

NHÓM QUAN HỆ LOGIC MỜ PHỤ THUỘC THỜI GIAN TRONG MÔ HÌNH CHUỖI THỜI GIAN MỜ

Nguyễn Công Điều^{1,2}

¹*Viện Công nghệ thông tin, Viện HLKH&CNVN*

²*Đại học Thăng Long, Đại Kim, Hoàng Mai, Hà Nội*

Email: ncdieu@joit.ac.vn

Đến Toà soạn: 10/3/2014; Chấp nhận đăng: 17/7/2014

TÓM TẮT

Mô hình chuỗi thời gian mờ đang có nhiều ứng dụng trong công tác dự báo, nhất là trong các dự báo kinh tế. Trong những năm gần đây khá nhiều công trình đã được hoàn thành theo hướng nâng cao độ chính xác và giảm khối lượng tính toán trong mô hình chuỗi thời gian mờ như các bài báo của Chen và Hsu, Huang, Kuo, Wu. Hầu hết những phương pháp trên đều dựa vào kĩ thuật tạo các nhóm quan hệ logic mờ của Chen để làm giảm khối lượng tính toán khi chỉ cần thực hiện các phép tính số học thay vì các phép tính min-max như trong các mô hình của Song-Chissom. Yu đã chú ý đến sự lặp lại các thành phần trong nhóm quan hệ logic mờ của Chen và đưa ra mô hình chuỗi thời gian mờ có trọng số.

Trong bài báo này, chúng tôi đưa ra một cách xác định nhóm quan hệ logic mờ phụ thuộc vào thứ tự thời gian. Nhờ khái niệm nhóm quan hệ mờ này và sử dụng mô hình chuỗi thời gian mờ có trọng, chúng tôi đã đưa ra một mô hình chuỗi thời gian mờ cải tiến. Tính hiệu quả của mô hình này được chứng minh khi áp dụng cho số liệu của số lượng sinh viên nhập học và dự báo cho chỉ số chứng khoán Đài Loan.

Từ khóa: chuỗi thời gian mờ, mô hình chuỗi thời gian mờ có trọng, nhóm quan hệ mờ.

1. MỞ ĐẦU

Chuỗi thời gian mờ và mô hình chuỗi thời gian mờ bậc nhất do Song và Chissom [1 - 3] phát triển từ năm 1993. Sau công trình này, một loạt các bài báo của nhiều tác giả khác nhau tiếp tục dựa trên ý tưởng này để dự báo chuỗi thời gian và ứng dụng trong nhiều lĩnh vực khác nhau như dự báo dân số, tài chính, nhiệt độ, nhu cầu điện, v.v. Gần đây có rất nhiều tác giả liên tục cải tiến mô hình chuỗi thời gian mờ để dự báo đạt kết quả chính xác hơn.

Chen [4] đã đưa ra phương pháp mới đơn giản và hữu hiệu hơn so với phương pháp của Song và Chissom bằng cách sử dụng các phép tính số học thay vì các phép tính hợp max-min phức tạp trong xử lý mối quan hệ mờ. Phương pháp của Chen chủ yếu dựa trên phương pháp xây dựng nhóm quan hệ logic mờ. Nhiều công trình tiếp theo đã sử dụng cách tiếp cận này để dự báo cho chuỗi thời gian. Huang [6] đã sử dụng các thông tin có trước trong tính chất của chuỗi thời gian như mức độ tăng giảm để đưa ra mô hình heuristic chuỗi thời gian mờ. Cũng dựa trên tư

tương này, Điều [9] đã sử dụng các hàm xác định mối quan hệ heuristic và cách xác định các điểm giải mờ.

Trong những năm gần đây, một số tác giả đã sử dụng nhiều kỹ thuật khác nhau để tìm mô hình hữu hiệu cho chuỗi thời gian mờ. Những kỹ thuật trong lý thuyết tính toán mềm, khai phá dữ liệu, mạng nơ ron và các giải thuật tiến hoá đều được đưa vào sử dụng. Một hướng khác là sử dụng phương pháp heuristic có tính đến xu hướng như [6, 9] hay sử dụng khái niệm tối ưu đám đông như trong các công trình [7, 11] để xây dựng các thuật toán trong mô hình chuỗi thời gian mờ. Mô hình bậc cao cũng đang được triển khai có hiệu quả bắt đầu từ bài báo của Chen [5] và đang tiếp tục nghiên cứu như trong [12]. Lee [8] và một số tác giả khác đã đưa ra mô hình chuỗi thời gian mờ 2 hay nhiều nhân tố.

