

MÔ HÌNH CHUỖI THỜI GIAN MỜ BẬC CAO HAI NHÂN TỐ KẾT HỢP VỚI TỐI ƯU BẦY ĐÀN CHO DỰ BÁO NHIỆT ĐỘ VÀ THỊ TRƯỜNG CHỨNG KHOÁN

Nghiêm Văn Tính^{1*}, Nguyễn Công Điều²

¹ Trường Đại học Kỹ thuật Công nghiệp - Đại học Thái Nguyên

² Trường Đại học Thăng Long

nghiemvantinh@tnut.edu.vn, ncdieu@yahoo.com

TÓM TẮT: Trong thực tế, một dãy quan sát có thể phụ thuộc và bị ảnh hưởng vào nhiều yếu tố khác nhau. Do đó, để dự báo chính xác hơn cần đưa thêm những yếu tố khác có tương quan tiềm năng vào mô hình. Nghiên cứu này đề xuất một mô hình dự báo mới dựa trên chuỗi thời gian mờ bậc cao hai nhân tố kết hợp với tối ưu bầy đàn (PSO) để dự báo nhiệt độ và thị trường chứng khoán. Khác với một số mô hình trước đây, trong mô hình đề xuất, nhóm quan hệ mờ phụ thuộc thời gian bậc cao được xây dựng để sử dụng trong quá trình dự báo. Còn PSO vẫn được áp dụng để tìm độ dài khoảng tối ưu trong tập nền của từng nhân tố. Hai tập dữ liệu kinh điển về thời tiết và thị trường chứng khoán ở Đài Loan với các nhân tố tương ứng là 'nhiệt độ', 'độ bao phủ của mây' và 'chỉ số TAIFEX', 'chỉ số TAIEX' được lựa chọn để minh chứng cho hiệu quả của mô hình đề xuất và so sánh với các mô hình hiện có. Từ kết quả cho thấy mô hình đề xuất đưa ra độ chính xác dự báo tốt hơn so với các mô hình dự báo trước đây dựa trên chuỗi thời gian mờ bậc cao hai nhân tố.

Từ khóa: Chuỗi thời gian mờ, Nhóm quan hệ mờ phụ thuộc thời gian hai nhân tố, Tối ưu bầy đàn, Nhiệt độ, TAIFEX.

I. GIỚI THIỆU

Chuỗi thời gian mờ (FTS) do Song và Chissom đề xuất trong [1], [2] được xem như một công cụ hữu hiệu cho phân tích chuỗi thời gian. Kể từ đó, nó là chủ đề của nhiều nghiên cứu trong lĩnh vực dự báo, đặc biệt là khi dữ liệu thu thập là không chính xác, mơ hồ và không có xu hướng xác định. Hai mô hình FTS ban đầu của Song và Chissom đưa ra khá phức tạp và mất nhiều thời gian tính toán khi giải quyết với ma trận quan hệ mờ lớn. Chúng được phát triển bởi một mô hình hiệu quả hơn trong Chen [3] bằng cách sử dụng các phép toán đơn giản thay vì các phép toán max-min phức tạp để dự báo tuyến sinh đại học tại trường đại học Alabama. Mô hình này trở thành cấu trúc FTS phổ biến và được nhiều nhà nghiên cứu cải tiến, áp dụng trong nhiều bài toán dự báo bao gồm: tuyển sinh đại học [3-16], sản xuất lúa gạo [17], chỉ số vận chuyển [18], ô nhiễm [19], giao dịch chứng khoán [20-25], nhiệt độ hàng ngày [21-23], vv. Song song với việc áp dụng các bài toán trên, nhiều tác giả đã đưa ra các mô hình dự báo hiệu quả hơn mô hình [3] dựa trên các cải tiến về độ dài khoảng chia [26-28], mờ hóa dữ liệu [29-31], quan hệ logic mờ [32-35] và kỹ thuật giải mờ [10, 16, 36]. Cụ thể với mục đích nâng cao độ chính xác dự báo, gần đây hơn các nhà nghiên cứu đã đề xuất các kỹ thuật khác nhau để giải quyết các bài toán phức tạp trong mô hình chuỗi thời gian mờ. Ví dụ: Tối ưu bầy đàn (Particle swarm optimization-PSO) được sử dụng cho việc điều chỉnh và tìm độ dài khoảng chia thích hợp trong các bài báo [8, 10-12, 24]. Bằng cách tương tự để tìm độ dài khoảng tối ưu, một số công trình khác dùng giải thuật di truyền GA [6, 7] kết hợp với chuỗi thời gian mờ cho dự báo tuyển sinh và giải thuật tìm kiếm hòa âm [37] để hiệu chỉnh độ dài khoảng chia trong giai đoạn mờ hóa. Kỹ thuật tìm kiếm Tuba và hệ thống suy diễn mờ [38] cũng được áp dụng để tìm độ dài khoảng chia từ tập nền trong mô hình chuỗi thời gian mờ N nhân tố cho dự báo sản xuất công nghiệp ô tô. Nhóm tác giả khác lại áp dụng các thuật toán phân cụm kết hợp với chuỗi thời gian mờ trong quá trình xây dựng mô hình dự báo như thuật toán phân cụm k-mean [15, 16, 26], thuật toán phân cụm tự động [5, 28]. Một số kỹ thuật khác để xác định khoảng và độ dài khoảng chia tối ưu cũng được trình bày trong các công trình [13, 14, 39, 40]. Một hướng phát triển khác có thể thấy trên thực tế, chuỗi dữ liệu chính còn phụ thuộc và bị ảnh hưởng bởi những số liệu khác. Do đó, việc kết hợp nhiều chuỗi số liệu (nhân tố) để dự báo cho nhân tố chính trong mô hình có thể cải thiện độ chính xác dự báo một cách đáng kể. Điều này được chứng minh trong các nghiên cứu [21, 24, 38, 41, 42, 43]. Mặt khác, một số nghiên cứu trong [4, 5, 21-24] đã chứng minh hiệu quả của mô hình dự báo được cải thiện bằng việc thiết lập quan hệ mờ bậc cao,... Từ các nghiên cứu trên cho thấy độ dài của khoảng chia, bậc của quan hệ mờ và yếu tố quan sát trong chuỗi thời gian mờ có ảnh hưởng rất lớn đến độ chính xác dự báo của mô hình. Vì vậy, nghiên cứu này đưa ra mô hình chuỗi thời gian mờ bậc cao hai nhân tố kết hợp với tối ưu bầy đàn cho dự báo nhiệt độ hàng ngày và thị trường chứng khoán. Trong mô hình đề xuất sử dụng khái niệm nhóm quan hệ mờ phụ thuộc thời gian [11] để thiết lập nhóm quan hệ mờ bậc cao hai nhân tố và áp dụng kỹ thuật tối ưu bầy đàn trong việc tìm khoảng chia tối ưu từ tập nền của mỗi nhân tố nhằm tăng độ chính xác dự báo của mô hình. Thử nghiệm trên hai bộ dữ liệu về nhiệt độ trung bình hàng ngày và thị trường chứng khoán tại Đài Bắc, Đài Loan cho thấy mô hình đề xuất đạt độ chính xác cao hơn so với các mô hình trước đây. Phần còn lại của bài báo được tổ chức như sau:

Sau phần giới thiệu là một số khái niệm liên quan đến chuỗi thời gian mờ trong phần 2. Phần 3 giới thiệu từng bước của mô hình dự báo kết hợp với PSO. Phần 4 đánh giá hiệu quả dự báo của mô hình đề xuất và so sánh với các mô hình dự báo trước đây. Cuối cùng là một số kết luận được đưa ra trong phần 5.

II. MỘT SỐ ĐỊNH NGHĨA VÀ CÁC THUẬT TOÁN

Trong phần này tóm tắt một số khái niệm cơ bản về chuỗi thời gian mờ [1-3] và thuật toán tối ưu bầy đàn [44].

2.1. Định nghĩa về chuỗi thời gian mờ (FTS-fuzzy time series)

Chuỗi thời gian mờ khác với chuỗi thời gian truyền thống là các quan sát không phải là giá trị số mà được biểu diễn bởi giá trị ngôn ngữ. Với khả năng làm việc với một tập dữ liệu quan sát nhỏ khiến cho việc dự báo chuỗi thời gian mờ ngày càng được quan tâm hơn. Một số định nghĩa cơ bản về chuỗi thời gian mờ một nhân tố và hai nhân tố được đưa ra như sau:

Định nghĩa 1: Chuỗi thời gian mờ [1]

Cho $Y(t) (t = \dots, 0, 1, 2, \dots)$ là một tập con của tập số thực và cũng là tập nền trên đó xác định các tập mờ $f_i(t)$. $F(t)$ là tập chứa các tập $f_i(t)$ ($i = 1, 2, \dots$). Khi đó ta gọi $F(t)$ là chuỗi thời gian mờ xác định trên tập nền $Y(t)$.

