiểm chứng lại quá trình huấn luyện hệ ANFIS, sử dụng các dữ liệu đã được chuẩn bị ở pha thứ nhất. Hai pha này thực hiện độc lập với nhau và được mô tả cụ thể như sau:

Pha 1: Pha huấn luyện. Đầu tiên, thực hiện chuẩn hóa tập dữ liệu huấn luyện ứng với khung thời gian ω bất kỳ. Tiến hành mờ hóa tập dữ liệu huấn luyện này bằng phương pháp Gaussian của tập mờ hình chuông dựa trên kỳ vọng μ và phương sai σ trên từng cột dữ liệu. Sau đó, thực hiện tính tích trọng số mờ dựa trên phép tích hợp các số mờ. Chuẩn hóa các số mờ trên đoạn $[0,1]$ nhằm đánh giá mức độ trọng số của dữ liệu. Xây dựng ma trận dự báo dựa vào kết quả có được từ phép tính tích trọng số mờ và chuẩn hóa. Trên cơ sở ma trận dự báo thực hiện việc huấn luyện

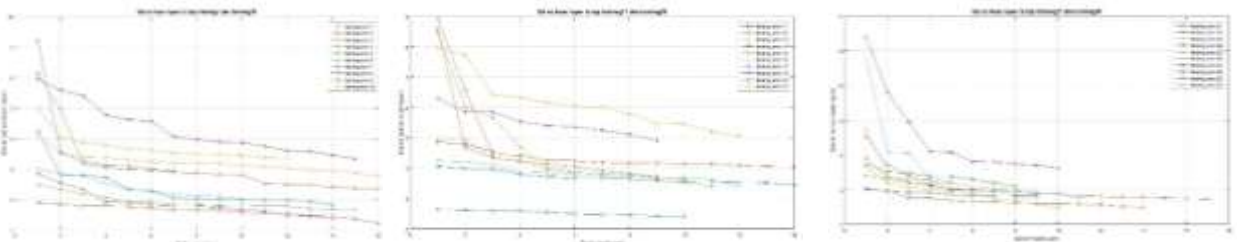
dữ liệu. Nếu độ lệch giữa giá trị thực tế và giá trị dự báo lớn hơn sai số ε cho trước thì thực hiện huấn luyện lại cho đến khi giá trị hàm lỗi nhỏ nhất thì dừng lại. Kết quả của quá trình huấn luyện là tập các ma trận hệ số suy luận $R = \{R_1, \dots, R_k\}$ phù hợp với khoảng thời gian của dữ liệu dự báo.

Pha 2: Pha kiểm thử. Thực hiện chuẩn hóa bộ dữ liệu kiểm thử với khung thời gian ω đã có trong pha huấn luyện. Tiến hành mờ hóa tập dữ liệu kiểm thử này bằng hàm mờ Gaussian hình chuông. Tính tích trọng số mờ dựa trên phép tích hợp các số mờ và chuẩn hóa các số mờ trên đoạn $[0,1]$. Với ma trận hệ số suy luận đã có trong pha huấn luyện thực hiện quá trình huấn luyện dựa trên hệ ANFIS và phương pháp suy luận Takagi-Sugeno sẽ thu được một vectơ dự báo. Từ đó, có thể dự báo liên tục nhiều ngày tương lai.

