

# PHƯƠNG PHÁP DỰ BÁO CHUỖI THỜI GIAN DỰA TRÊN CHUỖI THỜI GIAN MỜ THEO TIẾP CẬN ĐẠI SỐ GIA TỬ

Hoàng Tùng<sup>1</sup>, Nguyễn Đình Thuần<sup>2</sup>, Vũ Minh Lộc<sup>3</sup>

<sup>1</sup>Trường Đại học Đồng Nai

<sup>2</sup>Trường Đại học Công nghệ Thông tin, ĐHQG Tp. Hồ Chí Minh

<sup>3</sup>Trường Đại học Bà Rịa – Vũng Tàu

tung\_k51e@yahoo.com, thuannnd@uit.edu.vn, locvuminh@gmail.com

**TÓM TẮT** — Phương pháp dự báo chuỗi thời gian dựa trên chuỗi thời gian mờ là một vấn đề nghiên cứu nhận được nhiều sự quan tâm trong những năm qua. Trong các nghiên cứu về chuỗi thời gian mờ, các hạng từ dùng để định tính giá trị quan sát của chuỗi thời gian thường được định lượng bằng tập mờ. Các nghiên cứu này đã dùng tập mờ để tiếp cận tới giá trị quan sát, là các hạng từ, của chuỗi thời gian mờ. Ở bài báo này chúng tôi trình bày một phương pháp dự báo chuỗi thời gian dựa trên chuỗi thời gian mờ theo tiếp cận đại số gia tử (ĐSGT). Theo đó, ĐSGT đóng vai trò là một công cụ tiếp cận giúp định lượng các hạng từ là giá trị của chuỗi thời gian mờ. Với cách tiếp cận này, khái niệm về khoảng tính mờ của các hạng từ sẽ giúp ích cho việc xác định các khoảng chia hợp lý trên miền trị tham chiếu của chuỗi thời gian. Từ kết quả thực nghiệm, với số khoảng chia khác nhau, trên một số chuỗi thời gian cho thấy, phương pháp của bài báo mang tới kết quả dự báo chính xác hơn những phương pháp dự báo dựa trên chuỗi thời gian mờ được công bố thời gian gần đây.

**Từ khóa** — Chuỗi thời gian mờ, đại số gia tử, khoảng chia, khoảng chia hợp lý.

## I. GIỚI THIỆU

Có thể phân biệt hai nhóm phương pháp dùng cho dự báo chuỗi thời gian, nhóm thứ nhất dựa trên các mô hình thống kê chẳng hạn như ARMA, ARIMA, MA; nhóm thứ hai dựa trên chuỗi thời gian mờ. Theo [1], số lượng quan sát trên một chuỗi thời gian được gọi là nhỏ, được hiểu, là khi số giá trị lịch sử quan sát được trên chuỗi đó nhỏ hơn 50, trong trường hợp ngược lại được gọi là chuỗi có số lượng quan sát lớn. Từ [2-3] và một số nhận xét trong [4-5] có thể suy ra, mỗi nhóm phương pháp có một thế mạnh riêng, nhóm thứ nhất thường cho kết quả dự báo tốt hơn trên những chuỗi thời gian có số lượng quan sát lớn, ngược lại, nhóm thứ hai thường cho kết quả dự báo chính xác hơn trên những chuỗi thời gian có số lượng quan sát nhỏ. Nhóm phương pháp thứ hai trong suốt những năm qua đã trở thành một đề tài nghiên cứu thu hút được nhiều sự quan tâm.

Nghiên cứu đầu tiên về chuỗi thời gian mờ được công bố bởi hai tác giả Song và Chissom vào năm 1993 [6]. Tiếp theo, cũng hai tác giả này trong [7-8] đã dùng chuỗi thời gian mờ để dự báo lượng thí sinh đăng ký vào đại học Alabama. Từ đó chuỗi thời gian mờ trở thành một công cụ cho phép dự báo chuỗi thời gian. Theo những nghiên cứu này, phương pháp sử dụng chuỗi thời gian mờ trong dự báo chuỗi thời gian bao gồm nhiều bước, nhưng có thể nhóm thành ba pha, pha thứ nhất dùng để mờ hóa chuỗi thời gian cần dự báo; pha thứ hai dùng để xây dựng các suy diễn logic mờ; pha thứ ba thực hiện giải mờ để xác định giá trị dự báo. Độ chính xác của dự báo phụ thuộc nhiều vào pha thứ nhất và thứ ba; trong đó, pha thứ nhất có ảnh hưởng rất lớn tới kết quả dự báo. Thật vậy, ở pha thứ nhất có một việc cần được thực hiện là chia miền trị của chuỗi thời gian thành các khoảng, giá trị dự báo sẽ được lấy ra từ những khoảng chia này ở pha thứ ba. Nói chung độ chính xác trong dự báo của những nghiên cứu này còn khá thấp và độ phức tạp tính toán lại cao, nên các nghiên cứu tiếp sau chủ yếu tập trung cải tiến một tổ hợp các pha trong phương pháp dự báo ở trên nhằm mục tiêu: cải thiện độ chính xác của dự báo, hoặc giảm thiểu độ phức tạp tính toán khi áp dụng phương pháp dự báo.

Trong số những công trình nối tiếp các nghiên cứu của Song và Chissom đáng chú ý có công trình được công bố vào năm 1996 của Chen [9]. Ở nghiên cứu này Chen đề nghị một phương pháp chỉ cần sử dụng các phép toán số học đơn giản trên các khoảng chia để tính giá trị dự báo, thay vì phải dùng các phép toán trên các quan hệ với chi phí tính toán lớn ở pha thứ hai như các nghiên cứu trước. Công trình này đã hình thành nên một hướng nghiên cứu mới trên chuỗi thời gian mờ, ở đó các khoảng chia được nhìn nhận rõ hơn trong vai trò làm cơ sở để tính toán giá trị dự báo, ngoài vai trò cũ là cơ sở để xây dựng các tập mờ dùng ước lượng giá trị của các hạng từ dùng định tính chuỗi thời gian. Các nghiên cứu như [10-13] được dựa trên nguồn cảm hứng từ nghiên cứu của Chen, trong đó [10] là nghiên cứu đầu tiên nhấn mạnh ảnh hưởng của các khoảng chia tới kết quả dự báo.

