



DOI:10.22144/ctujos.2024.322

## XÂY DỰNG MÔ HÌNH DỰ BÁO CHO CHUỖI THỜI GIAN KHOẢNG DỰA VÀO CHUỖI ĐIỂM

Võ Văn Tài<sup>1</sup>, Nguyễn Huỳnh Luận<sup>2</sup>, Danh Ngọc Thắm<sup>3</sup>, Tăng Xuân Khánh<sup>1</sup> và Lê Đại Nghiệp<sup>4\*</sup>

<sup>1</sup>Khoa Khoa học Tự nhiên, Trường Đại học Cần Thơ

<sup>2</sup>Khoa Kỹ thuật Cơ - Điện và Máy tính, Trường Đại học Văn Lang

<sup>3</sup>Bộ môn Toán, Trường Đại học FPT Cần Thơ

<sup>4</sup>Khoa cơ bản, Trường Đại học Nam Cần Thơ

\*Tác giả liên hệ (Corresponding author): [ldnghiep@nctu.edu.vn](mailto:ldnghiep@nctu.edu.vn)

### Thông tin chung (Article Information)

Nhận bài (Received): 03/04/2024

Sửa bài (Revised): 03/06/2024

Duyệt đăng (Accepted): 03/07/2024

**Title:** Building a forecasting model for interval time series based on point series

**Author(s):** Vo Van Tai<sup>1</sup>, Nguyen Huynh Luan<sup>2</sup>, Danh Ngọc Thắm<sup>3</sup>, Tang Xuân Khánh<sup>1</sup> and Le Dai Nghiep<sup>4\*</sup>

**Affiliation(s):** <sup>1</sup>Can Tho University, <sup>2</sup>Van Lang University, <sup>3</sup>FPT University-Campus Can Tho, <sup>4</sup>Nam Can Tho University

### TÓM TẮT

Nghiên cứu này đề nghị mô hình dự báo cho chuỗi thời gian khoảng bằng cách tách nó thành hai chuỗi thời gian điểm. Với mỗi chuỗi thời gian điểm, nghiên cứu chuẩn hoá dữ liệu, chia nó thành các cụm thích hợp dựa vào kỹ thuật phân tích cụm mờ, xây dựng mối quan hệ mờ giữa các phần tử trong chuỗi với các cụm và thiết lập nguyên tắc dự báo. Mô hình đề nghị được trình bày chi tiết các bước thực hiện và được minh hoạ bởi ví dụ số. Nó cũng được áp dụng cho hai chuỗi khoảng thực tế và nhận được kết quả cạnh tranh so với các mô hình phổ biến khác.

**Từ khóa:** Chuỗi thời gian khoảng, mô hình dự báo, mối quan hệ mờ, phân tích cụm

### ABSTRACT

This study proposes a forecasting model for interval time series by splitting it into two point time series. For each point time series, the study normalizes the data, divides it into clusters with the appropriate numbers using the fuzzy clustering technique, establishes fuzzy relationships between elements in the series and clusters, and sets up a forecasting rule. The proposed model details the implementation steps and is illustrated by the numerical example. It is also applied to two real interval series and achieves competitive results compared to other popular models.

**Keywords:** Interval time series, forecasting model, fuzzy relationship, cluster analysis

## 1. GIỚI THIỆU

Chuỗi thời gian (CTG) là một kiểu dữ liệu mà giá trị của nó được lưu trữ theo thời gian như phút, giờ, ngày, tuần, tháng, năm, ... Trong sự phát triển của nhiều các lĩnh vực, CTG thường được lưu trữ rất phổ biến, do đó nó được xem là loại dữ liệu quan trọng được rất nhiều nhà thống kê quan tâm (Cai, 2015; Phamtoan et al., 2022). Trong phân tích CTG, dự báo là một hướng nghiên cứu được xem là quan

trọng nhất. Dự báo là nền tảng không thể thiếu cho các kế hoạch, chiến lược phát triển hợp lý của tất cả các lĩnh vực. Một dự báo đúng có thể mang lại nhiều lợi ích về nhiều mặt và một dự báo sai chắc chắn sẽ gây ra thiệt hại, đôi khi rất nghiêm trọng (Sulandari et al., 2020).

Để dự báo CTG, chúng ta thường sử dụng hai phương pháp chính: mô hình hồi quy và mô hình dành riêng cho kiểu dữ liệu này mà nó được gọi là

mô hình CTG (Maia et al., 2008; Vovan, 2019). Để dự báo bằng mô hình hồi quy hiệu quả, chúng ta phải biết được khuynh hướng dạng đồ thị của nó. Thực tế chúng ta thường không có được điều này, nên kết quả dự báo bằng mô hình hồi quy không cao (Aabbasov & Mamedova, 2003). Để khắc phục điều này, mô hình CTG đã được đề xuất. Mô hình CTG được phát triển theo 2 hướng: mờ và không mờ. Mô hình không mờ cho CTG được ứng dụng rất phổ biến ngày nay với các mô hình như tự hồi quy, trung bình trượt, tự hồi quy trung bình trượt và tự hồi quy tích hợp trung bình trượt (ARIMA), trong đó ARIMA được áp dụng phổ biến nhất (Neto et al., 2008). Khi xây dựng mô hình CTG không mờ, chúng ta cũng phải giả sử những điều kiện như sai số là ổn định và chuỗi thời gian phải dừng (Vovan & Lethithu, 2022). Vì thực tế dữ liệu khó thoả điều kiện này nên hiệu quả dự báo của CTG không mờ không cao trong nhiều trường hợp. Khắc phục mô hình CTG không mờ, mô hình CTG mờ đã được đề xuất. Mô hình này xây dựng mối liên kết mờ giữa mỗi phần tử trong chuỗi và định lượng mối quan hệ mờ này để xây dựng nguyên tắc dự báo. Mô hình CTG mờ được đánh giá có nhiều ưu điểm hơn CTG không mờ vì không đòi những điều kiện của dữ liệu và không cần nhiều dữ liệu quá khứ (De et al., 2019; Pant & Kumar, 2022). Mô hình CTG mờ cũng được phát triển theo hai hướng (i) xây dựng mô hình dự báo trực tiếp cho tương lai và (ii) mờ hoá dữ liệu để tạo ra chuỗi mới có nhiều sự liên kết hơn, sau đó sử dụng các mô hình (i) để dự báo cho chuỗi mới được thiết lập. Theo hướng (i) không có nhiều mô hình được đề xuất. Những mô hình tiêu biểu theo hướng (i) được đề xuất bởi Abbasov and Mamedova (2003), Tinh (2020) và IFTS (Tai, 2019). Theo hướng (ii) có rất nhiều mô hình được đề xuất (Alyousifi et al., 2021; Pant & Kumar, 2022). Yêu cầu chung của các mô hình này chỉ mờ hoá số liệu quá khứ mà không dự báo được cho tương lai.