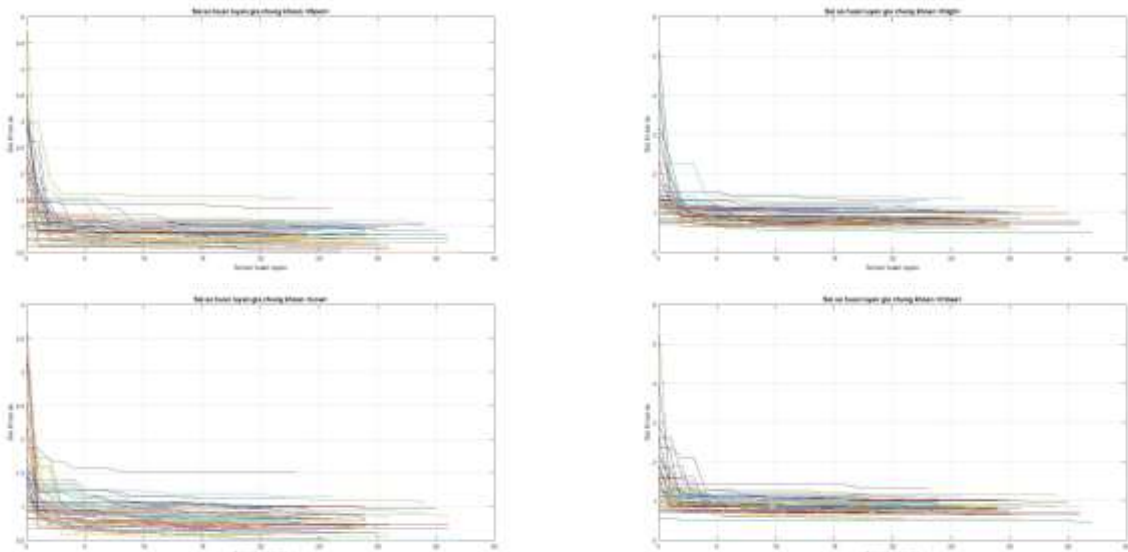
B. Kết quả thực nghiệm

Ứng dụng thực nghiệm gồm: pha huấn luyện và pha kiểm thử. Pha huấn luyện dữ liệu được thực thi trên Matlab phiên bản 2013 trên hệ điều hành Windows Server 2008 R2 Enterprise, bộ xử lý Intel(R) Xeon(R) CPU X3440 @ 2.53GHz – Ram 8GB. Pha kiểm thử dữ liệu được thực thi trên Matlab phiên bản 2008 trên hệ điều hành Windows 8.1 Professional, bộ xử lý Core i3 – Ram 4GB. Quá trình thực nghiệm được đo đạc trên đơn vị thời là giây theo chuẩn của Matlab. Toàn bộ các hàm hỗ trợ của thuật toán đều được cài đặt cụ thể trên Matlab ngoại trừ các hàm cơ bản. Các số liệu được đo đạc và biểu diễn đồ thị được thực hiện trực tiếp trên Matlab.

Dựa trên hàm lỗi đã mô tả như trên, bài báo đánh giá sai số trung bình huấn luyện được mô tả như sau:



Hình 3. Giá trị trung bình của độ lệch ứng với bộ dữ liệu huấn luyện sản lượng sữa

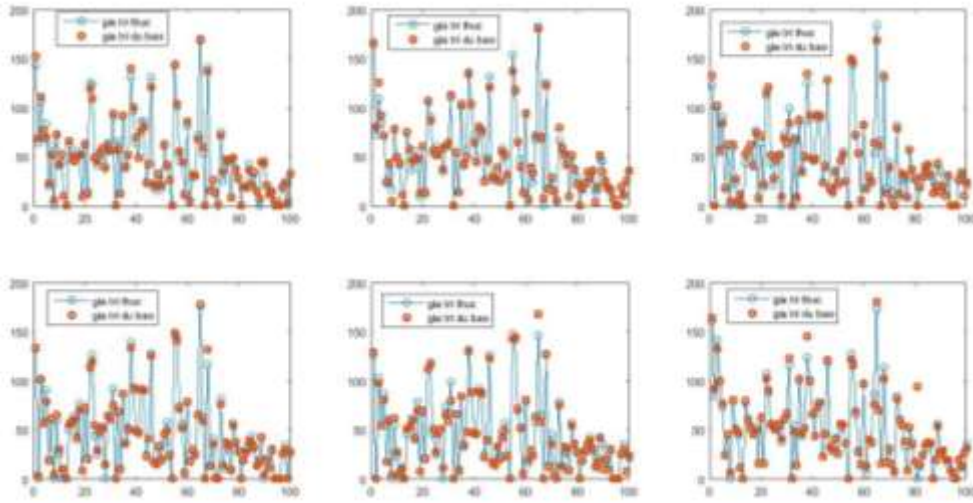


Hình 4. Giá trị trung bình của độ lệch ứng với bộ dữ liệu huấn luyện giá cổ phiếu