Qua các nghiên cứu trong hướng này có thể nhận thấy hai kiểu chia khoảng, kiểu thứ nhất vẫn sử dụng lại cách chia miền trị của chuỗi thời gian thành các khoảng bằng nhau như cách mà Song và Chissom đã thực hiện; kiểu thứ hai được thực hiện để tìm ra những khoảng chia không bằng nhau. Thường thì kiểu chia khoảng thứ hai là các nghiên cứu mới hơn và kết quả dự báo cũng chính xác hơn. Ở nghiên cứu của Chen (1996) còn đặt vấn đề định lượng bằng tập mờ các hạng từ dùng định tính các giá trị của chuỗi thời gian nhưng thực chất thao tác này không hỗ trợ cho việc tính toán giá trị dự báo, vì vậy mà ở nhiều những nghiên cứu sau đó thao tác này đã được bỏ qua và chỉ tập trung vào tìm ra phương pháp chia khoảng và xác định giá trị dự báo trên các khoảng chia sao cho hợp lý hơn.

Chuỗi thời gian mờ, về bản chất, là một tập các hạng tử, của một biến ngôn ngữ, được quan sát theo thời gian. Các hạng tử này, trong các nghiên cứu trước đây, thường được định lượng bằng tập mờ; thời gian gần đây, trong công trình [14] đã ứng dụng ĐSGT như một phương án định lượng khác, hình thành nên một tiếp cận mới tới ngữ nghĩa của các hạng tử và đã đem đến một cách nhìn mới về chuỗi thời gian mờ: chuỗi thời gian mờ theo tiếp cận ĐSGT. Trong [15] đã trình bày một phương pháp chia khoảng mới dựa trên chính những hạng tử dùng để quan sát chuỗi thời gian, những hạng tử này được định lượng bằng ĐSGT. Theo cách định lượng của ĐSGT, mỗi hạng tử là đại diện cho một tập giá trị thuộc vào khoảng mờ của nó. Các hạng tử, có thứ tự ngữ nghĩa, dùng để định tính các giá trị quan sát được của chuỗi thời gian bao trọn toàn bộ miền tham chiếu giá trị của chuỗi thời gian,  $U$ ; các khoảng mờ của chúng sẽ là các khoảng nằm liên tiếp nhau từ cận dưới tới cận trên của  $U$ . Như vậy, các khoảng mờ của các hạng tử này, một cách tự nhiên, hình thành nên các khoảng chia trên  $U$ . Cách tiếp cận này khác hoàn toàn với các phương pháp chia khoảng của các tài liệu đã đề cập ở trên, nó cho thấy mối liên hệ giữa các hạng tử và khoảng chia. Tuy vậy ở nghiên cứu này đề xuất cách chia khoảng dựa trên bất kỳ ĐSGT nào có thể dùng cho việc chuẩn hóa các hạng tử của một chuỗi thời gian mờ, cách làm này mang tính tổng quát nhưng có thể gây bối rối cho người áp dụng phương pháp, khi thực hiện với số khoảng chia lớn, vì phải thử sai các ĐSGT. Bài báo này cũng dựa trên cách chia khoảng như [15] nhưng đề nghị dùng thống nhất một ĐSGT chỉ bao gồm hai gia tử, một gia tử âm và một gia tử dương, để chia khoảng và đề nghị một cách mới, đơn giản nhưng khá hiệu quả, để tính giá trị dự báo cho chuỗi thời gian.

Phần còn lại của bài báo được tổ chức như sau: phần hai, trình bày một số khái niệm cơ bản của ĐSGT được tham khảo cho nghiên cứu này; phần ba, trình bày nội dung chính của bài báo, phương pháp dự báo chuỗi thời gian dựa trên chuỗi thời gian mờ theo tiếp cận ĐSGT, phần thứ tư trình bày những kết quả thực nghiệm của phương pháp đề nghị trên một số chuỗi thời gian, phần năm là phần trình bày kết luận của bài báo.

## II. ĐSGT VÀ CHUỖI THỜI GIAN MỜ THEO TIẾP CẬN ĐSGT

Trong phần này chúng tôi tham khảo tài liệu [14, 16] để trình bày vắn tắt một số vấn đề cơ bản trong ĐSGT, những vấn đề này được dùng làm cơ sở cho nghiên cứu của bài báo.

ĐSGT, được các tác giả N.C.Ho và cộng sự công bố trong các công trình nghiên cứu năm 1990 và 1992, là một tiếp cận mới để định lượng hạng tử khác biệt với cách tiếp cận bằng tập mờ. Một ĐSGT được ký hiệu là  $AX = (X, G, C, H, \leq)$  trong đó  $G = \{c^+, c^-\}$  là tập các phần tử sinh,  $C$  bao gồm các phần tử hằng 0, 1,  $W$  theo thứ tự là phần tử bé nhất, lớn nhất và trung hòa trong  $X$ ,  $H$  là tập các gia tử, " $\leq$ " là quan hệ cảm sinh từ ngữ nghĩa của các hạng tử trên  $X$ . Với mỗi hạng tử  $x \in X$  trong ĐSGT,  $H(x)$  là tập các hạng tử  $u \in X$  được sinh ra từ  $x$  bằng cách áp dụng các gia tử trong  $H$  và viết  $u = h_n \dots h_1 x$ , với  $h_n, \dots, h_1 \in H$ .