Tất cả các mô hình trên chỉ tập trung vào chuỗi thời gian điểm (PTS). Trong thực tế, chúng ta cũng lưu trữ nhiều dữ liệu bằng chuỗi thời gian khoảng (ITS), như các giá trị thấp nhất và cao nhất của nhiệt độ, độ ẩm, giá vàng, giá dầu và cổ phiếu trong một ngày (Lethikim & Vovan, 2022). Khi chuỗi thời gian được ghi lại không chắc chắn về giá trị cụ thể của các điểm dữ liệu tại các thời điểm nhất định, việc sử dụng chuỗi khoảng là cần thiết để biểu diễn mức độ không chắc chắn này (Han et al., 2008; Phamtoan & Vovan, 2022). Do đó, trong một số trường hợp chuỗi thời gian khoảng thì cần thiết hơn chuỗi thời gian điểm. Từ phân tích trên, chúng ta thấy rằng thực tế đòi hỏi một mô hình dự báo cho ITS. Hướng

nghiên cứu dự báo cho chuỗi thời gian khoảng hầu như chưa được quan tâm nhiều. Nghiên cứu này đề xuất mô hình dự báo cho chuỗi thời gian khoảng khi nó được tách thành hai chuỗi thời gian điểm và việc dự báo cho mỗi chuỗi điểm có sự cải tiến ở những giai đoạn quan trọng theo hướng mờ.

## 2. MÔ HÌNH ĐỀ NGHỊ

Cho chuỗi thời gian khoảng:

$$T(x, y) = \{(x_1; y_1), \dots, (x_n; y_n); x_i, y_i \in R\}.$$

Mô hình dự báo cho chuỗi thời gian khoảng

$T(x, y)$  đề nghị gồm các bước sau:

**Bước 1:** Tìm chuỗi cho trọng tâm:

$I(t) = \{I_1, I_2, \dots, I_n\}$  và chuỗi cho bán kính:

$R(t) = \{R_1, R_2, \dots, R_n\}$ , trong đó

$$I_i = \frac{x_i + y_i}{2}, R_i = y_i - I_i, 1 \leq i \leq n. \quad (1)$$

Mô hình dự báo khoảng lúc này sử dụng hai chuỗi điểm trên để thực hiện. Bởi vì các bước tiếp theo áp dụng giống nhau cho cả hai chuỗi điểm nên các chuỗi được kí hiệu chung là  $X(t) = \{X_1, X_2, \dots, X_n\}$ .

**Bước 2:** Chuẩn hóa cho  $X(t)$  theo phương pháp tính phần trăm chênh lệch giữa hai thời điểm liên tiếp:

$$N(X_{i+1}) = \frac{X_{i+1} - X_i}{X_i} \cdot 100; 1 \leq i \leq n - 1. \quad (2)$$

Sau Bước 2, chúng ta có được chuỗi

$$N(t) = \{N(X_2), N(X_3), \dots, N(X_n)\}.$$

**Bước 3:** Tìm số chum thích hợp cho chuỗi  $N(t)$  theo các bước sau:

**Bước 3.1:** Gọi  $V^{(t)}$  là chuỗi trọng tâm. Đối với lần lặp đầu tiên,  $t = 0$ , chúng ta đặt

$$V^{(0)} = \{v_2^{(0)}, v_3^{(0)}, \dots, v_n^{(0)}\} \\ = \{N(X_2), N(X_3), \dots, N(X_n)\},$$

và một số dương nhỏ  $\varepsilon = 0,0001$ .

**Bước 3.2:** Cập nhật trọng tâm  $V^{(t+1)}$ , trong đó mỗi phần tử của nó tính bởi:

$$v_i^{(t+1)} = \frac{\sum_{j=2}^n v_j^{(t)} \cdot f_\lambda(v_i^{(t)}, v_j^{(t)})}{\sum_{j=2}^n f_\lambda(v_i^{(t)}, v_j^{(t)})}, 2 \leq i \leq n, \quad (3)$$

với

$$f_\lambda(v_i^{(t)}, v_j^{(t)}) = \begin{cases} \exp\left[-\frac{d(v_i^{(t)}, v_j^{(t)})}{\lambda}\right] & \text{nếu } d(v_i^{(t)}, v_j^{(t)}) \leq d_s \\ 0 & \text{nếu } d(v_i^{(t)}, v_j^{(t)}) > d_s \end{cases} \quad (4)$$

$$d(v_i^{(t)}, v_j^{(t)}) = |v_i^{(t)} - v_j^{(t)}|,$$

$$d_s = \frac{1}{(n-1)(n-2)} \sum_{i < j} d(v_i^{(t)}, v_j^{(t)})$$

và  $\lambda$  là một hằng số.

**Bước 3.3:** Xác định điều kiện dừng theo nguyên tắc sau:

Nếu  $\max_i \{d(v_i^{(t+1)}, v_i^{(t)})\} < \varepsilon$  thì Bước 3 sẽ dừng, ngược lại  $V^{(t+1)}$  được đặt thành  $V^{(t)}$  và Bước 3.2 được lặp lại.

Khi Bước 3 kết thúc, nếu trong  $V^{(t)}$  có bao nhiêu phần tử thì  $N(t)$  sẽ được chia thành bấy nhiêu chùm. Các phần tử trong  $N(t)$  hội tụ về cùng một giá trị trong  $V^{(t)}$  sẽ được xếp vào một chùm. Giả sử khi Bước 3 kết thúc, chúng ta sẽ có  $k$  chùm  $\{w_1, w_2, \dots, w_k\}$ .