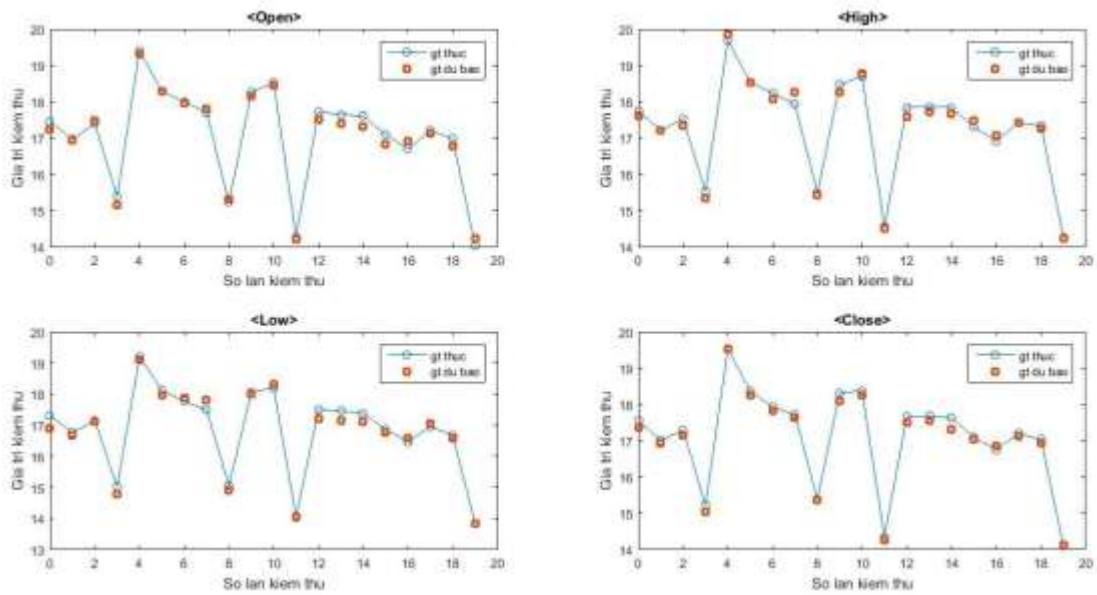
Tại **Hình 3** mô tả sai số huấn luyện của mạng neuro-fuzzy cho dự báo sản lượng sữa của từng bộ dữ liệu mẫu được lấy từ các trạm cung cấp sữa của công ty Vinamilk. Theo như thực nghiệm cho thấy giá trị hàm lỗi giảm tương đối nhanh tại các thời điểm ban đầu sau đó tốc độ hội tụ giảm dần. Tương tự tại **Hình 4** mô tả sai số huấn luyện của mạng neuro-fuzzy ứng với dữ liệu *Cophieu68*. Theo đồ thị cho thấy tốc độ hội tụ của mạng neuro-fuzzy tương đối nhanh. Từ đó cho thấy mô hình đã đề xuất phù hợp với bài toán dự báo dữ liệu dạng chuỗi thời gian.

Sau khi thực hiện huấn luyện mạng neuro-fuzzy, chúng tôi tiến hành quá trình kiểm thử nhằm dự báo kết quả. Việc kiểm thử này được thực hiện trên bài toán dự báo sản lượng sữa và dự báo giá cổ phiếu nhằm minh chứng tính tổng quát và tính đúng đắn của mô hình đã đề xuất. Tại **Hình 5** mô tả giá trị thực và giá trị dự báo về sản lượng sữa và tại **Hình 6** mô tả giá trị thực và giá trị dự báo của giá cổ phiếu với bốn trạng thái: Giá mở <Open>, Giá cao nhất <High>, Giá thấp nhất <Low> và giá đóng <Close>.

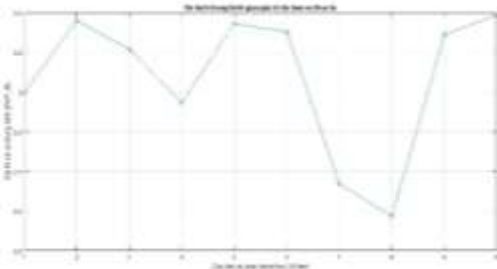
Kết quả của pha kiểm thử được mô tả theo **Hình 5** và **Hình 6** như sau:



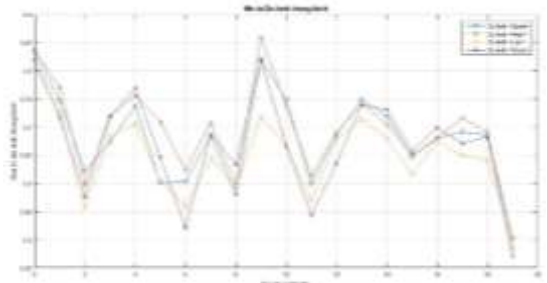
Hình 5. Mô tả giá trị dự báo và giá trị thực từ bộ kiểm thử về sản lượng sữa



Hình 6. Mô tả giá trị dự báo và giá trị thực từ bộ kiểm thử với 20 mã cổ phiếu

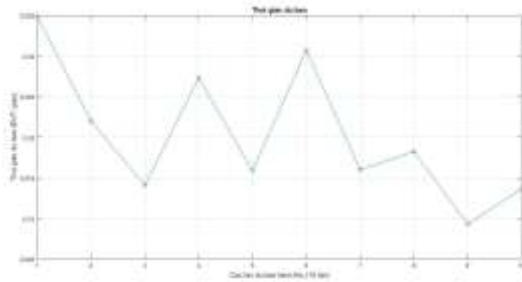


Hình 7. Độ lệch trung bình giữa giá trị dự báo và giá trị thực về sản lượng sữa

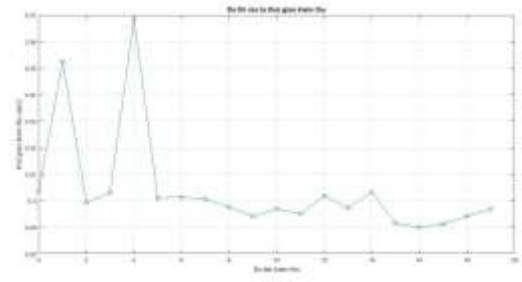


Hình 8. Độ lệch trung bình giữa giá trị dự báo và giá trị thực về giá cổ phiếu

Thời gian dự báo được đo đạc trực tiếp bằng phần mềm Matlab nhằm đánh giá chi phí về tốc độ dự báo nhằm phục vụ cho quá trình định hướng kinh doanh. Đồ thị mô tả thời gian dự báo như sau:



Hình 9. Đồ thị mô tả thời gian dự báo sản lượng sữa



Hình 10. Đồ thị mô tả thời gian dự báo giá cổ phiếu

C. So sánh kết quả thực nghiệm với các phương pháp khác

Bảng 1. So sánh sai số giữa phương pháp đề xuất và các phương pháp khác trong dự báo sản lượng sữa

Phương pháp	MSE			RMSE			MAE			MAPE		
	1	2	TB	1	2	TB	1	2	TB	1	2	TB
AR	15.6612	93.5912	54.6262	3.9574	9.6743	6.81585	2.9258	6.81585	3.9663	5.3127	5.7072	5.50995
ARMA	15.6616	93.7654	54.7135	3.9575	9.6833	6.8204	2.9258	5.0073	3.96655	5.3126	5.7072	5.5099
ANFIS	12.3017	79.9558	46.12875	3.5074	8.9418	6.2246	2.6506	4.0399	3.34525	4.7446	4.6049	4.67475
Phương pháp đề xuất	17.9444; 25.2930; 24.1698; 20.4683; 31.6572; 103.775; 15.9993; 14.9449; 27.7341; 27.5319		30.95179	4.2361; 5.0292; 4.9162; 4.5242; 5.6264; 10.187; 3.9999; 3.8658; 5.2663; 5.2471		5.28982	2.9977; 3.3611; 3.2162; 2.9462; 3.3458; 3.3050; 2.5376; 2.3742; 3.2905; 3.3907		3.0765	6.4892; 6.8751; 7.1831; 6.6168; 7.5374; 6.7119; 5.5827; 5.3558; 8.0169; 7.0238		6.73927

Bảng 2. So sánh sai số giữa phương pháp đề xuất và các phương pháp khác trong dự báo giá cổ phiếu

Loại sai số Phương pháp	MSE	MAE	MAPE
T-Norm dựa trên ANFIS	1.8069	1.3372	0.2534
ANFIS truyền thống	2.3729	1.7263	0.4012
ANFIS đề xuất	0.0296; 0.0233; 0.3079; 0.1087	0.1474; 0.1282; 0.5268; 0.3087	0.0088; 0.0073; 0.0306; 0.0178

V. LỜI CẢM ƠN

Nhóm tác giả xin chân thành cảm ơn Khoa Khoa học và Kỹ thuật Máy tính, trường Đại học Bách khoa Tp.HCM và Trung tâm Công nghệ Thông tin, trường Đại học Công nghiệp Thực phẩm Tp.HCM là nơi bảo trợ để thực hiện nghiên cứu này.