Tập  $H$  gồm các gia tử dương  $H^+$  và gia tử âm  $H^-$ . Các gia tử dương làm tăng ngữ nghĩa của một hạng tử mà nó tác động, còn gia tử âm có xu hướng ngược lại. Không mất tính tổng quát, có thể giả thiết rằng  $H = \{h_{-1} < h_{-2} < \dots < h_{-q}\}$  và  $H^+ = \{h_1 < h_2 < \dots < h_p\}$ .

Nếu tập  $X$  và  $H$  là các tập sắp thứ tự tuyến tính, khi đó  $AX = (X, G, C, H, \leq)$  là ĐSGT tuyến tính. Hơn nữa, chúng sẽ được gọi là ĐSGT tuyến tính đầy đủ nếu được trang bị thêm hai gia tử tới hạn  $\Sigma$  và  $\Phi$  lần lượt là cận trên và cận dưới đúng của tập  $H(x)$ , khi đó ĐSGT tuyến tính đầy đủ được ký hiệu là  $AX = (X, G, C, H, \Sigma, \Phi, \leq)$ .

ĐSGT đã thể hiện ở sự hợp lý trong việc xây dựng các khái niệm khó xác định trong lý thuyết tập mờ như tính mờ, khoảng tính mờ. Cụ thể, tính mờ của hạng tử ngôn ngữ  $x$  được hiểu là tính chất mà ngữ nghĩa của nó vẫn có thể thay đổi được khi tác động lên nó bằng các gia tử, còn khoảng tính mờ thì được định nghĩa một cách hình thức như sau:

**Định nghĩa 2.1.** Với  $AX = (X, G, C, H, \Sigma, \Phi, \leq)$  là một ĐSGT tuyến tính đầy đủ. Một ánh xạ  $fm: X \rightarrow [0, 1]$  được gọi là một khoảng tính mờ của các hạng tử trong  $X$  nếu:

1.  $fm(c^-) + fm(c^+) = 1$  và  $\sum_{h \in H} fm(hu) = fm(u)$ , với mọi  $u \in X$ ; trong trường hợp này  $fm$  được gọi là đầy đủ.

2. Với các hằng số 0,  $W$  và 1 thì  $fm(0) = fm(W) = fm(1) = 0$ ;

3. Với mọi  $x, y \in X$  và mọi  $h \in H$ ,  $\frac{fm(hx)}{fm(x)} = \frac{fm(hy)}{fm(y)}$ , tỷ số này không phụ thuộc vào  $x, y$  và nó chính là độ

đo tính mờ của gia tử  $h$ , ký hiệu là  $\mu(h)$ . Trong đó, điều kiện (1) thể hiện tính đầy đủ của các phần tử sinh và các gia tử cho việc biểu diễn ngữ nghĩa của miền thực đối với các biến, (2) thể hiện tính rõ của các hạng tử và (3) được chấp nhận với giả thiết các gia tử là độc lập với ngữ cảnh, có nghĩa là khi áp dụng một gia tử  $h$  lên các hạng tử khác nhau thì hiệu quả tác động tương đối làm thay đổi ngữ nghĩa của các hạng tử đó là như nhau.

Tính chất của khoảng tính mờ được làm rõ hơn thông qua mệnh đề 2.1 dưới đây.

**Mệnh đề 2.1.** Với mỗi khoảng tính mờ  $fm$  trên  $X$  những khẳng định sau đây là đúng:

1.  $fm(hx) = \mu(h)fm(x)$ , với mọi  $x \in X$ ;

$$2. fm(c^-) + fm(c^+) = 1;$$

$$3. \sum_{-q \leq i \leq p, i \neq 0} fm(h_i c) = fm(c), c \in \{c^-, c^+\};$$

$$4. \sum_{-q \leq i \leq p, i \neq 0} fm(h_i x) = fm(x)$$

$$5. \sum_{-q \leq i \leq -1} \mu(h_i) = \alpha \text{ và } \sum_{1 \leq i \leq p} \mu(h_i) = \beta, \alpha, \beta > 0 \text{ và } \alpha + \beta = 1$$

Trong [15] đã điều chỉnh định nghĩa về chuỗi thời gian mờ, theo đó, chuỗi thời gian mờ được nhìn theo cách mới, nhìn từ phía ĐSGT với một cách định lượng mới. Cách nhìn này không làm thay đổi bản chất của chuỗi thời gian mờ. Định nghĩa này được phát biểu như sau:

**Định nghĩa 2.2.** Định nghĩa chuỗi thời gian mờ theo tiếp cận ĐSGT

Tập nền  $X(t)$ , một tập con của  $R^1$ , là miền tham chiếu giá trị của các giá trị ngôn ngữ  $X_i(t)$ ,  $F(t)$  là một tập các  $X_i(t)$  thì  $F(t)$  được gọi là một chuỗi thời gian mờ trên  $X(t)$ .

Như đã trình bày trong phần Giới thiệu, chuỗi thời gian mờ về bản chất là một tập các hạng tử của một biến ngôn ngữ, được quan sát theo thời gian. Theo tiếp cận bằng tập mờ các hạng tử  $X_i(t)$  sẽ được định lượng bằng tập mờ. Còn theo tiếp cận ĐSGT các giá trị ngôn ngữ này sẽ được định lượng bằng ánh xạ định lượng ngữ nghĩa và khoảng tính mờ. Rõ ràng định nghĩa chuỗi thời gian mờ theo tiếp cận ĐSGT không khác về bản chất so với định nghĩa của Song và Chissom ở [6].