**Bước 4:** Tìm tâm của mỗi chùm và sắp xếp chúng theo thứ tự tăng dần, chúng ta có các giá trị  $\{c_1, c_2, \dots, c_k\}$ . Khi đó, tập nền ( $US$ ) được thiết lập như sau:

$$US = \left[ c_1 - \frac{c_2 - c_1}{2}; c_k + \frac{c_k - c_{k-1}}{2} \right].$$

Chia  $US$  thành  $k$  khoảng không bằng nhau:

$$U_1 = \left[ c_1 - \frac{c_2 - c_1}{2}; \frac{c_1 + c_2}{2} \right]. \quad (5)$$

$$U_j = \left[ \frac{c_{j-1} + c_j}{2}; \frac{c_{j+1} + c_j}{2} \right], 2 \leq j \leq k-1. \quad (6)$$

$$U_k = \left[ \frac{c_k + c_{k-1}}{2}; c_k + \frac{c_k - c_{k-1}}{2} \right] \quad (7)$$

**Bước 5:** Thiết lập các tập mờ  $A_j$  tương ứng với các khoảng  $U_j, 1 \leq j \leq k$  theo:

$$\mu_{ij} = \exp \left[ -\frac{(N(X_i) - c_j)^2}{2l_j^2} \right], 1 \leq i \leq n-1, \quad (8)$$

trong đó  $l_j$  là độ rộng của khoảng  $U_j$ , được tính bằng khoảng cách giữa giới hạn trên và giới hạn dưới của nó. Chúng ta cũng thiết lập mối quan hệ mờ giữa các tập. Điều này được thực hiện đơn giản bằng cách kiểm tra mỗi điểm dữ liệu nằm trong khoảng nào. Chẳng hạn, nếu  $N(X_1)$  thuộc  $U_3$  và  $N(X_2)$  thuộc  $U_1$  chúng ta có thể thiết lập được mối quan hệ  $A_3 \rightarrow A_1$ . Quá trình này được lặp lại với mỗi  $N(X_i), i = 2, 3, \dots, n$ .

**Bước 6:** Nội suy cho  $N(X)$ . Từ các tập mờ  $A_j$  và các hệ số của mối quan hệ mờ thu được ở Bước 5, các phần tử trong chuỗi  $N(t)$  được nội suy theo quy tắc sau:

$$f_i = \frac{\mu_{1j} \cdot N(X_1) + \dots + \mu_{ij} \cdot N(X_i)}{\mu_{1j} + \dots + \mu_{ij}} = \frac{\sum_{h=1}^i \mu_{hj} \cdot N(X_h)}{\sum_{h=1}^i \mu_{hj}}. \quad (9)$$

**Bước 7:** Nội suy dữ liệu thực tế  $N(t)$  được  $F(t) = \{F(X_2), F(X_3), \dots, F(X_n)\}$ , trong đó  $F(X_i)$  được tính như sau:

$$F(X_i) = \left[ \frac{f_i}{100} + 1 \right] \cdot N(X_i), 1 \leq i \leq n. \quad (10)$$

Giả sử rằng các giá trị nội suy của  $x_i, y_i, N(I_i), N(R_i)$  lần lượt được cho bởi  $F(x_i), F(y_i), F(I_i), F(R_i)$ . Khi đó, chúng ta có thể thu được các giá trị nội suy cho chuỗi gốc  $T(x, y)$  như sau:

$$F(x_i) = F(I_i) - F(R_i); F(y_i) = F(I_i) + F(R_i). \quad (11)$$

**Bước 8:** Dự báo cho tương lai theo nguyên tắc sau:

Nếu  $N(X_n) \in U_j$  và không có mối quan hệ mờ từ  $A_j, \hat{X}_{n+1}$  được xác định bởi:

$$\hat{X}_{n+1} = \left( \frac{1}{n-1} \sum_{i=1}^{n-1} f_i + 1 \right) \cdot N(X_n), \quad (12)$$

trong đó  $f_i$  là giá trị mờ của  $N(X_i), 1 \leq i \leq n-1$ .

Nếu  $N(X_n) \in U_j$  và có mối quan hệ mờ  $A_j \rightarrow A_p$  thì  $\hat{X}_{n+1}$  được xác định bởi:

$$\hat{X}_{n+1} = \left( \frac{c_p}{100} + 1 \right) \cdot N(X_n), \quad (13)$$

trong đó  $c_p$  là tâm của chùm  $A_p$ .

Nếu  $N(X_n) \in U_j$  và có  $q$  mối quan hệ mờ  $A_j \rightarrow A_{p1}$  thì  $A_{p2}, \dots, A_{pq}, \hat{X}_{n+1}$  được xác định bởi:

$$\hat{X}_{n+1} = \left( \frac{1}{q} \sum_{h=1}^q c_{ph} + 1 \right) \cdot N(X_n), \quad (14)$$

trong đó  $c_{ph}$  là tâm của cụm  $A_{ph}$  và  $q = q_1 + q_2 + \dots + q_h$ .

Khi đó, có thể dễ dàng tính được khoảng dự báo  $[\hat{x}_{n+1}, \hat{y}_{n+1}]$  dựa trên quy tắc này. Hơn nữa, có thể dự báo  $\hat{X}_{n+2}$  bằng cách gán  $\hat{X}_{n+1}$  đến  $X(t)$  và lặp lại các bước trên.

### 3. VÍ DỤ MINH HOẠ

Để minh họa thuật toán đề nghị, chuỗi nhiệt độ theo ngày từ ngày 01/10/2023 đến ngày 31/10/2023 của Thành phố Cần Thơ được xem xét. Dữ liệu được

lấy miễn phí tại trang website: [accuweather.com](http://accuweather.com) và được cho bởi Bảng 1.

Mô hình đề nghị thực hiện cho tập dữ liệu này gồm các bước sau:

**Bước 1:** Tính các giá trị tâm và bán kính theo công thức (9) và (10), ta có Bảng 2.

**Bước 2:** Chuẩn hóa chuỗi tâm  $I(t)$  và chuỗi bán kính  $R(t)$ , ta thu được Bảng 3.

**Bảng 1. Nhiệt độ của Thành phố Cần Thơ 10/2023**

| Ngày | $T(x, y)$ | Ngày | $T(x, y)$ |
|------|-----------|------|-----------|
| 01   | [24; 31]  | 17   | [25; 30]  |
| 02   | [25; 27]  | 18   | [26; 34]  |
| 03   | [24; 30]  | 19   | [24; 33]  |
| 04   | [26; 32]  | 20   | [24; 33]  |
| 05   | [25; 32]  | 21   | [26; 32]  |
| 06   | [25; 34]  | 22   | [25; 33]  |
| 07   | [25; 33]  | 23   | [24; 32]  |
| 08   | [26; 34]  | 24   | [24; 32]  |
| 09   | [25; 33]  | 25   | [24; 33]  |
| 10   | [24; 32]  | 26   | [25; 33]  |
| 11   | [26; 35]  | 27   | [25; 34]  |
| 12   | [25; 34]  | 28   | [24; 34]  |
| 13   | [25; 31]  | 29   | [25; 34]  |
| 14   | [26; 33]  | 30   | [24; 32]  |
| 15   | [25; 33]  | 31   | [25; 33]  |
| 16   | [25; 30]  |      |           |