VI. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Bài báo đã giải quyết được vấn đề dự báo đồng thời nhiều kết quả, giúp cải thiện đáng kể về tốc độ so với các phương pháp khác như: AR, ARMA, ANFIS, NARX, SANN, MLR, T-Norm dựa trên ANFIS. Điều này đã được minh chứng trong phần kết quả thực nghiệm. Hướng phát triển của bài báo là đề xuất hệ ANFIS có khả năng tự nhận biết các trường hợp dị biệt và có khả năng chịu lỗi cao để kết quả dự báo không bị ảnh hưởng bởi các giá trị bất thường.

TÀI LIỆU THAM KHẢO

[1] <http://www.cophieu68.vn/>. 2016.
 [2] A. Al-Hmouz, et al. (2012). 'Modeling and Simulation of an Adaptive Neuro-Fuzzy Inference System (ANFIS) for Mobile Learning'. *IEEE transactions on learning technologies*. 5(3), 226-237.
 [3] A. Arango, J.D. Velásquez (2014). 'Forecasting the Colombian Exchange Market Index (IGBC) using Neural Networks'. *IEEE latin america transactions*. 12(4), 718-724.
 [4] R. Lee (2016). *Detection of Dengue Epidemic in Dhaka, Bangladesh by a Neuro Fuzzy Approach*, M. Arifuzzaman, et al., In: *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*. Springer Switzerland. pp. 165-174.
 [5] S. Campisi-Pinto, J. Adamowski, G. Oron (2013). 'Forecasting Urban Water Demand Via Wavelet-Denoising and Neural Network Models. Case Study: City of Syracuse, Italy'. *Water Resources Management*. 26(12), 3539-3558.