### III. PHƯƠNG PHÁP ĐỀ NGHỊ

Phương pháp dự báo chuỗi thời gian mà bài báo này đề nghị bao gồm bốn bước. Khác với nghiên cứu [15] xuất phát từ một tập các hạng tử dùng để định tính chuỗi thời gian, sau đó dùng ĐSGT để sinh ra các hạng tử tương đương thay thế, từ những hạng tử này sẽ hình thành nên các khoảng chia. Phương pháp trong bài báo này tiếp cận theo một hướng khác. Mục đích của việc áp dụng phương pháp dự báo là tìm ra các giá trị dự báo càng chính xác càng tốt, như đã trình bày ở phần Giới thiệu, số khoảng chia hợp lý sẽ có ảnh hưởng lớn tới độ chính xác của dự báo, do khi thực hiện dự báo người ta có xu hướng lựa chọn số khoảng chia trước, các hạng tử được xác định sau. Phương pháp mà bài báo đề nghị cũng theo xu hướng này, có nghĩa, các bước của phương pháp sẽ theo tiến trình: xác định miền trị tham chiếu của chuỗi thời gian, ấn định số khoảng chia; dùng ĐSGT chỉ gồm hai gia tử, gia tử âm và dương, để tìm số hạng tử tương ứng với số khoảng chia; xác định giá trị dự báo.

Bài báo quy ước thao tác tìm số hạng tử bằng ĐSGT hai gia tử tương ứng với số khoảng chia, được ấn định từ trước, được gọi là thao tác chia khoảng. Dưới đây trình bày chi tiết các bước của phương pháp.

Phương pháp dự báo chuỗi thời gian dựa trên chuỗi thời gian mờ theo tiếp cận ĐSGT.

Bước 1: Xác định miền trị tham chiếu của chuỗi thời gian,  $F(t)$ ,  $U = [Dmin - D1, Dmax + D2]$ , trong đó  $Dmin$ ,  $Dmax$ ,  $D1$ ,  $D2$  lần lượt là giá trị lịch sử nhỏ nhất, lớn nhất của  $F(t)$ ;  $D1$  và  $D2$  là các giá trị được chọn sao cho các giá trị của  $F(t)$  sẽ thuộc vào  $U$ .

Bước 2:

(1) Ấn định số khoảng cần chia, giả sử là  $k$  ( $k \in N$ ).

(2) Dùng ĐSGT hai gia tử, một gia tử dương,  $h$ , và một gia tử âm,  $h'$ , chia  $U$  thành  $k$  khoảng. Chia từ trái qua phải, mỗi lượt chia ta thu được các hạng tử cùng độ dài, lặp lại với số lượt chia đủ lớn để đạt được  $k$  khoảng.

(3) Tính khoảng tính mờ của các hạng tử, theo định nghĩa 2.1 và mệnh đề 2.1, mỗi khoảng tính mờ của một hạng tử hình thành một khoảng chia trên  $U$ , các khoảng chia này sẽ nằm liên tiếp nhau từ  $Dmin - D1$  tới  $Dmax + D2$ .

(4) Loại bỏ những khoảng không chứa bất kỳ giá trị lịch sử nào của chuỗi thời gian. Giả sử số khoảng này là  $m$  ( $m \geq 1$ ).

(5)

(a) Tìm khoảng có số lượng lớn nhất các giá trị lịch sử của  $F(t)$  rơi vào, nằm trái nhất và có số phần tử thuộc khoảng đó khác nhau đôi một nhiều nhất; giả sử khoảng chia này tương ứng với hạng tử  $A_i$ , để chia thành hai khoảng. Số khoảng được xét để chia tiếp, cho đủ  $k$  khoảng, bao gồm tất cả các khoảng đã thu được ở bước (4) bớt đi một khoảng (khoảng tương ứng với  $A_i$ ) và thêm vào hai khoảng (những khoảng tương ứng với hai hạng tử được sinh từ  $A_i$  là  $hA_i$  và  $h'A_i$ ).

(b) Loại bỏ những khoảng không chứa bất kỳ giá trị lịch sử nào của chuỗi thời gian.

(c) Lặp lại (a) và (b) cho tới khi nào tìm được đủ  $m$  khoảng (để có được đủ  $k$  khoảng) hoặc chuyển sang (d) khi không thể chia tiếp được nữa.

(d) Thực hiện:

d.1. Lấy lại khoảng liền kề bên trái, theo hướng từ phải qua trái, đã bị loại ở bước 3 kết nạp vào số khoảng chia đã có.

d.2. Nếu số khoảng chia bằng  $k$  thì dừng. Nếu không lấy lại khoảng liền kề bên phải, theo hướng từ trái qua phải. Quay lại d.1.

Bước 3: Thiết lập các nhóm suy diễn logic mờ

Xây dựng các suy diễn logic mờ giữa các hạng từ dùng định tính giá trị của chuỗi thời gian, tại các thời điểm kế tiếp nhau theo thời gian. Các suy diễn logic mờ này sẽ có dạng  $A_t \rightarrow A_u$ . Tiếp theo, gom các suy diễn logic mờ có cùng về trái thành nhóm. Kết quả thu được sẽ là các suy diễn logic mờ có dạng  $A_t \rightarrow A_u(p) \dots A_v(q)$ , ở đây  $p, q$  là số lần xuất hiện của  $A_t$  và  $A_v$  trong các quan hệ logic mờ từ  $A_t$

Bước 4: Tính toán giá trị dự báo

Giả sử giá trị của chuỗi thời gian tại thời điểm  $t$  được định tính bởi hạng từ  $A_i$  và hạng từ này là về trái của quan hệ logic mờ  $A_i \rightarrow A_j(m) \dots A_k(n)$ , vậy thì giá trị dự báo tại thời điểm  $t+1$  sẽ bằng:

$$\frac{m * TB(A_j) + \dots + n * TB(A_k)}{m + \dots + n}$$

Trong đó  $TB(A_j), TB(A_k)$  lần lượt là trung bình cộng của những giá trị lịch sử của chuỗi thời gian trong khoảng mờ, lần lượt, tương ứng với các hạng từ  $A_j$  và  $A_k$ .