Định nghĩa 2: Chuỗi thời gian mờ bậc m ($m \geq 2$) một nhân tố [4]

$F(t)$ là một chuỗi thời gian mờ. Nếu $F(t)$ được suy ra bởi $F(t-1), F(t-2), \dots, F(t-m)$, thì quan hệ giữa chúng được biểu diễn bởi $F(t-1), F(t-2), \dots, F(t-m) \rightarrow F(t)$ và nó được gọi là chuỗi thời gian mờ bậc m một nhân tố.

Định nghĩa 3: Chuỗi thời gian mờ bậc m hai nhân nhân tố [21]

Giả sử $F_A(t)$ và $F_B(t)$ là hai chuỗi thời gian mờ. Nếu $F_A(t)$ được suy ra đồng thời bởi $(F_A(t-1), F_B(t-1)), (F_A(t-2), F_B(t-2)), \dots, (F_A(t-m), F_B(t-m))$ thì quan hệ giữa chúng được biểu diễn theo cách sau $(F_A(t-m), F_B(t-m)), \dots, (F_A(t-2), F_B(t-2)), (F_A(t-1), F_B(t-1)) \rightarrow F_A(t)$ và nó được gọi là chuỗi thời gian mờ bậc m hai nhân nhân tố. Trong đó, $F_A(t)$ được gọi là nhân tố chính và $F_B(t)$ gọi là nhân tố thứ hai của chuỗi thời gian mờ và $(F_A(t-m), F_B(t-m)), \dots, (F_A(t-2), F_B(t-2)), (F_A(t-1), F_B(t-1))$ gọi là trạng thái hiện tại của quan hệ, $F_A(t)$ gọi là trạng thái tương lai.

Định nghĩa 4: Nhóm quan hệ mờ bậc 1 một nhân tố (FLRGs) [3]

Một nhóm quan hệ mờ có thể được xem như một tập các quan hệ mờ có cùng về trái với nhau. Giả sử có các quan hệ logic mờ bậc một có cùng các tập mờ bên về trái như sau:

$$A_i \rightarrow A_{k1}; A_i \rightarrow A_{k2}; \dots; A_i \rightarrow A_{km}.$$

Theo Chen [3], các quan hệ này được gom thành một nhóm như sau: $A_i \rightarrow A_{k1}, A_{k2}, \dots, A_{km}$. Các quan hệ giống nhau (lặp lại) chỉ được tính duy nhất một lần khi tham gia vào nhóm quan hệ mờ.

Định nghĩa 5: Nhóm quan hệ mờ phụ thuộc thời gian bậc cao hai nhân tố

Dựa trên khái niệm nhóm quan hệ mờ phụ thuộc thời gian đề xuất trong tài liệu [11] và khái niệm về chuỗi mờ bậc cao hai nhân tố. Chúng tôi đưa ra khái niệm về quan hệ mờ phụ thuộc thời gian bậc m hai nhân tố như sau:

Quan hệ mờ giữa m quan sát liên tiếp của chuỗi thời gian hai nhân tố $(F_A(t-1), F_B(t-1)), (F_A(t-2), F_B(t-2)), \dots, (F_A(t-m), F_B(t-m))$ và $F_A(t)$ được biểu diễn bởi $(F_A(t-m), F_B(t-m)), \dots, (F_A(t-2), F_B(t-2)), (F_A(t-1), F_B(t-1)) \rightarrow F_A(t)$. Nếu, đặt $F_A(t) = A_j(t)$ và $F_A(t-q) = A_{iq}(t-q)$; $F_B(t-q) = B_{iq}(t-q)$, với $1 \leq q \leq m$. Thì quan hệ tại thời điểm t này được biểu diễn thành $(A_{im}(t-m), B_{km}(t-m)), \dots, (A_{i2}(t-2), B_{k2}(t-2)), (A_{i1}(t-1), B_{k1}(t-1)) \rightarrow A_j(t)$.

Nếu cũng tại thời điểm t , tồn tại các quan hệ mờ bậc m hai nhân tố có cùng về trái sau:

$$(A_{im}(t-m), B_{km}(t-m)), \dots, (A_{i2}(t-2), B_{k2}(t-2)), (A_{i1}(t-1), B_{k1}(t-1)) \rightarrow A_{j1}(t_1).$$

$$(A_{im}(t-m), B_{km}(t-m)), \dots, (A_{i2}(t-2), B_{k2}(t-2)), (A_{i1}(t-1), B_{k1}(t-1)) \rightarrow A_{jp}(t_p).$$

với $t_1, t_2, \dots, t_p \leq t$. Nghĩa là các quan hệ tại thời điểm t_1, t_2, \dots, t_p xảy ra trước quan hệ mờ ở tại thời điểm t , nhưng có cùng các tập mờ bên về trái. Khi đó các quan hệ tại thời điểm t này được gom thành một nhóm quan hệ mờ theo dạng sau:

$(A_{im}(t-m), B_{km}(t-m)), \dots, (A_{i2}(t-2), B_{k2}(t-2)), (A_{i1}(t-1), B_{k1}(t-1)) \rightarrow A_{j1}(t_1), \dots, A_{jp}(t_p), A_j(t)$ và được gọi là nhóm quan hệ mờ phụ thuộc vào thời gian bậc cao m ($m \geq 2$) hai nhân tố.

2.2. Thuật toán PSO [44]

PSO là thuật toán tiến hóa dựa trên việc mô phỏng hành vi tương tác của bầy chim hay đàn cá tìm nguồn thức ăn. Mỗi con chim (cá thể) trong đàn (quần thể) được đặc trưng bởi hai thành phần là vector vị trí x_i và vector vận tốc (dịch chuyển) v_i trong không gian (d-1) chiều. Ban đầu PSO được khởi tạo bởi vị trí và vận tốc một cách ngẫu nhiên. Sau mỗi bước dịch chuyển (lặp) mỗi cá thể đánh giá khả năng tìm kiếm bằng hàm đo độ thích nghi (fitness function). Đồng thời mỗi cá thể cập nhật vận tốc v_i và vị trí x_i của nó theo công thức (1) và (2). Cũng trong quá trình lặp, mỗi cá thể chịu ảnh hưởng bởi hai thông tin: Thông tin thứ nhất là vị trí tốt nhất mà mỗi cá thể ghi được cho tới thời điểm hiện tại (thông tin cục bộ), gọi là $P_{best,i}$ và bị tác động bởi thành phần tự nhận thức C_1 . Thông tin thứ hai là vị trí tốt nhất trong tất cả quá trình tìm kiếm của toàn bộ cá thể trong quần thể cho tới thời điểm hiện tại (thông tin toàn cục), gọi là G_{best} và bị tác động bởi thành phần xã hội C_2 .

$$V_i^{k+1} = \omega^k * V_i^k + C_1 * Rand() * (P_{best,i} - x_i^k) + C_2 * Rand() * (G_{best} - x_i^k) \tag{1}$$

$$x_i^{k+1} = x_i^k + v_i^{k+1} \tag{2}$$

$$\omega^k = \omega_{max} - \frac{k * (\omega_{max} - \omega_{min})}{iter_{max}} \tag{3}$$

Vị trí tốt nhất của mỗi cá thể i là một vector $P_{best,i} = [p_{i,1}, p_{i,2}, \dots, p_{i,n-1}]$ và giá trị $P_{best,i}$ được tính như sau:

$$P_{best,i}^{k+1} = f(x) = \begin{cases} P_{best,i}^k, & \text{if } fitness(x_i^{k+1}) > P_{best,i}^k \\ fitness(x_i^{k+1}), & \text{if } fitness(x_i^{k+1}) \leq P_{best,i}^k \end{cases} \tag{4}$$

$$\text{Giá trị } G_{best} \text{ tại lần lặp thứ } k \text{ là: } G_{best} = \min(P_{best,i}^k) \tag{5}$$

Trong đó;

- V_i^k, x_i^k là vận tốc và vị trí của cá thể thứ i ở lần lặp thứ k .
- ω^k là trọng số quán tính ở thời điểm thứ k , được giảm tuyến tính từ giá trị $\omega_{max} = 0.9$ đến $\omega_{min} = 0.4$ và giới hạn bởi số lần lặp tối đa ($iter_{max}$) theo công thức (3).