**Bảng 2. Tâm và bán kính các khoảng**

| $I(t)$ | $R(t)$ | $I(t)$ | $R(t)$ |
|--------|--------|--------|--------|
| 27,5   | 3,5    | 29,0   | 4,0    |
| 26,0   | 1,0    | 30,0   | 4,0    |
| 27,0   | 3,0    | 28,5   | 4,5    |
| 29,0   | 3,0    | 28,5   | 4,5    |
| 28,5   | 3,5    | 29,0   | 3,0    |
| 29,5   | 4,5    | 29,0   | 4,0    |
| 29,0   | 4,0    | 28,0   | 4,0    |
| 30,0   | 4,0    | 28,0   | 4,0    |
| 29,0   | 4,0    | 28,5   | 4,5    |
| 28,0   | 4,0    | 29,0   | 4,0    |
| 30,5   | 4,5    | 29,5   | 4,5    |
| 29,5   | 4,5    | 29,0   | 5,0    |
| 28,0   | 3,0    | 29,5   | 4,5    |
| 29,5   | 3,5    | 28,0   | 4,0    |
| 29,0   | 4,0    | 29,0   | 4,0    |
| 27,5   | 2,5    |        |        |

**Bước 3:** Thực hiện Bước 3 cho chuỗi  $N(I)$ , sau 105 lần lặp, thuật toán dừng lại. Khi đó, 30 phần tử của  $N(I)$  hội tụ về 5 phần tử như sau  $\{-4,266; -1,959; 0,102; 2,958; 8,137\}$

Do đó  $N(I)$  được chia thành 5 chòm:

**Bảng 3. Kết quả cho chuỗi  $N(I)$  và  $N(R)$**

| $N(I)$ | $N(R)$  | $N(I)$ | $N(R)$  |
|--------|---------|--------|---------|
| -      | -       | 5,455  | 60,00   |
| -5,455 | -71,429 | 3,448  | 0,000   |
| 3,846  | 200,000 | -5,000 | 12,500  |
| 7,407  | 0,000   | 0,000  | 0,000   |
| -1,724 | 16,667  | 1,754  | -33,333 |
| 3,509  | 28,571  | 0,000  | 33,333  |
| -1,695 | -11,111 | -3,448 | 0,000   |
| 3,448  | 0,000   | 0,000  | 0,000   |
| -3,333 | 0,000   | 1,786  | 12,500  |
| -3,448 | 0,000   | 1,754  | -11,111 |
| 8,929  | 12,500  | 1,724  | 12,500  |
| -3,279 | 0,000   | -1,695 | 11,111  |
| -5,085 | -33,333 | 1,724  | -10,000 |
| 5,357  | 16,667  | -5,085 | -11,111 |
| -1,695 | 14,286  | 3,571  | 0,000   |
| -5,172 | -37,500 |        |         |

$$w_1 = \{-5,455; -3,333; -3,448; -3,279; -5,085; -5,172; -5,000; -3,448; -5,085\}$$

$$w_2 = \{-1,724; -1,695; -1,695; -1,695\}$$

$$w_3 = \{0; 0; 0; 0\}$$

$$w_4 = \{3,846; 3,509; 3,448; 5,357; 5,455, 3,448; 1,754; 1,786; 1,754; 1,724; 1,724; 3,571\}$$

$$w_5 = \{7,407; 8,929\}$$

Tương tự, ta cũng có chuỗi  $N(R)$  hội tụ về 6 phần tử

$\{-71,429; -34,423; 2,287; 29,712; 59,995, 200,00\}$  sau 80 lần lặp. Do đó chuỗi  $N(R)$  được chia thành 6 chòm:

$$w_1 = \{-71,429\}$$

$$w_2 = \{-33,333; -37,500; -33,333\}$$

$$w_3 = \{0; -11,111; 0; 0; 0; 0; 0; 0; 0; 0; 0\}$$

$$w_4 = \{16,667; 28,571; 12,500; 16,667; 14,286; 12,500; 33,333; 12,500; 12,500; 11,111\}$$

$$w_5 = \{60,000\}, w_6 = \{200,000\}$$

**Bước 4:** Từ kết quả của Bước 3, ta xác định tập nền của chuỗi  $N(I)$  và dãy  $N(R)$  như sau:

$$US(I) = [-5,700; 10,695],$$

$$US(R) = [-89,782; 270,000].$$

Tính tâm các chòm, xác định các khoảng  $U_j$  và độ dài của chúng cho hai chuỗi ta thu được Bảng 4.

**Bảng 4. Tâm của các chùm các khoảng và độ dài của chúng đối với  $N(I)$  của chuỗi nhiệt độ**

| Chuỗi  | Trọng tâm | $U_j$              | Độ dài $U_j$ |
|--------|-----------|--------------------|--------------|
| $N(I)$ | -4,367    | [-5,700; -3,035]   | 2,665        |
|        | -1,702    | [-3,035; -0,851]   | 2,184        |
|        | 0,000     | [-0,851; 1,558]    | 2,409        |
|        | 3,115     | [1,557; 5,641]     | 4,084        |
|        | 8,168     | [5,641; 10,695]    | 5,053        |
| $N(R)$ | -71,429   | [-89,782; -53,075] | 36,706       |
|        | -34,722   | [-53,075; -18,909] | 34,167       |
|        | -3,095    | [-18,909; 6,984]   | 25,893       |
|        | 17,064    | [6,984; 38,532]    | 31,548       |
|        | 60,000    | [38,532; 130,00]   | 91,468       |
|        | 200,000   | [130,000; 270,00]  | 140,00       |

**Bước 5:** Các tập mờ  $A_j$  được thiết lập lần lượt với các khoảng  $U_j, 1 \leq j \leq k$ . Vì  $N(I)$  được chia thành 5 chùm nên có 5 tập mờ. Tương tự,  $N(R)$  có 6 tập mờ. Khi đó ta có mối quan hệ mờ của các phần tử trong hai chuỗi và các chùm xây dựng được trình bày trong Bảng 6 và Bảng 7.