Với phương pháp được trình bày ở trên chúng ta có thể thực hiện dự báo với số khoảng chia bất kỳ trên miền trị tham chiếu của chuỗi thời gian. Tuy nhiên, nếu số khoảng chia lớn tới mức mỗi khoảng chia chỉ chứa một giá trị lịch sử của chuỗi thời gian sẽ làm mất đi ý nghĩa của việc dùng chuỗi thời gian mờ cho dự báo chuỗi thời gian. Bởi vì dùng các hạng từ để định tính các giá trị của chuỗi thời gian là nhằm gom nhóm những giá trị có cùng chung một tính chất nào đó, nếu mỗi nhóm như thế chỉ có một giá trị thì tính mờ không còn “rõ” nữa. Thêm nữa, việc chia quá nhiều khoảng dường như không thực tế, vì mỗi khoảng làm cơ sở để định lượng một hạng từ; số khoảng sẽ tương ứng bằng với số hạng từ được dùng, thông thường người ta thường chỉ dùng số hạng từ hạn chế để định tính các giá trị của một biến ngôn ngữ.

Ở tài liệu [15] trình bày cách tính giá trị dự báo dựa vào ánh xạ định lượng ngữ nghĩa của các hạng từ,  $A_i$ , và có xét tới khoảng cách giữa  $A_i$  cùng với ánh xạ ngữ nghĩa của hạng từ  $hA_i$  và  $h'A_i$  tới trung bình của các giá trị lịch sử của chuỗi thời gian rơi vào khoảng mờ của  $A_i$ . Cách tính giá trị dự báo này khác với cách tính giá trị dự báo mà chúng tôi đã đề nghị ở Bước 4 trong phương pháp trình bày ở trên. Có thể nói cách tính giá trị dự báo của bài báo này đơn giản hơn khá nhiều.

#### IV. KẾT QUẢ THỰC NGHIỆM VÀ THẢO LUẬN

Trong phần này bài báo trình bày kết quả thực nghiệm khi áp dụng phương pháp đã đề nghị vào việc dự báo: chuỗi thời gian ghi nhận lượng thí sinh đăng ký vào đại học Alabama trong các năm từ năm 1971 tới năm 1992, chuỗi thời gian TAIEX ghi nhận chỉ số chứng khoán Đài Loan với mốc thời gian từ 1/12/1992 tới 29/12/1992, chuỗi thời gian ghi nhận tỷ lệ thất nghiệp, UNE, cũng ở Đài Loan giai đoạn từ 01/01/2013 tới 12/01/2013. Từ đây bài báo sẽ gọi những chuỗi thời gian này lần lượt là Alabama, TAIEX, UNE; trong đó chuỗi Alabama là chuỗi thời gian được các tác giả Song và Chissom sử dụng ở những nghiên cứu đầu tiên về chuỗi thời gian mờ và đã được nhiều tài liệu sau đó sử dụng lại, chuỗi TAIEX và UNE là các chuỗi thời gian được tham khảo từ tài liệu [13].

Ở phần trình bày dưới đây bài báo dùng một số ký hiệu:  $covfm(x)$  là ánh xạ khoảng mờ của hạng từ  $x$  trong đoạn  $[0, 1]$  lên miền trị tham chiếu,  $U$ , của chuỗi thời gian đang xét;  $LU$  là độ rộng của miền  $U$ . Bài báo cũng dùng thông nhất ĐSGT  $AX = (X, G, C, H, \leq)$  với  $G = \{Low, Hight\}$ ,  $C = \{0, 1, W\}$ ,  $H = \{Very, Little\}$  cho dự báo các chuỗi thời gian nêu trên.

Để kiểm nghiệm tính chính xác của dự báo, công thức đánh giá sai số lỗi bình phương trung bình (RMSE) thường được sử dụng.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x'_i - x_i)^2}$$

ở đây  $x'_i$  là giá trị dự báo,  $x_i$  là giá trị lịch sử và  $n$  là số lượng giá trị đã dự báo. Bài

báo này cũng sẽ dùng chỉ số RMSE để so sánh tính chính xác trong dự báo giữa phương pháp đề nghị của bài báo với các phương pháp của Wang và cộng sự (2013), Chen (2013), Wang và cộng sự (2014), Lu và cộng sự (2015).

##### 4.1. Kết quả thực nghiệm trên chuỗi Alabama

Bài báo cũng dùng lại miền trị tham chiếu của chuỗi Alabama,  $U$ , giống như các nghiên cứu trước đây, tức  $U = [13000, 20000]$ , ở đây  $Dmin = 13055$ ,  $Dmax = 19337$ ,  $D1 = 55$ ,  $D2 = 663$ ,  $LU = 7000$ . Để so sánh kết quả dự báo của phương pháp đề nghị với một số phương pháp của những nghiên cứu nêu trên, bài báo sẽ lần lượt sử dụng số

khoảng chia giống như các nghiên cứu này. Cụ thể, các khoảng chia 7, 17 và 22 sẽ được sử dụng, trong đó số khoảng chia là 7 được sử dụng rộng rãi nhất.

Với 7 khoảng chia trên  $U$ , áp dụng các bước của phương pháp đề nghị ta có các kết quả sau:

Nếu coi lượng thí sinh đăng ký học nhỏ hơn 16000 là thấp thì ta có thể thiết lập các tham số:  $fm(low) = \frac{16000 - 13000}{20000 - 13000} = 0.428$ , suy ra  $fm(high) = 0.572$ . Ánh xạ ngược lại miền  $U$  ta có bề rộng của  $cofm(low)$  và  $cofm(high)$  lần lượt là:  $fm(low) \times LU = 0.428 \times 7000 = 2996$ ,  $fm(high) \times LU = 0.572 \times 7000 = 4004$ .