III. MÔ HÌNH DỰ BÁO DỰA TRÊN FTS BẬC CAO HAI NHÂN TỐ VÀ PSO

Trong phần này, chúng tôi trình bày mô hình dự báo dựa trên chuỗi thời gian mờ bậc cao hai nhân tố kết hợp với tối ưu bầy đàn. Mô hình dự báo đề xuất sử dụng hai chuỗi thời gian mờ $F_A(t)$ và $F_B(t)$, trong đó $F_A(t)$ được gọi là nhân tố chính và $F_B(t)$ là nhân tố thứ hai. Trước tiên mô hình cần phải xây dựng cách phân khoảng tập nền. Trong một số mô hình trước đây, tập nền được chia thành những khoảng đều nhau. Những năm gần đây, việc phân khoảng có độ dài khác nhau được sử dụng thường xuyên nhằm nâng cao độ chính xác dự báo của mô hình. Một số phương pháp thường được sử dụng như các phương pháp phân cụm, giải thuật di truyền và phương pháp tối ưu bầy đàn. Bài báo này, chúng tôi sử dụng PSO để tìm độ dài khoảng tối ưu cho từng nhân tố trong tập nền nhằm tăng độ chính xác dự báo của mô hình. Cấu trúc của mô hình dự báo và thuật toán tìm khoảng tối ưu bằng PSO (thuật toán này được đặt tên là **PSO-optimal interval**) được trình bày theo các bước sau:

Bước 1: Xác định tập nền cho hai chuỗi dữ liệu quan sát $F_A(t)$ và $F_B(t)$.

Giả sử $U_A(t) = [D_{min} - D_1, D_{max} + D_2]$ và $V_B(t) = [E_{min} - E_1, E_{max} + E_2]$ là hai tập nền tương ứng của nhân tố chính $F_A(t)$ và nhân tố thứ hai $F_B(t)$.

Trong đó; D_{min}, D_{max} là giá trị nhỏ nhất và lớn nhất của chuỗi dữ liệu quan sát $F_A(t)$ và D_1, D_2 là hai số dương được chọn sao cho tập nền $U_A(t)$ bao chọn vẹn chuỗi dữ liệu lịch sử và đảm bảo nhiều của dữ liệu kiểm thử của nhân tố chính. Tương tự E_{min}, E_{max} là giá trị nhỏ nhất và lớn nhất của chuỗi dữ liệu quan sát $F_B(t)$ và E_1, E_2 là hai số dương được chọn sao cho tập nền $U_B(t)$ chứa tất cả các quan sát của nhân tố thứ hai.

Bước 2: Chia tập nền $U_A(t)$ và $V_B(t)$ thành n và m khoảng tối ưu

Trước tiên mỗi tập nền chứa mỗi chuỗi thời gian có thể được chia thành các khoảng có độ dài bằng nhau với số lượng khoảng bất kỳ. Một cách tổng quát có thể chia tập nền $U_A(t)$ của nhân tố chính $F_A(t)$ thành n khoảng, tập nền $V_B(t)$ của nhân tố thứ hai $F_B(t)$ thành m khoảng với độ dài tương ứng là $L_1 = \frac{(D_{max}-D_{min})+(D_2+D_1)}{n}$ và $L_2 = \frac{(E_{max}-E_{min})+(E_2+E_1)}{m}$. Sau đó gọi thuật toán **PSO-optimal interval** để hiệu chỉnh độ dài của các khoảng này thành các khoảng có độ dài khác nhau phù hợp hơn.

Bước 3: Xác định các tập mờ cho hai nhân tố dựa trên các khoảng tối ưu đạt được trong bước 2

Bước 4: Mờ hóa dữ liệu cho từng nhân tố thành các tập mờ

Bước 5: Thiết lập quan hệ mờ bậc cao m ($m \geq 2$) hai nhân tố dựa trên các tập mờ xác định ở bước 3.

Bước 6: Tạo nhóm quan hệ mờ phụ thuộc thời gian bậc cao hai nhân tố

Bước 7: Tính toán và giải mờ dự báo từ các nhóm quan hệ mờ bậc cao được thiết lập trong bước 6

Bước 8: Tính sai số dự báo

Để xác minh tính hiệu quả của mô hình dự báo, sai số trung bình (Mean Square Error- MSE) và sai số trung bình tuyệt đối (Mean Absolute Percentage Error – MAPE) được sử dụng làm tiêu chí đánh giá độ chính xác dự báo của mô hình.

$$MSE = \frac{1}{n} \sum_{i=m}^n (F_i - R_i)^2 \tag{6}$$

$$MAPE = \frac{1}{n} \sum_{i=m}^n \left| \frac{F_i - R_i}{R_i} \right| * 100\% \tag{7}$$

Thuật toán PSO-optimal interval:

Thuật toán được chia làm hai bước chính để tìm khoảng tối ưu cho mô hình dự báo như sau:

Bước 1: Khởi tạo và cấu trúc khoảng cho các cá thể trong PSO

Giả sử số lượng khoảng chia ban đầu trên tập nền $U_A(t) = [D_{min} - D_1, D_{max} + D_2]$ của nhân tố chính là n và trên tập nền $V_B(t) = [E_{min} - E_1, E_{max} + E_2]$ của nhân tố thứ hai là m .

Trong nghiên cứu này, vị trí của mỗi cá thể trong PSO biểu diễn các khoảng trên tập nền của nhân tố chính $F_A(t)$ và nhân tố thứ hai $F_B(t)$. Cụ thể nó được cấu trúc thành một mảng gồm $n + m - 2$ phần tử. Trong đó gồm:

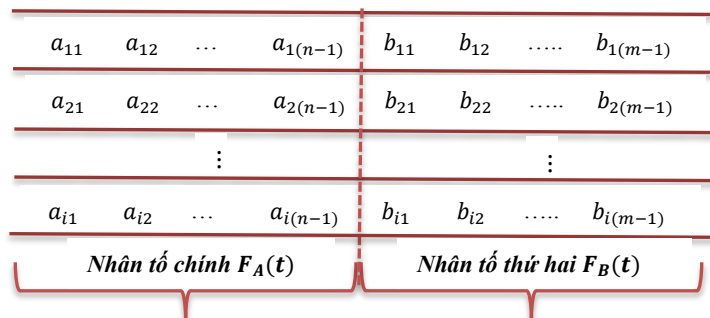
- ❖ $n-1$ phần tử đặc trưng cho nhân tố chính xác định trên tập nền $U_A(t)$ là $\{ a_{i1}, a_{i2}, \dots, a_{ik}, \dots, a_{i(n-1)} \}$, với $1 \leq k \leq n-1, a_{i(k-1)} < a_{ik}$ và $1 \leq i \leq pn$ (pn là số lượng cá thể trong quần thể);
- ❖ $m-1$ phần tử đặc trưng cho nhân tố thứ hai xác định trên tập nền $V_B(t)$ là $\{ b_{i1}, a_{i2}, \dots, a_{ij}, \dots, a_{i(m-1)} \}$, với $1 \leq j \leq m-1, a_{i(j-1)} < a_{ij}$.

Dựa vào $n + m - 2$ phần tử này, xác định được $n+m$ khoảng là $u_1 = (D_{min} - D_1, a_{i1}]$, $u_2 = (a_{i1}, a_{i2}]$, ..., $u_n = (a_{i(n-1)}, D_{max} + D_2]$ và $v_1 = (E_{min}, b_{i1}]$, $v_2 = (b_{i1}, b_{i2}]$, ..., $v_m = (b_{i(m-1)}, D_{max}]$

Khi mỗi cá thể di chuyển đến vị trí khác thì các phần tử trong mảng tương ứng với vị trí mới cần được sắp xếp theo thứ tự tăng dần tương ứng với từng nhân tố. Cấu trúc mã hóa của hai nhân tố theo vị trí của các cá thể trong PSO được biểu diễn như Hình 1.

Trong đó, - vị trí của cá thể thứ nhất được khởi tạo đúng bằng các giá trị của các khoảng trên tập nền tương ứng với từng nhân tố của chuỗi thời gian.

- ❖ Vị trí của các cá thể còn lại khởi tạo ngẫu nhiên trong tập nền tương ứng với từng nhân tố.
- ❖ Khởi tạo vị trí tốt nhất ban đầu $P_{best,i} = [p_{i,1}, p_{i,2}, \dots, p_{i,n-1}]$ của cá thể thứ i đúng bằng vị trí của các cá thể ban đầu.
- ❖ Vận tốc của các cá thể được khởi tạo ngẫu nhiên trong khoảng $[-V_{max}, V_{max}]$, Với $V_{max} = 0.5 * (\frac{D_{max} - D_{min} + D_2 - D_1}{2})$ được chọn cho nhân tố chính, $V_{max} = 0.5 * (\frac{E_{max} - E_{min} + E_2 - E_1}{2})$ được chọn cho nhân tố thứ hai.



Hình 1: Cấu trúc mã hóa của i cá thể trong PSO

Bước 2: Tính giá trị hàm mục tiêu cho mỗi cá thể i dựa vào công thức (6) hoặc (7) bằng các bước con sau:

- Thiết lập số lần lặp ban đầu $k=1$ ($1 \leq k \leq iter_max$).