**Bước 6:** Nội suy cho  $N(I)$  và  $N(R)$ . Chẳng hạn  $N(I_3) = 7,407$  thuộc khoảng  $U_5, \mu_{15} = 0,026, \mu_{25} = 0,693$  và  $\mu_{35} = 0,989$ . Khi đó

$$f_3 = \frac{\mu_{15} \cdot N(I_1) + \mu_{25} \cdot N(I_2) + \mu_{35} \cdot N(I_3)}{\mu_{15} + \mu_{25} + \mu_{35}} = 5,763.$$

Các giá trị còn lại trong hai dãy  $N(I)$  và  $N(R)$  được tính toán tương tự với kết quả trình bày tại cột  $f_i$  của Bảng 8 và Bảng 9.

**Bước 7:** Tính các giá trị nội suy của  $I(t)$  là  $\hat{I}(t)$ , và  $R(t)$  là  $\hat{R}(t)$ , ta có Bảng 8, Bảng 9 và Bảng 10.

**Bước 8:** Dự báo cho tương lai.

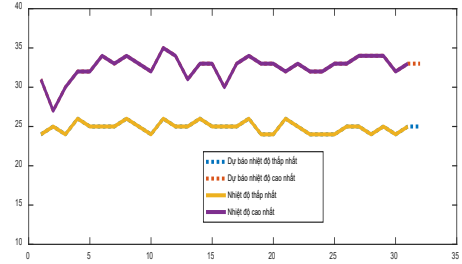
Cho  $I(t)$ : Vì  $f_{30}$  thuộc  $U_4$  và có quan hệ mờ  $A_4 \rightarrow \{3A_1, 3A_2, A_3, 3A_4, A_5\}$ , các giá trị dự báo  $\hat{I}_{32}$  được tính:

$$\hat{I}_{32} = \left[ \frac{\frac{1}{11}(3c_1 + 3c_2 + c_3 + 3c_4 + c_5)}{100} + 1 \right] \cdot I_{31} = 28,982.$$

Đối với  $R(t)$ : Vì  $f_{30}$  thuộc  $U_4$  và có quan hệ mờ  $A_4 \rightarrow \{2A_2, 6A_3, 5A_4\}$ , các giá trị dự báo  $\hat{R}_{32}$  được tính:

$$\hat{R}_{32} = \left[ \frac{\frac{1}{13}(2c_2 + 6c_3 + 5c_4)}{100} + 1 \right] \cdot R_{31} = 3,992.$$

Như vậy, nhiệt độ của Thành phố Cần Thơ vào ngày 1/11/2023 được dự báo là [25; 33].



**Hình 1. Đồ thị dự báo nhiệt độ ngày 1/11/2023**

Hình 1 cho thấy kết quả thực hiện từ mô hình đề nghị khá tốt vì kết quả dự báo và thực tế hầu như trùng nhau.

## 4. MỘT SỐ ÁP DỤNG

### 4.1. Số liệu và phương pháp thực hiện

Trong phần này, nghiên cứu thực hiện hai ứng dụng cho mô hình đề nghị. Chúng là dự báo chứng khoán VN30 index và dự báo giá vàng SJC. Số liệu VN30 được lấy từ ngày 05/09/2023 đến 24/10/2023 từ trang website <https://Investing.com> và số liệu giá vàng SJC được lấy từ 17/07/2023 đến 17/10/2023 trên website <https://webgia.com/gia-vang/sjc/bieu-do-3-thang.html>. Minh họa cho 2 chuỗi thời gian khoảng này được cho bởi Hình 2 và Hình 3.

Phản ứng dụng này được thực hiện theo các bước sau:

**Bước 1:** Chia mỗi tập dữ liệu thành 2 phần, tập huấn luyện và tập kiểm tra với tỷ lệ lần lượt là 80% và 20%.

**Bước 2:** Xây dựng các mô hình trên tập huấn luyện và sử dụng tập huấn luyện để dự báo cho tập kiểm tra để lựa chọn mô hình tốt nhất căn cứ vào tham số Sai số phần trăm tuyệt đối trung bình:

$$MAPE = \frac{1}{2} \left[ \frac{1}{n} \sum_{i=1}^n \left( \frac{|\hat{X}_{il} - X_{il}|}{X_{il}} \cdot 100 \right) + \frac{1}{n} \sum_{i=1}^n \left( \frac{|\hat{X}_{ir} - X_{ir}|}{X_{ir}} \cdot 100 \right) \right].$$

Trong các mô hình dự báo được xây dựng, mô hình nào có MAPE nhỏ nhất thì mô hình đó sẽ tốt nhất.

**Bước 3:** Lấy toàn bộ số liệu, sử dụng mô hình tốt nhất ở trên để dự báo cho 5 ngày tiếp theo.

**Bảng 6. Mối quan hệ mờ của  $N(I)$  của chuỗi nhiệt độ**

| $N(I)$      | $A_1$ | $A_2$ | $A_3$ | $A_4$ | $A_5$ |
|-------------|-------|-------|-------|-------|-------|
| $N(I_1)$    | 0,920 | 0,228 | 0,077 | 0,111 | 0,026 |
| $N(I_2)$    | 0,009 | 0,040 | 0,279 | 0,984 | 0,694 |
| $N(I_3)$    | 0,000 | 0,000 | 0,009 | 0,576 | 0,989 |
| $N(I_4)$    | 0,612 | 1,000 | 0,774 | 0,496 | 0,147 |
| $N(I_5)$    | 0,013 | 0,058 | 0,346 | 0,995 | 0,654 |
| $N(I_6)$    | 0,605 | 1,000 | 0,781 | 0,500 | 0,149 |
| $N(I_7)$    | 0,014 | 0,062 | 0,359 | 0,997 | 0,647 |
| $N(I_8)$    | 0,928 | 0,757 | 0,384 | 0,288 | 0,075 |
| $N(I_9)$    | 0,942 | 0,726 | 0,359 | 0,275 | 0,071 |
| $N(I_{10})$ | 0,000 | 0,000 | 0,001 | 0,363 | 0,989 |
| $N(I_{11})$ | 0,920 | 0,771 | 0,396 | 0,294 | 0,077 |
| $N(I_{12})$ | 0,964 | 0,301 | 0,108 | 0,133 | 0,032 |
| $N(I_{13})$ | 0,001 | 0,005 | 0,084 | 0,860 | 0,857 |
| $N(I_{14})$ | 0,605 | 1,000 | 0,781 | 0,500 | 0,149 |
| $N(I_{15})$ | 0,955 | 0,283 | 0,100 | 0,128 | 0,031 |
| $N(I_{16})$ | 0,001 | 0,005 | 0,077 | 0,849 | 0,866 |
| $N(I_{17})$ | 0,014 | 0,062 | 0,359 | 0,997 | 0,647 |
| $N(I_{18})$ | 0,972 | 0,320 | 0,116 | 0,139 | 0,034 |
| $N(I_{19})$ | 0,261 | 0,738 | 1,000 | 0,748 | 0,271 |
| $N(I_{20})$ | 0,072 | 0,286 | 0,767 | 0,946 | 0,447 |
| $N(I_{21})$ | 0,261 | 0,738 | 1,000 | 0,748 | 0,271 |
| $N(I_{22})$ | 0,942 | 0,726 | 0,359 | 0,275 | 0,071 |
| $N(I_{23})$ | 0,261 | 0,738 | 1,000 | 0,748 | 0,271 |
| $N(I_{24})$ | 0,070 | 0,279 | 0,760 | 0,948 | 0,450 |
| $N(I_{25})$ | 0,072 | 0,286 | 0,767 | 0,946 | 0,447 |
| $N(I_{26})$ | 0,073 | 0,292 | 0,774 | 0,944 | 0,444 |
| $N(I_{27})$ | 0,605 | 1,000 | 0,781 | 0,500 | 0,149 |
| $N(I_{28})$ | 0,073 | 0,292 | 0,774 | 0,944 | 0,444 |
| $N(I_{29})$ | 0,964 | 0,301 | 0,108 | 0,133 | 0,032 |
| $N(I_{30})$ | 0,012 | 0,054 | 0,333 | 0,994 | 0,661 |