Có thể chọn:  $\mu(Little) = 0.4$ ,  $\mu(Very) = 0.6$ . Từ  $\mu(Little)$  và  $\mu(Very)$  ta có  $\alpha = 0.4$ ,  $\beta = 0.6$ . Ta có:

Các hạng từ tương ứng với 7 khoảng chia: *very very low* (A1), *little very low* (A2), *very little low* (A3), *little little low* (A4), *little little high* (A5), *very little high* (A6) và *very high* (A7). Các khoảng mờ, cũng là các khoảng chia trên miền trị  $U$  của Alabama lần lượt tương ứng với các hạng từ trên, và trung bình các giá trị lịch sử của chuỗi Alabama trên các khoảng đó được trình bày trong bảng dưới đây:

**Bảng 1.** Các khoảng chia và trung bình giá trị lịch sử trên các khoảng chia

Số thứ tự	Khoảng chia	Trung bình của các giá trị lịch sử trên khoảng
1	[13000, 14079)	13478
2	[14079, 14798)	14696
3	[14798, 15517)	15335
4	[15517, 15996)	15816
5	[15996, 16637)	16388
6	[16637, 17598)	16862
7	[17598, 20000]	18932

Từ các khoảng chia ta có được bảng sau:

**Bảng 2.** Lượng thí sinh đăng ký vào đại học Alabama được mờ hóa

Năm	Lượng thí sinh đăng ký	Lượng thí sinh đăng ký được mờ hóa
1971	13055	A1
1972	13563	A1
1973	13867	A1
1974	14696	A2
1975	15460	A3
1976	15311	A3
1977	15603	A4
1978	15861	A4
1979	16807	A6
1980	16919	A6
1981	16388	A5
1982	15433	A3
1983	15497	A3
1984	15145	A3
1985	15163	A3
1986	15984	A4
1987	16859	A6
1988	18150	A7
1989	18970	A7
1990	19328	A7
1991	19337	A7

Từ Bảng 1 ta có được các nhóm quan hệ:

**Bảng 3.** Các nhóm quan hệ giữa các hạng từ ngôn ngữ

Nhóm 1	$A1 \rightarrow A1$ (2), $A1 \rightarrow A2$
Nhóm 2	$A2 \rightarrow A3$
Nhóm 3	$A3 \rightarrow A3$ (4), $A3 \rightarrow A4$ (2)
Nhóm 4	$A4 \rightarrow A4$ , $A4 \rightarrow A6$ (2)
Nhóm 5	$A5 \rightarrow A3$
Nhóm 6	$A6 \rightarrow A5A6A7$
Nhóm 7	$A7 \rightarrow A7$ (4)

Áp dụng cách tính giá trị dự báo của phương pháp đề nghị ta có được kết quả dự báo, thể hiện cùng với kết quả dự báo của một số phương pháp được công bố trong thời gian gần đây [13, 15], trong Bảng 4 như dưới đây.

**Bảng 4.** So sánh kết quả dự báo trên chuỗi Alabama với 7 khoảng chia

Năm	Dữ liệu thực tế	Wang et al 2013	Wang et al 2014	Chen 2013	Lu et al 2015	Phương pháp của [15]	Phương pháp đề nghị
1972	13563	13486	13944	14347	14279	13820	13884
1973	13867	14156	13944	14347	14279	13820	13884
1974	14696	15215	13944	14347	14279	13820	13884
1975	15460	15906	15328	15550	15392	15402	15335
1976	15311	15906	15753	15550	15392	15536	15495
1977	15603	15906	15753	15550	15392	15536	15495
1978	15861	15906	15753	15550	16467	16461	16513
1979	16807	16559	16279	16290	16467	16461	16513
1980	16919	16559	17270	17169	17161	17444	17394
1981	16388	16559	17270	17169	17161	17444	17394
1982	15433	16559	16279	16209	14916	15402	15335
1983	15497	15906	15753	15550	15392	15536	15495
1984	15145	15906	15753	15550	15392	15536	15495
1985	15163	15906	15753	15550	15392	15536	15495
1986	15984	15906	15753	15550	15470	15536	15495
1987	16859	16559	16279	16290	16467	16461	16513
1988	18150	16559	17270	17169	17161	17444	17394
1989	18970	19451	19466	18907	19257	19135	18932
1990	19328	18808	18933	18907	19257	19135	18932
1991	19337	18808	18933	18907	19257	19135	18932
1992	18876	18808	18933	18907	19257	19135	18932
RMSE		578.3	506.0	486.3	445.2	441.3	440.3

Tính toán tương tự ta có kết quả dự báo trên chuỗi Alabama với số khoảng chia 17 và 22. Kết quả dự báo này được tổng hợp với kết quả dự báo của [13] và kết quả dự báo áp dụng phương pháp của [15] thành các Bảng 5, Bảng 6 như sau.