Bước 2.1: Xác định các tập mờ cho các quan sát của chuỗi thời gian mờ hai nhân tố

Mỗi khoảng đạt được trong bước 1 chứa một nhân ngôn ngữ đặc trưng bởi tập mờ chứa một hay nhiều giá trị quan sát của chuỗi thời gian. Để xác định các tập mờ cho mỗi quan sát của từng nhân tố trong chuỗi thời gian, bằng cách gán mỗi khoảng tương ứng đạt được trong bước 1 bởi một tập mờ. Đối với n khoảng của nhân tố chính $F_A(t)$, xác định được n tập mờ A_i được biểu diễn theo công thức (8).

$$A_i = \frac{e_{i1}}{u_1} + \frac{e_{i2}}{u_2} + \dots + \frac{e_{ij}}{u_j} + \dots + \frac{e_{in}}{u_n} \tag{8}$$

$$e_{ij} = \begin{cases} 1 & j == i \\ 0.5 & \text{if } j == i - 1 \text{ or } j == i + 1 \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

Trong đó $e_{ij} \in [0,1]$ ($1 \leq i \leq n, 1 \leq j \leq n$) được biểu thị là cấp độ thuộc của khoảng u_j vào tập mờ A_i . Để thuận tiện cho việc giải mờ đầu ra dự báo, trong bài báo này chúng tôi sử dụng hàm thuộc tam giác với cấp độ thuộc nhận ba giá trị 0, 0.5, 1 theo công thức (9).

Một cách tương tự đối với m khoảng của nhân tố thứ hai $F_B(t)$, xác định m tập mờ B_k ($1 \leq k \leq m$) được biểu diễn bởi công thức (10) như sau.

$$\begin{aligned}
 B_1 &= \frac{1}{v_1} + \frac{0.5}{v_2} + \frac{0}{v_3} + \dots + \frac{0}{v_{m-1}} + \frac{0}{v_m} \\
 B_2 &= \frac{0.5}{v_1} + \frac{1}{v_2} + \frac{0.5}{v_3} + \dots + \frac{0}{v_{m-1}} + \frac{0}{v_m} \\
 &\dots\dots\dots \\
 B_m &= \frac{0}{v_1} + \frac{0}{v_2} + \frac{0}{v_3} + \dots + \frac{0.5}{v_{m-1}} + \frac{1}{v_m}
 \end{aligned} \tag{10}$$

Bước 2.2. Mờ hóa dữ liệu lịch sử của chuỗi thời gian mờ hai nhân tố trên cơ sở các tập mờ đã xác định ở bước 2.1

Việc mờ hóa có thể coi là một quá trình xác định mối liên hệ giữa các giá trị lịch sử của chuỗi thời gian với các tập mờ được xác định trong bước 2.1. Để mờ hóa tất cả các giá trị lịch sử trên từng nhân tố của chuỗi thời gian, trước tiên cần gán các giá trị ngôn ngữ cho mỗi khoảng theo cấp độ thuộc cao nhất mà mỗi dữ liệu lịch sử của mỗi nhân tố thuộc vào khoảng này. Các giá trị ngôn ngữ được gán ở đây là các tập mờ đã xác định trên bước 3.

Bước 2.3. Xác định các quan hệ mờ bậc m hai nhân tố

Dựa trên định nghĩa 3 về chuỗi thời gian mờ hai nhân tố, thì một quan hệ logic mờ có thể được xây dựng bởi hai hay nhiều tập mờ liên tiếp trong mỗi chuỗi thời gian. Sau khi các giá trị quan sát của hai chuỗi thời gian $F_A(t)$ và $F_B(t)$ được mờ hóa, chúng ta có thể xác định được tất cả các quan hệ logic mờ hai nhân tố bậc m theo cách thức sau: $(F_A(t - m), F_B(t - m)), \dots, (F_A(t - 2), F_B(t - 2)), (F_A(t - 1), F_B(t - 1)) \rightarrow F_A(t)$; trong đó $(F_A(t - m), F_B(t - m)), \dots, (F_A(t - 2), F_B(t - 2)), (F_A(t - 1), F_B(t - 1))$ được gọi là trạng thái hiện tại của quan hệ và $F_A(t)$ gọi là trạng thái tương lai của quan hệ. Sau đó các quan hệ này có thể được thay bởi các tập mờ tương ứng trên từng nhân tố.

Bước 2.4: Thiết lập nhóm quan hệ mờ phụ thuộc thời gian bậc m hai nhân tố

Dựa trên định nghĩa 5 về nhóm quan hệ mờ phụ thuộc thời gian bậc cao hai nhân tố. Giả sử tồn tại các quan hệ mờ bậc hai với hai nhân tố được biểu diễn bởi các tập mờ A_i và B_j tại các thời điểm khác nhau như sau:

- Tại $t = t_1$ có mỗi quan hệ mờ là $(A_1, B_2), (A_2, B_2) \rightarrow A_5$
- Tại $t = t_2$ có mỗi quan hệ mờ là $(A_2, B_3), (A_2, B_2) \rightarrow A_4$
- Tại $t = t_3$ có mỗi quan hệ mờ là $(A_1, B_2), (A_2, B_2) \rightarrow A_4$

. Khi đó, tại thời điểm $t = t_1, t_2, t_3$ (với $t_1 < t_2 < t_3$) chúng ta lần lượt nhận được các nhóm quan hệ mờ bậc hai hai nhân tố với cùng trạng thái bên trái là $G_1: (A_1, B_2), (A_2, B_2) \rightarrow A_5$; $G_2: (A_2, B_3), (A_2, B_2) \rightarrow A_4$ và $G_3: (A_1, B_2), (A_2, B_2) \rightarrow A_5, A_4$.

Bước 2.5: Giải mờ và tính giá trị dự báo đầu ra

Để giải mờ dữ liệu đã mờ hóa và tính toán giá trị cho nhóm quan hệ mờ bậc cao hai nhân tố. Trước tiên, một phương pháp giải trong công trình [16] được sử dụng để tính toán giá trị dự báo cho các nhóm quan hệ mờ hai nhân tố với các bậc khác nhau trong giai đoạn huấn luyện. Sau đó, chúng tôi sử dụng quy tắc giải mờ trong [23] để tính toán giá trị dự báo cho nhóm quan hệ mờ trong giai đoạn thử nghiệm. Các kỹ thuật giải mờ đưa ra bởi các quy tắc sau:

Quy tắc 1: Tính toán giá trị dự báo cho các nhóm quan hệ mờ trong pha huấn luyện

Để tính giá trị đầu ra cho nhóm quan hệ mờ phụ thuộc thời gian bậc m và giá trị dự báo tương ứng với từng thời điểm thuộc nhóm quan hệ mờ. Chúng tôi sử dụng phương pháp giải mờ [16] bằng việc xem xét thêm thông tin của các tập mờ xuất hiện bên vế phải của các quan hệ mờ trong cùng nhóm. Ý tưởng chính của quy tắc dự báo như sau:

Giả nhóm quan hệ mờ phụ thuộc thời gian bậc m hai nhân tố có dạng $(A_{im}, B_{im}), (A_{i(m-1)}, B_{i(m-1)}), \dots, (A_{i1}, B_{i1}) \rightarrow A_{j1}, A_{j2}, \dots, A_{jp}$. Trong đó, $A_{im}, A_{i(m-1)}, \dots, A_{i1}$ biểu diễn giá trị mờ hóa của nhân tố chính và $B_{im}, B_{i(m-1)}, \dots, B_{i1}$ biểu thị giá trị mờ hóa của nhân tố thứ hai. Khi đó, đối với mỗi nhóm quan hệ này chúng tôi chia các khoảng tương ứng với các tập mờ bên vế phải của nhóm quan hệ thành q khoảng con có độ dài bằng nhau. Sau đó tính toán giá trị cho các nhóm này bằng tổng trung bình giá trị điểm giữa của các khoảng con thứ q tương ứng với p tập mờ bên vế phải của nhóm quan hệ theo công thức (11).

$$\text{Forecasted} = \frac{1}{p} \sum_{j=1}^p \text{sub_mid}_{jk} \tag{11}$$

Trong đó, ($1 \leq j \leq p, 1 \leq k \leq q$)

- ✓ p là tổng số tập mờ bên vế phải của nhóm;

- ✓ sub_mid_{jk} là điểm giữa của một trong q khoảng con (điểm giữa của khoảng con thứ k) tương ứng với tập mờ thứ j bên vế phải của nhóm quan hệ.

Quy tắc 2: Tính giá trị dự báo cho nhóm quan hệ mờ rỗng (Nhóm quan hệ có vế phải chưa xác định tập mờ).