**Bảng 7. Mối quan hệ mờ của  $N(R)$  của chuỗi nhiệt độ**

| $N(R)$      | $A_1$ | $A_2$ | $A_3$ | $A_4$ | $A_5$ | $A_6$ |
|-------------|-------|-------|-------|-------|-------|-------|
| $N(R_1)$    | 1,000 | 0,562 | 0,031 | 0,020 | 0,356 | 0,153 |
| $N(R_2)$    | 0,000 | 0,000 | 0,000 | 0,000 | 0,310 | 1,000 |
| $N(R_3)$    | 0,151 | 0,597 | 0,993 | 0,864 | 0,806 | 0,360 |
| $N(R_4)$    | 0,056 | 0,323 | 0,747 | 1,000 | 0,894 | 0,424 |
| $N(R_5)$    | 0,025 | 0,180 | 0,473 | 0,936 | 0,943 | 0,473 |
| $N(R_6)$    | 0,259 | 0,788 | 0,953 | 0,671 | 0,739 | 0,321 |
| $N(R_7)$    | 0,151 | 0,597 | 0,993 | 0,864 | 0,806 | 0,360 |
| $N(R_8)$    | 0,151 | 0,597 | 0,993 | 0,864 | 0,806 | 0,360 |
| $N(R_9)$    | 0,151 | 0,597 | 0,993 | 0,864 | 0,806 | 0,360 |
| $N(R_{10})$ | 0,073 | 0,385 | 0,834 | 0,990 | 0,874 | 0,408 |
| $N(R_{11})$ | 0,151 | 0,597 | 0,993 | 0,864 | 0,806 | 0,360 |
| $N(R_{12})$ | 0,584 | 0,999 | 0,506 | 0,279 | 0,594 | 0,249 |
| $N(R_{13})$ | 0,056 | 0,323 | 0,747 | 1,000 | 0,894 | 0,424 |
| $N(R_{14})$ | 0,066 | 0,358 | 0,798 | 0,996 | 0,883 | 0,415 |
| $N(R_{15})$ | 0,652 | 0,997 | 0,414 | 0,224 | 0,567 | 0,237 |
| $N(R_{16})$ | 0,002 | 0,021 | 0,051 | 0,396 | 1,000 | 0,607 |
| $N(R_{17})$ | 0,151 | 0,597 | 0,993 | 0,864 | 0,806 | 0,360 |

| $N(R)$      | $A_1$ | $A_2$ | $A_3$ | $A_4$ | $A_5$ | $A_6$ |
|-------------|-------|-------|-------|-------|-------|-------|
| $N(R_{18})$ | 0,073 | 0,385 | 0,834 | 0,990 | 0,874 | 0,408 |
| $N(R_{19})$ | 0,151 | 0,597 | 0,993 | 0,864 | 0,806 | 0,360 |
| $N(R_{20})$ | 0,584 | 0,999 | 0,506 | 0,279 | 0,594 | 0,249 |
| $N(R_{21})$ | 0,017 | 0,138 | 0,372 | 0,876 | 0,958 | 0,492 |
| $N(R_{22})$ | 0,151 | 0,597 | 0,993 | 0,864 | 0,806 | 0,360 |
| $N(R_{23})$ | 0,151 | 0,597 | 0,993 | 0,864 | 0,806 | 0,360 |
| $N(R_{24})$ | 0,073 | 0,385 | 0,834 | 0,990 | 0,874 | 0,408 |
| $N(R_{25})$ | 0,259 | 0,788 | 0,953 | 0,671 | 0,739 | 0,321 |
| $N(R_{26})$ | 0,073 | 0,385 | 0,834 | 0,990 | 0,874 | 0,408 |
| $N(R_{27})$ | 0,080 | 0,407 | 0,860 | 0,982 | 0,867 | 0,403 |
| $N(R_{28})$ | 0,247 | 0,770 | 0,965 | 0,692 | 0,746 | 0,325 |
| $N(R_{29})$ | 0,259 | 0,788 | 0,953 | 0,671 | 0,739 | 0,321 |
| $N(R_{30})$ | 0,151 | 0,597 | 0,993 | 0,864 | 0,806 | 0,360 |

**Bảng 8. Kết quả nội suy cho  $N(I)$  của chuỗi nhiệt độ**

| $N(I)$ | $f_i$  | $\hat{I}(t)$ | $N(I)$ | $f_i$  | $\hat{I}(t)$ |
|--------|--------|--------------|--------|--------|--------------|
| -      | -      | -            | 5,455  | 2,408  | 29,698       |
| -5,455 | -      | -            | 3,448  | 2,519  | 30,756       |
| 3,846  | 2,906  | 27,785       | -5,000 | -3,777 | 27,424       |
| 7,407  | 5,763  | 30,671       | 0,000  | -0,649 | 28,315       |
| -1,724 | -2,221 | 27,867       | 1,754  | 2,192  | 29,636       |
| 3,509  | 3,190  | 30,441       | 0,000  | -0,344 | 28,900       |
| -1,695 | -1,852 | 28,463       | -3,448 | -3,509 | 27,018       |
| 3,448  | 2,721  | 30,816       | 0,000  | -0,425 | 27,881       |
| -3,333 | -3,248 | 28,058       | 1,786  | 1,817  | 29,018       |
| -3,448 | -3,295 | 27,078       | 1,754  | 1,813  | 29,526       |
| 8,929  | 4,999  | 32,025       | 1,724  | 1,808  | 30,033       |
| -3,279 | -3,292 | 28,529       | -1,695 | -1,698 | 28,508       |
| -5,085 | -3,584 | 26,997       | 1,724  | 1,701  | 30,002       |
| 5,357  | 2,471  | 30,229       | -5,085 | -3,344 | 27,064       |
| -1,695 | -2,496 | 28,276       | 3,571  | 1,753  | 29,509       |
| -5,172 | -3,632 | 26,501       |        |        |              |