**Bảng 5.** So sánh kết quả dự báo trên chuỗi Alabama với 17 khoảng chia

Năm	Giá trị lịch sử	Giá trị dự báo của Lu et al	Giá trị dự báo của [15]	Phương pháp đề nghị
1972	13563	13678	13544	13563
1973	13867	13678	13906	13867
1974	14696	14602	14683	14696
1975	15460	15498	15443	15447
1976	15311	15192	15395	15404
1977	15603	15641	15620	15603
1978	15861	15827	15919	15923
1979	16807	16744	16827	16833
1980	16919	17618	17559	17535
1981	16388	16392	16406	16388
1982	15433	15410	15433	15447
1983	15497	15498	15395	15404
1984	15145	15192	15160	15154
1985	15163	15567	15540	15539
1986	15984	15567	15540	15539
1987	16859	16744	16827	16833
1988	18150	17618	17559	17535
1989	18970	19036	19060	18970
1990	19328	19574	19167	19152
1991	19337	19146	19167	19152
1992	18876	19146	18878	18876
RMSE		256.3	237.7	237.6

**Bảng 6.** So sánh kết quả dự báo trên chuỗi Alabama với 22 khoảng chia

Năm	Giá trị lịch sử	Dự báo của Lu et al	Dự báo của [15]	Phương pháp đề nghị
1972	13563	13512	13544	13563
1973	13867	13998	13865	13867
1974	14696	14601	14724	14696
1975	15460	15462	15460	15460
1976	15311	15305	15291	15311
1977	15603	15641	15620	15603
1978	15861	15827	15873	15861
1979	16807	16715	16805	16807
1980	16919	17212	16905	16919
1981	16388	16392	16406	16388
1982	15433	15409	15436	15433
1983	15497	15514	15498	15497
1984	15145	15037	15143	15145
1985	15163	15195	15185	15163
1986	15984	16054	15984	15984
1987	16859	16861	16863	16859
1988	18150	18024	18213	18150
1989	18970	19036	18961	18970
1990	19328	19241	19329	19328
1991	19337	19666	19336	19337
1992	18876	18718	18878	18876
RMSE		120.6	18.4	0.0

#### 4.2. Kết quả thực nghiệm trên chuỗi TAIEX và UNE

Áp dụng phương pháp đề nghị để dự báo trên hai chuỗi TAIEX với 07 khoảng chia và UNE với 09 khoảng chia, số khoảng chia được thực hiện giống như phương pháp ở [13], ta có các bảng kết quả như dưới đây.

**Bảng 7.** So sánh kết quả dự báo trên chuỗi thời gian TAIEX với 7 khoảng chia

Ngày	Dữ liệu thực tế	Wang et al 2013	Chen et al 2013	Wang et al 2014	Lu et al 2015	Phương pháp của [15]	Phương pháp đề nghị
02/12/1992	3635.7	3629.3	3740.9	3564.5	3693.1	3709.8	3630.7
03/12/1992	3614.1	3629.3	3740.9	3564.5	3693.1	3709.8	3630.7
04/12/1992	3651.4	3629.3	3740.9	3564.5	3693.1	3709.8	3655.5
05/12/1992	3727.9	3629.3	3740.9	3564.5	3693.1	3709.8	3630.7
07/12/1992	3755.8	3629.3	3740.9	3859.9	3693.1	3709.8	3736.0
08/12/1992	3761	3629.3	3740.9	3859.9	3693.1	3709.8	3736.0
09/12/1992	3776.6	3629.3	3740.9	3859.9	3693.1	3709.8	3755.1
10/12/1992	3746.8	3629.3	3740.9	3859.9	3693.1	3709.8	3755.1
11/12/1992	3734.3	3629.3	3740.9	3859.9	3693.1	3709.8	3736.0
12/12/1992	3742.6	3629.3	3740.9	3859.9	3693.1	3709.8	3736.0
14/12/1992	3696.8	3629.3	3740.9	3859.9	3693.1	3709.8	3736.0
15/12/1992	3688.3	3629.3	3740.9	3564.5	3693.1	3709.8	3676.3
16/12/1992	3674.9	3629.3	3740.9	3564.5	3693.1	3709.8	3676.3
17/12/1992	3668.7	3629.3	3740.9	3564.5	3693.1	3709.8	3676.3
18/12/1992	3658	3629.3	3740.9	3564.5	3693.1	3709.8	3630.7
21/12/1992	3576.1	3629.3	3740.9	3564.5	3693.1	3709.8	3630.7
22/12/1992	3578	3629.3	3477.1	3564.5	3519.4	3442.3	3447.2
23/12/1992	3448.2	3629.3	3477.1	3564.5	3519.4	3442.3	3447.2
24/12/1992	3456	3629.3	3477.1	3413.3	3519.4	3442.3	3447.2
28/12/1992	3327.7	3629.3	3477.1	3413.3	3519.4	3442.3	3447.2
29/12/1992	3377.1	3629.3	3368.1	3413.3	3519.4	3491.4	3487.1
RMSE		114.2	85.7	107.2	75.7	68.9	53.6

**Bảng 8.** So sánh kết quả dự báo trên chuỗi thời gian UNE với 9 khoảng chia

Ngày	Dữ liệu thực tế	Wang et al 2013	Chen et al 2013	Wang et al 2014	Lu et al 2015	Phương pháp của [15]	Phương pháp đề nghị
02/01/2013	7.7	7.39	7.60	7.62	7.58	7.51	7.7
03/01/2013	7.5	7.39	7.60	7.62	7.58	7.51	7.5
04/01/2013	7.5	7.39	7.60	7.62	7.58	7.51	7.5
05/01/2013	7.5	7.39	7.60	7.62	7.58	7.51	7.5
06/01/2013	7.5	7.39	7.60	7.62	7.58	7.51	7.5
07/01/2013	7.3	7.39	7.60	7.62	7.58	7.51	7.5
08/01/2013	7.2	7.39	7.12	7.13	7.07	6.99	7.2
09/01/2013	7.2	6.89	7.12	7.13	7.07	6.99	7.1
10/01/2013	7.2	6.89	7.12	7.13	7.07	6.99	7.1
11/01/2013	7.0	6.89	7.12	7.13	7.07	6.99	7.1
12/01/2013	6.7	6.89	7.12	7.13	7.07	6.99	6.7
	RMSE	0.20	0.18	0.19	0.17	0.16	0.07

Từ kết quả dự báo trên ba chuỗi Alabama, TAIEX, UNE, lấy chỉ số RMSE để đánh giá ta có thể thấy rằng phương pháp đề nghị của bài báo này có dự báo chính xác hơn những phương pháp dự báo khác được công bố thời gian gần đây. Ngoài ra, chúng tôi cho rằng phương pháp dự báo đề nghị khá dễ hiểu và cũng dễ dàng khi áp dụng, đồng thời chi phí tính toán thấp vì chỉ cần dùng các phép toán số học với các thao tác tính toán đơn giản trên các bước.