Để tính toán giá trị dự báo cho nhóm quan hệ mờ trong giai đoạn thử nghiệm, tức là nhóm quan hệ được hình thành từ các quan hệ mà vế bên phải của nó chưa có dữ liệu để mờ hóa. Chúng tôi áp dụng phương pháp giải mờ trong [23] bằng việc xem xét biến động của quan hệ gần nhất về tương lai với sự chênh lệch của các biến động liên kế theo bậc trong nhân tố chính.

Giả sử xuất hiện nhóm quan hệ mờ bậc m hai nhân tố như sau: $(A_{im}, B_{im}), (A_{i(m-1)}, B_{i(m-1)}), \dots, (A_{i1}, B_{i1}) \rightarrow \#$. Khi đó dự báo cho mẫu chưa biết (dấu #) vào ngày thứ i được tính theo công thức (12) sau:

$$\text{Forecasted}(\#) = M_{i1} + \sum_{k=2}^m \frac{M_{i(k-1)} - M_{ik}}{2^{k-1}} \quad (12)$$

Trong đó, $M_{i1}, M_{i2}, \dots, M_{ik}, (2 \leq k \leq m)$ là điểm giữa của các khoảng $u_{i1}, u_{i2}, \dots, u_{ik}$ tương ứng với các tập mờ của nhân tố chính.

Bước 2.6. Cập nhật vị trí tốt nhất của mỗi cá thể i, $P_{best_i} = [p_{i,1}, p_{i,2}, \dots, p_{i,n-1}]$ theo công thức (4) và vị trí tốt nhất trong số các vị trí tốt của toàn bộ cá thể trong quần thể (G_{best}) tại lần lặp thứ k theo (5) là: $G_{best} = \min(P_{best_i}^k)$

Trong PSO mỗi cá thể trong quần thể được đặc trưng bởi hai thành phần tốt nhất đó là: Vector vị trí tốt nhất của từng cá thể i là P_{best_i} và vị trí tốt nhất trong cả quần thể là G_{best} . Mỗi cá thể i được cập nhật theo giá trị của hàm mục tiêu MAPE (7). Nếu giá trị của MAPE ở thời điểm hiện tại k nhỏ hơn giá trị MAPE ở thời điểm trước đó k-1 thì $P_{best_i} = \text{MAPE}(x_i^k)$ và $G_{best} = \min(P_{best_i})$.

Bước 2.7. Cập nhật các thành phần trong PSO theo các công thức sau:

- Cập nhật vận tốc của mỗi cá thể theo công thức (1)
- Cập nhật vị trí của mỗi cá thể theo công thức (2)
- Cập nhật trọng số ω của mỗi cá thể theo công thức (3)

Bước 2.8: Kiểm tra điều kiện dừng

Nếu số lượng lần lặp hiện tại còn nhỏ hơn số lần lặp tối đa ($k < \text{iter_max}$) hay (chưa tìm thấy giải pháp tối ưu), quay lại bước Bước 2. Trái lại đưa ra giá trị G_{best} và các khoảng tối ưu u_i của nhân tố chính và các khoảng tối ưu v_j của nhân tố thứ hai xác định trên từng tập nền tương ứng.

IV. KẾT QUẢ THỰC NGHIỆM

Mục này đưa ra một số kết quả dự báo của mô hình đề xuất trên hai tập dữ liệu. Trong kết quả thực nghiệm thứ nhất, chúng tôi áp dụng mô hình đề xuất cho dự báo nhiệt độ trung bình hàng ngày tại Đài Bắc Đài Loan với toàn bộ dữ liệu [22] từ 1/6/1996 đến 30/9/1996. Trong đó, “nhiệt độ trung bình” là nhân tố chính và “mật độ của mây” được đặt là nhân tố thứ hai để tham gia vào dự báo cho nhân tố nhiệt độ. Đối với kết quả thực nghiệm thứ hai, chúng tôi áp dụng mô hình đề xuất để dự báo thị trường chứng khoán TAIFEX [22] chứa toàn bộ dữ liệu lịch sử từ ngày 03/8/1998 đến ngày 30/9/1998. Trong đó, TAIFEX (Taiwan futures exchange) được đặt làm nhân tố chính và TAIEX (Taiwan stock exchange capitalization weighted stock index) đặt làm nhân tố thứ hai. Để so sánh hiệu quả dự báo của mô hình đề xuất với một số mô hình dự báo gần đây trên từng bộ dữ liệu tương ứng, chúng tôi thực hiện mô phỏng chương trình 20 lần chạy trên từng bậc và ghi lại kết quả tốt nhất trong tất cả các chạy làm kết quả dự báo cuối cùng. Các tham số cần thiết để thực hiện mô hình dự báo trên từng bộ dữ liệu được đưa ra trong Bảng 1.

Bảng 1. Các tham số sử dụng trong mô hình dự báo trên hai tập dữ liệu

Các tham số của mô hình	Giá trị cho dự báo nhiệt độ		Giá trị cho dự báo TAIFEX	
	Nhiệt độ	Mật độ của mây	Nhân tố TAIFEX	TAIEX
Số lượng cá thể trong quần thể: $pn =$	30		30	
Số lần lặp tối đa (số thế hệ): $iter_max =$	150		150	
Trọng số quán tính ω (Giảm tuyến tính)	$\omega_{max} = 0.9$ to $\omega_{min} = 0.4$		$\omega_{max} = 0.9$ to $\omega_{min} = 0.4$	
Các hệ số tự tin cá thể và hệ số xã hội $C_1 = C_2$	2		2	
Miền giới hạn vị trí của mỗi cá thể: $X =$	[23, 32]	[0, 100]	[6200, 7600]	[6200, 7600]
Miền giới hạn vận tốc của mỗi cá thể: $V =$	[-5, 5]	[-50, 50]	[-700, 700]	[-700, 700]

4.1. Kết quả dự báo nhiệt độ trung bình tại Đài Bắc Đài Loan

Phần này, chúng tôi áp dụng mô hình đề xuất để dự báo nhiệt độ trung bình hàng ngày tại Đài Bắc, Đài Loan với toàn bộ các quan sát [22] từ ngày 1/6/1996 đến 30/9/1996 về nhiệt độ và độ che phủ của mây được sử dụng để dự báo cho nhân tố chính “nhiệt độ”. Từ các tham số trong Bảng 1, chúng tôi thực hiện mô phỏng và đưa ra kết quả dự báo nhiệt độ trung bình hàng ngày trong tháng 6/1996 dựa trên mô hình chuỗi thời gian bậc 3 hai nhân tố với 9 khoảng chia

cho nhân tố “*hiệt độ*” và 7 khoảng chia cho nhân tố “*mật độ của mây*”. Kết quả dự báo được ghi ra trong Bảng 2 như sau.

Bảng 2. Kết quả dự báo nhiệt độ trung bình hàng ngày trong tháng 6/1996 tại Đài Bắc, Đài Loan.

Ngày tháng	Dữ liệu nhiệt độ	Mật độ của mây	Giá trị dự báo
June-1-96	26.1	36	Not forecasted
June-2-96	27.6	23	Not forecasted
June-3-96	29	23	Not forecasted
June-4-96	30.5	10	30.53
June-5-96	30	13	29.99
June-6-96	29.5	30	29.5
-----	---	---	---
June-28-96	27.8	55	27.8
June-29-96	29	29	28.96
June-30-96	30.2	19	30.26

Để xác minh tính ưu việt của mô hình đề xuất dựa trên chuỗi thời gian mờ bậc cao hai nhân tố với số lượng khoảng chia bằng 9 cho nhân tố chính “*hiệt độ*” và bằng 7 cho nhân tố thứ hai “*mật độ của mây*”, các mô hình trong các bài báo [43, 21, 22, 23, 5] được lựa chọn cho việc so sánh. Kết quả so sánh bằng độ chính xác dự báo MAPE (7) của mô hình đề xuất với các mô hình trước đây trên từng tháng (từ tháng 6 đến tháng 9 năm 1996) dựa trên mô hình chuỗi thời gian mờ bậc cao hai nhân tố (từ bậc 2 đến bậc 8) được đưa ra từ Bảng 3 đến Bảng 6 như sau:

Bảng 3. So sánh độ chính xác dự báo trong tháng 6 giữa mô hình đề xuất và các mô hình khác dựa trên các bậc khác nhau