**Bảng 9. Kết quả nội suy  $N(R)$  của chuỗi nhiệt độ**

| $N(R)$  | $f_i$   | $\hat{R}(t)$ | $N(R)$  | $f_i$   | $\hat{R}(t)$ |
|---------|---------|--------------|---------|---------|--------------|
| -       | -       | -            | 60,00   | 10,554  | 4,422        |
| -71,429 | -       | -            | 0,000   | 1,581   | 4,063        |
| 200,000 | 164,048 | 7,921        | 12,500  | 7,587   | 4,841        |
| 0,000   | -2,145  | 2,936        | 0,000   | 2,146   | 4,597        |
| 16,667  | 8,107   | 3,784        | -33,333 | -11,525 | 2,654        |
| 28,571  | 14,899  | 5,170        | 33,333  | 7,897   | 4,633        |
| -11,111 | 4,126   | 4,165        | 0,000   | 1,588   | 4,316        |
| 0,000   | 3,149   | 4,126        | 0,000   | 1,491   | 4,064        |
| 0,000   | 2,546   | 4,102        | 12,500  | 7,375   | 4,060        |
| 0,000   | 2,136   | 4,086        | -11,111 | 1,334   | 4,832        |
| 12,500  | 6,634   | 4,799        | 12,500  | 6,991   | 4,053        |
| 0,000   | 2,951   | 4,633        | 11,111  | 7,193   | 4,815        |
| -33,333 | -10,748 | 2,678        | -10,000 | 1,662   | 5,360        |
| 16,667  | 5,890   | 3,706        | -11,111 | 1,099   | 4,575        |
| 14,286  | 6,709   | 4,268        | 0,000   | 1,050   | 4,044        |
| -37,500 | -11,871 | 2,203        |         |         |              |

**Bảng 10. Kết quả nội suy giá trị của chuỗi nhiệt độ từ mô hình đề nghị**

| $T(x, y)$ | $F(x, y)$        | $T(x, y)$ | $F(x, y)$        |
|-----------|------------------|-----------|------------------|
| [24; 31]  | -                | [25; 33]  | [25,276; 34,120] |
| [25; 27]  | -                | [26; 34]  | [26,692; 34,120] |
| [24; 30]  | [19,863; 35,706] | [24; 33]  | [22,582; 34,819] |
| [26; 32]  | [27,736; 33,607] | [24; 33]  | [23,718; 32,912] |
| [25; 32]  | [24,083; 31,651] | [26; 32]  | [26,982; 32,290] |
| [25; 34]  | [25,270; 35,611] | [25; 33]  | [24,584; 33,216] |
| [25; 33]  | [24,298; 32,628] | [24; 32]  | [22,954; 31,081] |
| [26; 34]  | [26,690; 34,942] | [24; 32]  | [23,822; 31,941] |
| [25; 33]  | [23,956; 32,160] | [24; 33]  | [24,186; 33,850] |
| [24; 32]  | [22,992; 31,163] | [25; 33]  | [25,472; 33,579] |
| [26; 35]  | [27,226; 36,823] | [25; 34]  | [25,219; 34,848] |
| [25; 34]  | [23,896; 33,162] | [24; 34]  | [23,148; 33,867] |
| [25; 31]  | [24,319; 29,674] | [25; 34]  | [25,427; 34,577] |
| [26; 33]  | [26,523; 33,935] | [24; 32]  | [23,020; 31,108] |
| [25; 33]  | [24,008; 32,545] | [25; 33]  | [25,467; 33,551] |
| [25; 30]  | [24,298; 28,704] |           |                  |

**4.2. Kết quả thực hiện**

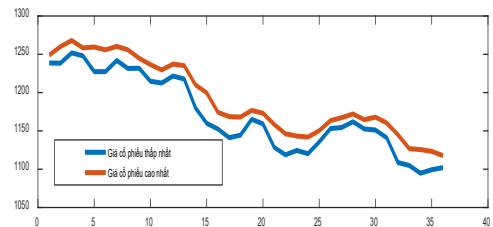
Thực hiện theo các bước đã đề nghị cho hai tập dữ liệu, ta có giá trị MAPE trong Bảng 11.

**Bảng 11. Giá trị MAPE của các mô hình**

| Mô hình                 | VN30  | SJC   |
|-------------------------|-------|-------|
| Mô hình đề nghị         | 0,005 | 0,002 |
| ARIMA                   | 0,012 | 0,007 |
| Abbasov-Mamedova (2013) | 0,835 | 0,294 |
| Vovan (2019)            | 0,741 | 0,252 |

Bảng 11 cho thấy MAPE của mô hình đề nghị có giá trị nhỏ nhất cho cả hai tập dữ liệu, nên mô hình

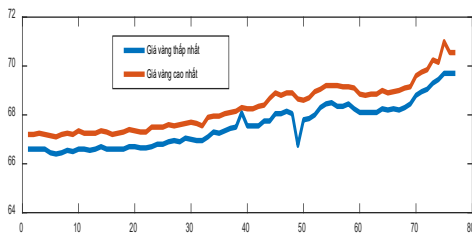
đề nghị là tốt nhất. Do đó nghiên cứu sử dụng mô hình đề nghị và toàn bộ dữ liệu để dự báo cho 5 ngày tiếp theo. Kết quả thực hiện được trình bày trong Bảng 12.