- [6] R. Cao, X. Liang, Z. Ni (2012). *Stock Price Forecasting with Support Vector Machines Based on Web Financial Information Sentiment Analysis*. Paper presented at 8th International Conference, ADMA 2012 Nanjing, December 15-18, 2012, China.
- [7] C. L. P. Chen, Chi-Hsu Wang (2014). 'A New Learning Algorithm for a Fully Connected Neuro-Fuzzy Inference System'. *IEEE transactions on neural networks and learning systems*. 25(10), 1741-1757.
- [8] W. Pedrycz, S.-M. Chen (2013). *A Best-Match Forecasting Model for High-Order Fuzzy Time Series*, Y.-C. Cheng, S.-T. Li, In: *Time Series Analysis, Modeling and Applications*. Springer Berlin pp. 331-345.
- [9] A.S. George, P. G. Maria, Z. D. Constantinou (2012). 'Milk Production Forecasting by a Neuro-Fuzzy Model'. *Research Topics in Agricultural and Applied Economics*. 3, 3-11.
- [10] C. Mao, T. Yokomori (2006). *Fuzzy Forecasting with DNA Computing*, D. J. -F. Jeng, et al., In: *DNA Computing*. Springer Berlin. pp. 324-336.
- [11] L. Jing (2013). 'A neural fuzzy inference system'. *Journal of electronics*. 30(4), 401-410.
- [12] K. Deep, et al. (2012). *A Computational Method of Forecasting Based on Intuitionistic Fuzzy Sets and Fuzzy Time Series*, B.P. Joshi, S. Kumar, In: *Advances in Intelligent and Soft Computing*. Springer India. pp. 993-1000.
- [13] S. Kara, S. Das, P.K. Ghosh (2014). 'Applications of neuro fuzzy systems: A brief review and future outline'. *Applied Soft Computing*. 15, 243-259.
- [14] P. Koprinkova, H. V. Mladenov, N. K. Kasabov. (2015). *Artificial Neural Networks*. Springer. New York.
- [15] Y. Liao, et al. (2013). 'Stock Price Forecast Using Tree Augmented Naïve (TAN) Bayes'. *Advances in Intelligent Systems and Computing*. 212, 1013-1019.
- [16] O. Maimon, L. Rokach. (2005). *Data mining and knowledge discovery handbook*. Springer. New York.
- [17] K. N. Das, et al. (2015). *Vehicle Classification Using Adaptive Neuro-Fuzzy Inference System (ANFIS)*, A. K. Maurya, D. K. Patel, In: *Proceedings of Fourth International Conference on Soft Computing for Problem Solving Book Subt*. Springer India. pp. 137-152.
- [18] C. G. Helmis, P. T. Nastos (2013). *24-Hours Ahead Forecasting of PM10 Concentrations Using Artificial Neural Networks in the Greater Athens Area, Greece*, K. Moustiris, et al., In: *Advances in Meteorology, Climatology and Atmospheric Physics*. Springer Berlin pp. 1121-1126.
- [19] M. D. Murphy, et al. (2014). 'Comparison of modeling techniques for milk-production forecasting'. *American Dairy Science Association*. 97, 3352-3363.
- [20] M. D. Murphy, et al. (2014). 'Comparison of modeling techniques for milk-production forecasting'. *J. Dairy Sci*. 97, 3352-3363.
- [21] T. Ravichandra, C. Thingom (2016). 'Stock Price Forecasting Using ANN Method'. *Advances in Intelligent Systems and Computing*. 435, 599-605.
- [22] I. E. Shepelev, et al. (2015). 'Comparing Ambient Temperature Account Methods in Neural Network Based City ShortTerm Load Forecasting'. *Optical Memory and Neural Networks*. 24(3), 220-229.
- [23] P. Singh (2016). *Two-Factors High-Order Neuro-Fuzzy Forecasting Model*, P. Singh, In: *Applications of Soft Computing in Time Series Forecasting*. Springer Switzerland. pp. 83-97.
- [24] V. Vijay, et al. (2015). *Ground-Based Measurement for Solar Power Variability Forecasting Modeling Using Generalized Neural Network*, V. P. Singh, et al., In: *Systems Thinking Approach for Social Problems*. Springer India. pp. 49-61.
- [25] I. Svalina, et al. (2013). 'An adaptive network-based fuzzy inference system (ANFIS) for the forecasting: The case of close price indices'. *Expert Systems with Applications*. 40(15), 6055-6063.
- [26] V. H. Nguyen, A. C. Le, V. N. Huynh (2015). *A New Approach to Multi-variable Fuzzy Forecasting Using Picture Fuzzy Clustering and Picture Fuzzy Rule Interpolation Method*, P. H. Thong, L. H. Son, In: *Knowledge and Systems Engineering*. Springer: New York. pp. 679-690.
- [27] B. V. Babu, et al. (2014). *Fault Diagnosis of Ball Bearings Using Support Vector Machine and Adaptive Neuro Fuzzy Classifier*, R. Tiwari, P. K. Kankar, V. K. Gupta, In: *Advances in Intelligent Systems and Computing*. Springer India. pp. 1477-1482.
- [28] L. Y. Wei (2016). 'A hybrid ANFIS model based on empirical mode decomposition for stock time series forecasting'. *Applied Soft Computing*. 42, 368-376.
- [29] F. Zhang, Z. Liao (2015). *Stock Price Forecasting Based on Multi-Input Hamacher T-Norm and ANFIS*. Paper presented at The Ninth International Conference on Management Science and Engineering Management, July 21 to 23, 2015, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany.

THE ANFIS FOR TIME-SERIES FORECASTING AND APPLICATION

Nguyen Thi Thu Hien, Le Huu Ha, Tran Thi Van Anh, Van The Thanh, Quan Thanh Tho

ABSTRACT— Prediction of data trend in future is one of major data mining tasks, which poses much challenges in order to obtain results of high accuracy. In particular, for the domain of time-series data, the inference system ANFIS (Adaptive Neuro-fuzzy Inference System) promises a powerful technique. Then, this paper proposes a multi-value prediction model based on ANFIS, which is very useful for time-series analysis. On the base of proposed model, we present a training algorithm and a testing algorithm. In order to show the effectiveness of proposed method, we developed two applications including (i) prediction of the milk quantities of Vinamilk products; and (ii) prediction of the stock prices, using data from Cophieu68 company.