Mô hình	Bậc của mô hình						
	Bậc 2	Bậc 3	Bậc 4	Bậc 5	Bậc 6	Bậc 7	Bậc 8
Mô hình [43]	0.28%	0.29%	0.30%	0.29%	0.29%	0.28%	0.29%
Mô hình [21]	0.8%	0.76%	0.79%	0.76%	0.79%	0.79%	0.81%
Mô hình [22]	$\alpha=0.25$	0.44%	0.42%	0.42%	0.44%	0.40%	0.40%
	$\alpha=0.5$	0.50%	0.45%	0.42%	0.38%	0.43%	0.39%
	$\alpha=0.9$	0.46%	0.42%	0.44%	0.42%	0.41%	0.46%
Mô hình [23]	0.36%	0.34%	0.32%	0.31%	0.31%	0.28%	0.29%
Mô hình [5]	0.25%	0.26%	0.27%	0.27%	0.27%	0.28%	0.29%
Mô hình đề xuất	0.08%	0.069%	0.06%	0.071%	0.071%	0.055%	0.078%

Bảng 4. So sánh độ chính xác dự báo trong tháng 7 giữa mô hình đề xuất và các mô hình khác dựa trên các bậc khác nhau

Mô hình	Bậc của mô hình						
	Bậc 2	Bậc 3	Bậc 4	Bậc 5	Bậc 6	Bậc 7	Bậc 8
Mô hình [43]	0.34%	0.35%	0.34%	0.34%	0.35%	0.33%	0.32%
Mô hình [21]	0.96%	0.96%	0.98%	0.97%	1.00%	0.98%	0.99%
Mô hình [22]	$\alpha=0.25$	0.44%	0.42%	0.42%	0.44%	0.40%	0.40%
	$\alpha=0.5$	0.50%	0.42%	0.38%	0.43%	0.39%	0.46%
	$\alpha=0.9$	0.46%	0.44%	0.42%	0.41%	0.46%	0.39%
Mô hình [23]	0.34%	0.33%	0.33%	0.32%	0.32%	0.34%	0.33%
Mô hình [5]	0.28%	0.29%	0.28%	0.29%	0.30%	0.29%	0.30%
Mô hình đề xuất	0.081%	0.068%	0.059%	0.06%	0.055%	0.051%	0.062%

Bảng 5. So sánh độ chính xác dự báo trong tháng 8 giữa mô hình đề xuất và các mô hình khác dựa trên các bậc khác nhau

Mô hình	Bậc của mô hình						
	Bậc 2	Bậc 3	Bậc 4	Bậc 5	Bậc 6	Bậc 7	Bậc 8
Mô hình [43]	0.23%	0.22%	0.22%	0.22%	0.23%	0.23%	0.22%
Mô hình [21]	1.07%	1.06%	1.08%	1.08%	1.09%	1.07%	1.07%
Mô hình [22]	$\alpha=0.25$	0.44%	0.42%	0.42%	0.44%	0.40%	0.40%
	$\alpha=0.5$	0.50%	0.42%	0.38%	0.43%	0.39%	0.46%
	$\alpha=0.9$	0.46%	0.44%	0.42%	0.41%	0.46%	0.39%
Mô hình [23]	0.31%	0.34%	0.33%	0.33%	0.33%	0.34%	0.35%
Mô hình [5]	0.35%	0.35%	0.35%	0.36%	0.37%	0.35%	0.36%
Mô hình đề xuất	0.068%	0.067%	0.067%	0.058%	0.056%	0.056%	0.067%

Bảng 6. So sánh độ chính xác dự báo trong tháng 9 giữa mô hình đề xuất và các mô hình khác dựa trên các bậc khác nhau

Mô hình	Bậc của mô hình						
	Bậc 2	Bậc 3	Bậc 4	Bậc 5	Bậc 6	Bậc 7	Bậc 8
Mô hình [43]	0.51%	0.49%	0.51%	0.51%	0.53%	0.50%	0.51%
Mô hình [21]	1.01%	0.90%	0.94%	0.96%	0.95%	0.95%	0.95%
Mô hình [22]	$\alpha=0.25$	0.44%	0.42%	0.44%	0.40%	0.40%	0.40%
	$\alpha=0.5$	0.50%	0.42%	0.38%	0.43%	0.39%	0.46%
	$\alpha=0.9$	0.46%	0.44%	0.42%	0.41%	0.46%	0.39%
Mô hình [23]	0.54%	0.56%	0.54%	0.50%	0.51%	0.52%	0.41%
Mô hình [5]	0.36%	0.35%	0.35%	0.37%	0.38%	0.37%	0.38%
Mô hình đề xuất	0.144%	0.121%	0.119%	0.114%	0.106%	0.095%	0.113%

Quan sát sai số dự báo MAPE trên từng bậc của mô hình từ Bảng 3 đến Bảng 6, cho thấy mô hình dự báo đề xuất hiệu quả hơn so với các mô hình trước đây, khi áp dụng để dự báo nhiệt độ trung bình hàng ngày tại Đài Bắc, Đài Loan từ tháng 6 đến tháng 9 năm 1996. Cụ thể mô hình đề xuất đã đưa ra được giá trị MAPE nhỏ hơn nhiều so với các mô hình dự báo đã công bố trong các bài báo [43,21,22,23,5] trên tất cả các bậc của mô hình. Đặc biệt, mô hình đề xuất đưa ra sai số dự báo nhỏ nhất dựa trên quan hệ mờ bậc 7 với sai số phần trăm trên từng tháng, từ tháng 6 đến tháng 9 tương ứng là 0.055%, 0.051%, 0.056%, 0.095%. Từ kết quả đáng tin cậy này cho thấy mô hình đề xuất ưu việt hơn các mô hình trước đây về độ chính xác dự báo.

4.2. Kết quả dự báo chỉ số chứng khoán TAIEX

Ở mục này, mô hình dự báo đề xuất được áp dụng để dự báo TAIEX [22] từ ngày 8/3/1998 đến 9/30/1998 với sự tham gia của chuỗi nhân tố thứ hai TAIEX. Toàn bộ dữ liệu về chỉ số TAIEX và TAIEX được chia làm hai giai đoạn để thực hiện so sánh giữa mô hình đề xuất với các mô hình trước đây theo các bậc và khoảng khác nhau. Trong giai đoạn thứ nhất, dữ liệu TAIEX và TAIEX từ ngày 8/3/1998 đến 9/23/1998 dùng làm giai đoạn huấn luyện. Dữ liệu còn lại được sử dụng làm giai đoạn thử nghiệm.

4.2.1. Kết quả trong giai đoạn huấn luyện

Để xác minh hiệu quả dự báo của mô hình đề xuất dựa trên chuỗi thời gian bậc cao hai nhân tố, các mô hình đưa ra trong các công trình như: Mô hình L06 [21], Mô hình [5], Mô hình [43], mô hình L08 [22], mô hình MPTSO [23], mô hình [38], và mô hình THPSO [24] được lựa chọn để so sánh và đánh giá về độ chính xác dự báo MSE (6). Từ các tham số trong Bảng 2, mô hình được thực hiện 20 lần chạy và ghi lại lần chạy có giá trị MSE nhỏ nhất làm kết quả dự báo cuối cùng. Kết quả và sai số dự báo giữa mô hình đề xuất và các mô hình so sánh được đưa ra trong Bảng 7. Phân khoảng tối ưu trên tập nền của mỗi nhân tố được ghi trong Bảng 8, trong đó tập nền của nhân tố chính TAIEX và nhân tố thứ hai TAIEX được chia thành 16 khoảng giống như số khoảng chia của các mô hình được biểu diễn trong các bài báo [5, 22, 23, 24]. Từ kết quả trong Bảng 7 cho thấy, mô hình đề xuất đưa ra giá trị dự báo sát nhất và có thể xem là hiệu quả hơn cả trong số các mô hình so sánh trên cùng bảng. Cụ thể với số khoảng chia là 16 khoảng cho mỗi nhân tố, mô hình đề xuất nhận được giá trị MSE = 9.29 nhỏ hơn nhiều so với các mô hình so sánh dựa trên nhóm quan hệ mờ bậc 7. Chi tiết đánh giá dựa trên hai điểm khác chính là cách nhóm quan hệ mờ và phương pháp tìm khoảng tối ưu từ tập nền. Trong cách thứ nhất, mô hình đề xuất sử dụng nhóm quan hệ mờ phụ thuộc thời gian để phục vụ trong việc tính toán giá trị dự báo đầu ra, trong khi 7 mô hình còn lại đều sử dụng nhóm quan hệ mờ của Chen [3]. Khác về việc phân khoảng và tìm khoảng tối ưu từ tập nền, thì hai mô hình biểu diễn trong [5, 43] sử dụng phân cụm tự động để tìm khoảng tối ưu, mô hình [38] sử dụng tìm kiếm Tuba với hệ thống suy diễn mờ để điều chỉnh độ dài khoảng chia tập nền. Còn lại các mô hình L06 [21], L08 [22], MPTSO [23], THPSO [24] và mô hình đề xuất đều áp dụng thuật toán tối ưu để điều chỉnh khoảng ban đầu và tìm khoảng chia tối ưu trên mỗi tập nền của từng nhân tố, trong đó hai mô hình trước là sử dụng giải thuật di truyền còn 3 mô hình còn lại sử dụng PSO.