**Hình 2. Đồ thị của dữ liệu cổ phiếu VN30**

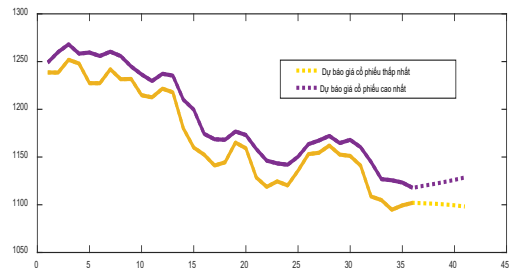
**Bảng 12. Kết quả dự báo cho 5 ngày tiếp theo**

| Dữ liệu  | Biên | Ngày 1  | Ngày 2  | Ngày 3  | Ngày 4  | Ngày 5  |
|----------|------|---------|---------|---------|---------|---------|
| VN index | Dưới | 1101,55 | 1101,12 | 1100,42 | 1099,44 | 1098,15 |
|          | Trên | 1119,82 | 1121,72 | 1123,79 | 1126,08 | 1128,63 |
| SJC      | Dưới | 69,79   | 69,87   | 69,96   | 70,04   | 70,12   |
|          | Trên | 70,62   | 70,70   | 70,77   | 70,84   | 70,91   |



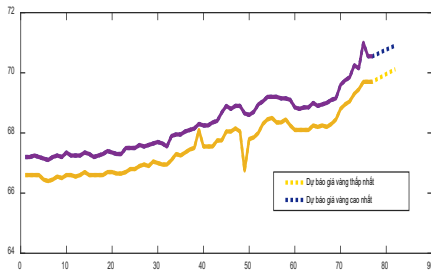
**Hình 3. Đồ thị của dữ liệu giá vàng SJC**

Các giá trị thực tế, dự báo của VN index và SJC được minh họa lần lượt bởi Hình 4 và Hình 5.



**Hình 4. Đồ thị dự báo giá cổ phiếu VN30 cho 5 ngày tiếp theo**





**Hình 5. Đồ thị dự báo giá vàng SJC cho 5 ngày tiếp theo**

Có thể thấy từ Hình 4 rằng dự báo giá cổ phiếu VN30 có sự tăng nhẹ sau thời gian có xu hướng bị giảm xuống. Trong khi đó, từ Hình 5 dự báo giá vàng thấp nhất và cao nhất đang có xu hướng tăng mạnh. Hơn nữa, sự khác biệt giữa giá thấp nhất và giá cao nhất hầu như có sự thay đổi tương đối cao.

### TÀI LIỆU THAM KHẢO

- Abbasov, A. M. & Mamedova, M.H. (2003). Application of fuzzy time series to population forecasting. *Vienna University of Technology*, 12, 545-552.
- Alyousifi, Y., Othman, M., Husin, A., & Rathnayake, U. (2021). A new hybrid fuzzy time series model with an application to predict PM10 concentration. *Ecotoxicology and Environmental Safety*, 227, 112875.
- Cai, Q., Zhang, D., Zheng, W., & Leung, S. C. (2015). A new fuzzy time series forecasting model combined with ant colony optimization and auto-regression. *Knowledge-Based Systems*, 74, 61-68.
- De Lima Silva, P. C., Sadaei, H. J., Ballini, R., & Guimarães, F. G. (2019). Probabilistic forecasting with fuzzy time series. *IEEE Transactions on Fuzzy Systems*, 28(8), 1771-1784.
- Han, A., Hong, Y., Lai, K. K., & Wang, S. (2008). Interval time series analysis with an application to the sterling-dollar exchange rate. *Journal of Systems Science and Complexity*, 21, 558-573.
- Lethikim, N., & Vovan, T. (2022). Fuzzy cluster analysis for interval data based on the overlap distance. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 30(04), 625-648.
- Maia, A. L. S., de Carvalho, F. D. A., & Ludermir, T. B. (2008). Forecasting models for interval-valued time series. *Neurocomputing*, 71(16-18), 3344-3352.
- Neto, E. D. A. L., & De Carvalho, F. D. A. (2008). Centre and range method for fitting a linear regression model to symbolic interval data. *Computational Statistics & Data Analysis*, 52(3), 1500-1515.
- Nguyen-Huynh, L., & Vo-Van, T. (2023). A new fuzzy time series forecasting model based on clustering technique and normal fuzzy function. *Knowledge and Information Systems*, 65(8), 3489-3509.
- Pant, M., & Kumar, S. (2022). Particle swarm optimization and intuitionistic fuzzy set-based novel method for fuzzy time series forecasting. *Granular Computing*, 7(2), 285-303.
- Phamtoan, D., Nguyenthinh, D., & Vovan, T. (2022). Improving the ANFIS forecasting model for time series based on the fuzzy cluster analysis algorithm. *International Journal of Fuzzy System Applications (IJFSA)*, 11(1), 1-20
- Phamtoan, D., & Vovan, T. (2023). The fuzzy cluster analysis for interval value using genetic algorithm and its application in image recognition. *Computational Statistics*, 38(1), 25-51.
- Sulandari, W., Subanar, S., Lee, M. H., & Rodrigues, P. C. (2020). Time series forecasting using singular spectrum analysis, fuzzy systems and neural networks. *MethodsX*, 7, 101015.
- Tinh, N. V. (2020). Enhanced forecasting accuracy of fuzzy time series model based on combined fuzzy C-mean clustering with particle swarm optimization. *International Journal of Computational Intelligence and Applications*, 19(2), 2050017.
- Vovan, T. (2019). An improved fuzzy time series forecasting model using variations of data. *Fuzzy Optimization and Decision Making*, 18, 151-173.
- Vovan, T., & Lethithu, T. (2022). A fuzzy time series model based on improved fuzzy function and cluster analysis problem. *Communications in Mathematics and Statistics*, 10(1), 51-66.

### 5. KẾT LUẬN

Từ một chuỗi thời gian khoảng, nghiên cứu đã tách thành chuỗi thời gian tâm và chuỗi thời gian sai số. Do đó, việc dự báo cho chuỗi thời gian khoảng có thể được thực hiện thông qua hai chuỗi thời gian điểm. Với mỗi chuỗi thời gian điểm, nghiên cứu chọn tập nền là sự biến đổi phần trăm giữa hai giá trị liên tiếp. Tập nền này sau đó được chia thành các khoảng không bằng nhau với một số lượng thích hợp dựa vào mức độ biến đổi nhiều hay ít các phần tử trong chuỗi. Mỗi phần tử sau đó được thiết lập mối quan hệ mờ với khoảng được chia để từ đó xây dựng nguyên tắc dự báo. Kết hợp các vấn đề trên, một nguyên tắc dự báo cho chuỗi thời gian khoảng được đề nghị. Mô hình đề nghị đã được kiểm tra trên hai chuỗi thực tế: chỉ số VN30 và giá vàng SJC. Kết quả cho thấy mô hình đề nghị cho kết quả tốt hơn các mô hình phổ biến gần đây.