## V. KẾT LUẬN

Trong bài báo này chúng tôi đã đề nghị một phương pháp dự báo chuỗi thời gian dựa trên chuỗi thời gian mờ theo tiếp cận ĐSGT. Phương pháp được đề nghị gồm bốn bước, bước một: xác định miền trị tham chiếu,  $U$ , của chuỗi thời gian cần dự báo, bước hai: xác định số khoảng chia trên  $U$  và dùng ĐSGT hai gia tử để tìm ra số hạng tử và các khoảng chia, là các khoảng tính mờ, tương ứng; bước ba: xây dựng các nhóm suy diễn logic mờ, bước bốn: tính giá trị dự báo. Với ĐSGT hai gia tử ta có thể đạt được với số khoảng chia tùy ý, bên cạnh đó, cách xác định giá trị dự báo dựa trên trung bình các giá trị lịch sử của chuỗi thời gian rơi vào các khoảng khá đơn giản nhưng lại mang lại kết quả dự báo khá tốt.

Kết quả thực nghiệm trên ba chuỗi thời gian, Alabama, TAIEX và UNE cho thấy phương pháp đề nghị của bài báo này cho kết quả dự báo chính xác hơn một số phương pháp khác được công bố thời gian gần đây. Thêm nữa, tính đơn giản khi áp dụng và chi phí tính toán thấp cũng được thể hiện trong phương pháp đề nghị.

## TÀI LIỆU THAM KHẢO

- [1] Fang-Mei Tseng et al. "Fuzzy ARIMA model for forecasting the foreign exchange market". *Fuzzy Sets and Systems*, vol.118, pp.9–19, 2001
- [2] Chi-Chen Wang. "A comparison study between fuzzy time series model and ARIMA model for forecasting Taiwan export". *Expert Systems with Applications*, vol.38, no.8, pp.9296-9304, 2011.
- [3] K. Senthamarai Kannan and E. Sakthivel. "Fuzzy Time Series Model and ARIMA Model – A Comparative Study". *Indian Journal of Applied Research*, vol.4, no.8, pp.624-636, 2014
- [4] Tahseen Ahmed Jilani, Syed Muhammad Aqil Burney. "A refined fuzzy time series model for stock market forecasting". *Physica A*, vol.387, pp.2857–2862, 2008
- [5] Ufuk Yolcu et al. "A new approach for determining the length of intervals for fuzzy time series". *Applied Soft Computing*, vol.9, pp. 647–651, 2009
- [6] Qiang Song and Brad S. Chissom. "Fuzzy time series and its models". *Fuzzy Sets and Systems*, vol.54, pp.269–277, 1993.
- [7] Qiang Song and Brad S. Chissom. "Forecasting enrollments with fuzzy time series – Part I". *Fuzzy Sets and Systems*, vol.54, pp.1–9, 1993.
- [8] Qiang Song and Brad S. Chissom. "Forecasting enrollments with fuzzy time series – Part II". *Fuzzy Sets and Systems* vol.62, pp.1–8,1994.
- [9] Shyi-Ming Chen. "Forecasting enrollments based on fuzzy time series". *Fuzzy Sets and Systems*, vol.81, pp.311-319, 1996
- [10] Kunhuang Huang. "Effective lengths of intervals to improve forecasting in fuzzy time series". *Fuzzy Sets and Systems*, vol.123, pp.387–394, 2001
- [11] Lizhu Wang, Xiaodong Liu, Witold Pedrycz. "Effective intervals determined by information granules to improve forecasting in fuzzy time series". *Expert Systems with Applications*, vol.40, pp.5673–5679, 2013.
- [12] 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.



- [13] 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.
- [14] Nguyễn Cát Hồ, Nguyễn Công Điều, Vũ Như Lâm. “Ứng dụng đại số gia tử trong dự báo chuỗi thời gian mờ”. *Tạp chí Khoa học và Công nghệ*, Vol 54, No.2, 2016.
- [15] Hoàng Tùng, Nguyễn Đình Thuận, Vũ Minh Lộc. “Phương pháp chia khoảng dựa trên đại số gia tử trong dự báo chuỗi thời gian mờ”. Bài gửi đăng và đã được chấp nhận về mặt nội dung trên tạp chí Khoa học và công nghệ.
- [16] Nguyen Cat Ho, Nguyen Van Long. “Fuzziness measure on complete hedge algebras and quantifying semantics of terms in linear hedge algebras”, *Fuzzy Sets and Systems*, Vol.158, pp.452 – 471, 2007.

## METHOD OF FORECASTING TIME SERIES BASED ON HEDGE ALGEBRAS BASED FUZZY TIME SERIES

**Hoang Tung, Nguyen Dinh Thuan, Vu Minh Loc**

**ABSTRACT** — *Method of using fuzzy time series for forecasting time series is the interesting research topic. Normally, in this research field, linguistic terms (linguistic values) usually are quantified by fuzzy sets. So the fuzzy set is the way to quantify linguistic terms of fuzzy time series. In this paper, we present a method of forecasting based on hedge algebra (HA) for fuzzy time series. HA is used as a tool for quantifying linguistic terms that are values of fuzzy time series. By the this way of quantifying, concept of fuzziness interval of linguistic terms used to determine reasonable intervals on universes of discourse. The experimental results show that the proposed method greatly improves forecasting quality.*

**Keywords** — *Fuzzy time series, hedge algebras, divided intervals, reasonable intervals.*