Bảng 7. So sánh kết quả dự báo của mô hình đề xuất với các mô hình khác dựa trên chuỗi thời gian mờ bậc cao khác nhau hai nhân tố với 16 khoảng chia cho mỗi nhân tố

Ngày tháng	Dữ liệu thực	L06 [21]	Mô hình [5]	Mô hình [43]	L08 [22]	MPTSO [23]	Mô hình [38]	THPSO [24]	Mô hình đề xuất
8/3/1998	7552	--	--	--	--	--	--	--	--
8/4/1998	7560	--	--	--	--	--	--	--	--
8/5/1998	7487	--	--	--	--	--	--	--	--
8/6/1998	7462	7450	7456.63	7474.5	--	--	--	--	--
8/7/1998	7515	7550	7515.13	7515.8	--	--	--	--	--
8/10/1998	7365	7350	7381.13	7413.5	--	--	--	--	--
8/11/1998	7360	7350	7381.13	7362.5	--	--	--	--	--
8/12/1998	7330	7350	7325	7345	7329	7325.28	7325.5	7325	7331.62
8/13/1998	7291	7250	7280.33	7295.5	7289.5	7287.48	7292.5	7287.5	7292.22
8/14/1998	7320	7350	7325	7315	7329	7325.28	7325.5	7325	7319.27
8/15/1998	7300	7350	7280.33	7315	7289.5	7287.48	7292.5	7287.5	7299.54
8/17/1998	7219	7250	7192.29	7222	7215	7221.26	7221.5	7221.3	7217.05

8/18/1998	7220	7250	7192.29	7222	7215	7221.26	7221.5	7221.3	7217.05
8/19/1998	7285	7250	7280.33	7282.5	7289.5	7287.48	7292.5	7287.5	7284.79
8/20/1998	7274	7250	7280.33	7282.5	7289.5	7287.48	7272	7287.5	7272.36
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
9/22/1998	6926	6950	6957.09	6940.5	6904.5	6906.7	6923	6916	6928.9
9/23/1998	6852	6850	6845.62	6833.5	6848	6842.05	6850.5	6848.2	6854.41
9/24/1998	6890	6850	6898.29	6888.5	6904.5	6906.7	6899.5	6880.5	6890.19
9/25/1998	6871	6850	6845.62	6888.5	6848	6864.96	6850.5	6880.5	6872.3
9/28/1998	6840	6850	6845.62	6833.5	6848	6842.05	6850.5	6848.2	6841.58
9/29/1998	6806	6850	6781.93	6833.5	6796	6781.01	6796.6	6794.3	6806.26
9/30/1998	6787	6750	6781.93	6788.5	6796	6781.01	6796.6	6794.3	6790.24
MSE		1364.56	287.32	252.47	105.02	92.17	69.98	55.96	9.29

Bảng 8. Kết quả phân khoảng tối ưu cho nhân tố chính TAIFEX và nhân tố thứ hai TAIEX của cá thể thứ 10 trong 30 cá thể.

TAIFEX index	a ₁	a ₂	a ₃	a ₄	a ₅	a ₁₃	a ₁₄	a ₁₅	
6200	6260.61	6397.58	6435.79	6650.78	6688.12	7291	7300.74	7350.13	7600
TAIEX index	b ₁	b ₂	b ₃	b ₄	b ₅	b ₁₃	b ₁₄	b ₁₅	
6200	5993.28	6348.81	6480.15	6618.64	6725.32	7450.63	7672.09	7853.22	7600

4.2.2. Kết quả trong giai đoạn thử nghiệm

Trong tiểu mục này, chúng tôi sử dụng dữ liệu tốt nhất được đào tạo trong pha luyện để dự báo chỉ số TAIFEX mới trong pha kiểm thử. Trong nghiên cứu này, dữ liệu từ ngày 8/3/1998 đến 9/30/1998 được sử dụng trong pha luyện, dữ liệu còn lại từ ngày 9/24/1998 đến ngày 9/30/1998 được dùng làm pha kiểm thử. Để minh chứng hiệu quả của mô hình đề xuất trong pha kiểm thử, các mô hình trong các tài liệu [22, 8, 23, 20] được lựa chọn cho việc so sánh. Giống như các bài báo [8, 20, 23], mô hình đề xuất cũng sử dụng PSO để tìm khoảng tối ưu và sử dụng cùng số khoảng chia là 16. Điểm khác giữa mô hình đề xuất và các mô hình so sánh này là cách nhóm quan hệ mờ hai nhân tố và phương pháp giải mờ. Kết quả đạt được giữa mô hình đề xuất và các mô hình so sánh được đưa ra trong Bảng 9. Dựa vào kết quả này có thể kết luận rằng mô hình dự báo đề xuất đưa ra độ chính xác tốt hơn các mô hình so sánh dựa vào chuỗi thời gian mờ bậc 7 với cùng số khoảng chia tập nền là 16.

Bảng 9. So sánh kết quả và độ chính xác dự báo giữa mô hình đề xuất và các mô hình khác dựa chuỗi thời gian mờ hai nhân tố bậc 7

Date	Actual data	Model [22]	Model [8]	Model [23]	Model [20]	Proposed model
9/24/1998	6890	6959.07	6861.0	6916.62	6878.3	6887.0
9/25/1998	6871	6833.52	6897.8	6886.0	6906.1	6854.16
9/28/1998	6840	6896.95	6912.8	6892.4	6866.8	6852.14
9/29/1998	6806	6863.76	6858.4	6871.54	6844.3	6808.58
9/30/1998	6787	6823.38	6800.5	6859.12	6820.8	6779.73
MSE		2815.69	1955.9	2635.23	938.9	99.89

V. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Thị trường chứng khoán và nhiệt độ hàng ngày là các số liệu biến động phức tạp trong tự nhiên và rất khó để tạo ra mối quan hệ tiềm năng như một mô hình toán học. Do vậy, chuỗi thời gian mờ đã chỉ ra khả năng hiệu quả để giải quyết các bài toán thực tế này. Bài báo đã đưa ra một mô hình dự báo mới dựa trên chuỗi thời gian mờ bậc cao hai nhân tố kết hợp với tối ưu bầy đàn để dự báo nhiệt độ trung bình hàng ngày và thị trường chứng khoán TAIFEX. Kết quả mô phỏng trên mỗi tập dữ liệu này cho thấy mô hình đề xuất có độ chính xác dự báo tốt hơn so với các mô hình dự báo trước đây. Trong nghiên cứu này, chúng tôi dự báo giá tương lai của chứng khoán cũng như nhiệt độ cho ngày tiếp theo chỉ dựa trên xu hướng của hai nhân tố. Việc thêm nhiều nhân tố có chính sách rõ ràng có thể khiến độ chính xác dự báo của mô hình thuận lợi hơn nếu chúng có tương quan tiềm năng với nhau. Do đó, dự báo đa nhân tố dựa trên mô hình chuỗi thời gian mờ sẽ được xem xét cho các nghiên cứu tiếp theo.

VI. REFERENCES

- [1] Q. Song, B. S. Chissom, “Forecasting Enrollments with Fuzzy Time Series - Part I,” Fuzzy set and systems, vol. 54, pp.1-9. 1993a.
- [2] Song, Q., Chissom, B. S., Fuzzy time series and its models, Fuzzy Sets and Systems, vol.54(3), 269-277, 1993b.
- [3] S. M. Chen, “Forecasting Enrollments based on Fuzzy Time Series,” Fuzzy set and systems, vol. 81, pp. 311-319. 1996.
- [4] S. M. Chen, “Forecasting Enrollments based on hight-order Fuzzy Time Series”, Int. Journal: Cybernetic and Systems, N.33, pp. 1-16, 2002.
- [5] S. M. Chen, K Tanuwijaya “Fuzzy forecasting based on high- order fuzzy logical relationships and automatic clustering techniques”, Expert Systems with Applications, 38, 15425–15437, 2011b.

- [6] Chen, S. M., & Chung, N. Y. "Forecasting enrollments of students by using fuzzy time series and genetic algorithms". *International Journal of Information and Management Sciences*, 17, 1-17, 2006a.
- [7] Chen, S. M., Chung, N. Y. Forecasting enrollments using high-order fuzzy time series and genetic algorithms. *International of Intelligent Systems*, 21, 485-501, 2006b.
- [8] Kuo, I. H., et al, "An improved method for forecasting enrollments based on fuzzy time series and particle swarm optimization", *Expert systems with applications*, 36, 6108-6117, 2009.
- [9] Hoang Tung, Nguyen Dinh Thuan, Vu Minh Loc. The partitioning method based on hedge algebras for fuzzy time series forecasting, *Journal of Science and Technology*, 54 (5), 571-583, 2016.
- [10] Huang, Y. L. et al. A hybrid forecasting model for enrollments based on aggregated fuzzy time series and particle swarm optimization. *Expert Systems with Applications*, 38, 8014-8023, 2011
- [11] Nguyen Cong Dieu, Nghiem Van Tinh. Fuzzy time series forecasting based on time-depending fuzzy relationship groups and particle swarm optimization, In: *Proceedings of the 9th National conference on Fundamental and Applied Information Technology Research (FAIR'9)*, pp.125-133, 2016
- [12] Chen, S. M, Bui Dang H. P. Fuzzy time series forecasting based on optimal partitions of intervals and optimal weighting vectors. *Knowledge-Based Systems*, 118, 204-216, 2017.
- [13] Lizhu Wang et al. "Determination of temporal information granules to improve forecasting in fuzzy time series". *Expert Systems with Applications*, vol.41, pp.3134-3142, 2014
- [14] Wei Lu et al. "Using interval information granules to improve forecasting in fuzzy time series". *International Journal of Approximate Reasoning*, vol.57, pp.1-18, 2015.
- [15] Zhiqiang Zhang, Qiong Zhu. "Fuzzy time series forecasting based on k-means clustering", *Open Journal of Applied Sciences*, 2,100-103, 2012.
- [16] Nghiem Van Tinh & Nguyen Cong Dieu, improving the forecasted accuracy of model based on fuzzy time series and k-means clustering, *journal of science and technology: issue on information and communications technology*, No.2, 51-60, 2017
- [17] Singh S. R. A simple method of forecasting based on fuzzy time series. *Applied Mathematics and Computation*, 186:330-339, 2007.
- [18] Duru, O. A fuzzy integrated logical forecasting model for dry bulk shipping index forecasting: An improved fuzzy time series approach. *Expert Systems with Applications*, 37, 5372-5380, 2010.
- [19] Domanska, D., & Wojtylak, M. Application of fuzzy time series models for forecasting pollution concentrations. *Expert Systems with Applications*, 39, 7673-7679, 2012.
- [20] Kuo I. H, et al. Forecasting TAIEX based on fuzzy time series and particle swarm optimization. *Expert Systems with Applications*, 237:1494-1502, 2010.
- [21] Lee, L. W. et al. Handling forecasting problems based on two-factors high-order fuzzy time series. *IEEE Transactions on Fuzzy Systems*, 14, 468-477, 2006.
- [22] Lee, L. W. Wang, L. H., & Chen, S. M, "Temperature prediction and TAIEX forecasting based on high order fuzzy logical relationship and genetic simulated annealing techniques," *Expert Systems with Applications*, 34, 328-336, 2008.
- [23] Ling-Yuan Hsu et al. Temperature prediction and TAIEX forecasting based on fuzzy relationships and MTPSO techniques, *Expert Syst. Appl*, 37, 2756-2770, 2010.
- [24] Park J. I., Lee, D. J., Song C. K., Chun M. G. TAIEX and KOSPI 200 forecasting based on two-factors high-order fuzzy time series and particle swarm optimization, *Expert Systems with Applications*, 37, 959-967, 2010.
- [25] M. Bose, K. Mali, A novel data partitioning and rule selection technique for modelling high-order fuzzy time series. *Applied Soft Computing*, <https://doi.org/10.1016/j.asoc.2017.11.011>, 2017.
- [26] C. Kai, F. F. Ping, C. W. Gang, A novel forecasting model of fuzzy time series based on k-means clustering, in: *Second International Workshop on Education Technology and Computer Science*, China, pp. 223-225, 2010.
- [27] H. T. Liu, M. L. Wei, An improved fuzzy forecasting method for seasonal time series, *Expert Syst. Appl*, 37, 6310-6318, 2010.
- [28] S. M. Chen, K. Tanuwijaya, Multivariate fuzzy forecasting based on fuzzy time series and automatic clustering techniques, *Expert Syst. Appl*, 38, 10594-10605, 2011a.
- [29] J. R. Hwang, S. M. Chen, C. H. Lee, Handling forecasting problems using fuzzy time series, *Fuzzy Sets Syst*, 100, 217-228, 1998.
- [30] M. Sah, K. Degtiarev, Forecasting enrollment model based on first-order fuzzy time series, in: *Proceedings of World Academy of Sciences*, in: *Engineering and Technology*, vol. 1, pp. 132-135, 2005.
- [31] C. H. Cheng, J. R. Chang, C. Yeh, Entropy-based and trapezoid fuzzification-based fuzzy time series approaches for forecasting IT project cost, *Technol. Forecast. Soc. Change*, 73, 524-542, 2006.
- [32] C. H. Cheng, T. L. Chen, C. H. Chiang, Trend-weighted fuzzy time-series model for TAIEX forecasting, in: I. King, J. Wang, L. W. Chan, D. Wang (Eds.), *Neural Information Processing*, vol. 4234, Springer-Verlag, pp. 469-477, 2006.
- [33] J. W. Wang, J. W. Liu. Weighted fuzzy time series forecasting model, in: *Proceedings of the Second International Conference on Intelligent Information and Database Systems: Part I*, Springer-Verlag, Hue, Vietnam, pp. 408-415, 2010.

- [34] C. H. Aladag, M. A. Basaran, E. Egrioglu, U. Yolcu, V. R. Uslu, Forecasting in high order fuzzy times series by using neural networks to define fuzzy relations, *Expert Syst. Appl.*, 36, 4228-4231, 2009.
- [35] C. H. Aladag, U. Yolcu, E. Egrioglu, A high order fuzzy time series forecasting model based on adaptive expectation and artificial neural networks, *Math. Comput. Simul.*, 81, 875-882, 2010.
- [36] W. Qiu, X. Liu, H. Li, A generalized method for forecasting based on fuzzy time series, *Expert Syst. Appl.* 38, 10446-10453, 2011.
- [37] Jiang P, Dong Q, Li P, Lian L. A novel high-order weighted fuzzy time series model and its application in nonlinear time series prediction. *Applied Soft Computing*, 55:44-62, 2017.
- [38] Avazbeigi, M., Hashemi Doulabi, S. H., & Karimi, B. Choosing the appropriate order in fuzzy time series: A new N-factor fuzzy time series for prediction of the auto industry production. *Expert Systems with Applications*, 37, 5630–5639, 2010.
- [39] Wang, L., Liu, X., & Pedrycz, W. Effective intervals determined by information granules to improve forecasting in fuzzy time series. *Expert Systems with Applications*, 40, 5673-5679, 2013.
- [40] Egrioglu, E., Aladag, C. H., Yolcu, U., Uslu, V. R., & Basaran, M. A. Finding an optimal interval length in high order fuzzy time series. *Expert Systems with Applications*, 37, 5052-5055, 2010.
- [41] K. Huarng, H. Yu. A type 2 fuzzy time series model for stock index forecasting, *Physica A*, 353, 445-462, 2005.
- [42] W. K. Wong, E. Bai, A. W. Chu, Adaptive time-variant models for fuzzy-time-series forecasting, *IEEE Trans. Syst. Man Cybern., Part B, Cybern.*, 40, 1531–1542, 2010.
- [43] Nai-Yi Wang a, Shyi-Ming Chen. Temperature prediction and TAIEX forecasting based on automatic clustering techniques and two-factors high-order fuzzy time series. *Expert Systems with Applications*, 36, 2143-2154, 2009.
- [44] Kennedy, J., & Eberhart, R. Particle swarm optimization. *Proceedings of IEEE international Conference on Neural Network*, 1942-1948, 1995.

A TWO FACTOR HIGH-ORDER FUZZY TIME SERIES MODEL COMBINED PARTICLE SWARM OPTIMIZATION FOR FORECASTING TEMPERATURE AND STOCK INDEX

Nghiêm Văn Tinh, Nguyễn Công Điều

ABSTRACT: *In fact, an observation series can be depended on and affected by many different factors. Therefore, to forecast more accurately, it is necessary to consider adding multi factors that have potential correlations in the model. This study proposes a new forecasting model based on two factor high-order fuzzy time series combined with particle swarm optimization(PSO) for forecasting the daily average temperature and stock index. Different from some previous models, In the proposed model, time - variant high- order fuzzy logical relationship groups are built for using in the forecasting process. While, PSO is still used to find the optimal interval in the universe of discourse of each factor. Two the weather temperature and stock market data sets with corresponding factors are "the temperature", "the cloud density" and "TAIFEX index", and "TAIEX index" which are selected to demonstrate the effectiveness of the proposed model and compare it with existing models. From the experimental results shown that the proposed model has better forecasting accuracy than previous forecasted models based on two factor high-order fuzzy time series.*