Trong công trình này, chúng tôi đưa ra khái niệm mới là nhóm quan hệ mờ phụ thuộc thời gian. Khi xác định nhóm quan hệ logic mờ, Chen chỉ xác định các các tập mờ có cùng về trái trong mỗi quan hệ mờ mà không đề ý đến thời điểm xuất hiện của từng thành phần của nhóm quan hệ trong về phải. Như vậy có thể xảy ra trường hợp có những phần tử trong nhóm quan hệ mờ xuất hiện sau thời điểm t nhưng cũng có mặt để tham gia dự báo phần tử tại thời điểm t . Điều này là không hợp lý. Trong định nghĩa nhóm quan hệ logic mờ mới chúng tôi đưa vào những phần tử trong về phải mà xuất hiện trước thời điểm xuất hiện của thành phần về trái của nhóm quan hệ. Với khái niệm nhóm quan hệ logic mờ phụ thuộc thời gian và sử dụng mô hình chuỗi thời gian mờ có trọng số chúng tôi đã xây dựng được một thuật toán mới để dự báo. Độ chính xác của mô hình chuỗi thời gian mờ mới này được cải tiến khá nhiều so với mô hình của Chen và của Yu thông qua hai thí dụ kinh điển là số lượng sinh viên nhập học và chỉ số chứng khoán Đài Loan.

Báo cáo này có 5 mục và phần kết luận. Sau phần mở đầu sẽ là phần đưa ra các khái niệm liên quan đến mô hình chuỗi thời gian mờ, đồng thời mô tả các thuật toán cơ bản liên quan đến dự báo thông qua mô hình chuỗi thời gian mờ. Đó là các thuật toán cơ bản của Chen, mô hình có trọng của Yu. Mục 4 đưa ra một cải biên để xác định nhóm quan hệ logic mờ phụ thuộc vào quá trình lịch sử. Mô hình cải biên chuỗi thời gian mờ Mục thứ 5 áp dụng mô hình cải tiến để dự báo số sinh viên nhập học của Đại học Alabama, dự báo chỉ số chứng khoán Đài Loan và xét tính hiệu quả của thuật toán.

2. MỘT SỐ KHÁI NIỆM

Trong phần này, các khái niệm về mô hình chuỗi thời gian mờ phương được Song và Chissom [1 - 3] phát triển và được Chen [4] cải tiến để xây dựng mô hình dự báo cho chuỗi thời gian được trình bày tóm tắt.

Một số định nghĩa sau liên quan đến chuỗi thời gian mờ [4].

Định nghĩa 1: $Y(t)$ ($t = \dots, 0, 1, 2, \dots$) là một tập con của R^I . $Y(t)$ 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: Tại các thời điểm t và $t-1$ có tồn tại một mối quan hệ mờ giữa $F(t)$ và $F(t-1)$ sao cho $F(t) = F(t-1) * R(t-1, t)$ trong đó $*$ là kí hiệu của một toán tử xác định trên tập mờ. $R(t-1, t)$ là *mối quan hệ mờ*. Ta cũng có thể kí hiệu mối quan hệ mờ giữa $F(t)$ và $F(t-1)$ bằng $F(t-1) \rightarrow F(t)$.

Nếu đặt $F(t-1) = A_i$ và $F(t) = A_j$ thì ta kí hiệu *mối quan hệ logic mờ* giữa chúng như sau: $A_i \rightarrow A_j$. Viết như thế này có thể hiểu là tập mờ A_j được suy ra từ A_i .

Định nghĩa 3: Nhóm các *mối quan hệ mờ* theo Chen.

Các *mối quan hệ logic* có thể gộp lại thành một nhóm nếu trong kí hiệu trên nếu các quan hệ logic có cùng một vế trái và có vế phải khác nhau. Thí dụ nếu ta có các *mối quan hệ*:

$$A_i \rightarrow A_k$$

$$A_i \rightarrow A_m$$

thì chúng có thể gộp chúng thành *nhóm các mối quan hệ logic mờ* sau:

$$A_i \rightarrow A_k, A_m$$

Định nghĩa 4: Nhóm các *mối quan hệ logic mờ* có lập theo Yu [10].

Nếu ta có các *mối quan hệ* :

$$A_i \rightarrow A_k; A_i \rightarrow A_m; A_i \rightarrow A_k$$

Thì nhóm quan hệ logic mờ theo Yu sẽ được định nghĩa như sau:

$$A_i \rightarrow A_k, A_m, A_k$$

Nhóm quan hệ logic mờ theo Yu khác định nghĩa của Chen là trong vế phải tập mờ A_k được lặp lại hai lần.

3. MỘT SỐ THUẬT TOÁN TRONG MÔ HÌNH CHUỖI THỜI GIAN MỜ

Thuật toán của Song và Chissom khá phức tạp vì phải tính giá trị max-min trong *mối quan hệ mờ*. Chen đã có một số cải tiến nên để tính *mối quan hệ mờ* chỉ cần sử dụng các phép tính số học đơn giản.

Thuật toán của Chen [4] cải tiến thuật toán của Song-Chissom bao gồm một số bước sau:

1. Xác định tập U bao gồm khoảng giá trị của chuỗi thời gian. Khoảng này xác định từ giá trị nhỏ nhất đến giá trị lớn nhất có thể của chuỗi thời gian.
2. Chia khoảng giá trị và xác định các tập mờ trên tập U . Vấn đề độ dài của khoảng chưa đặt ra và số lượng khoảng lấy bất kì.
3. Mờ hoá các dữ liệu chuỗi thời gian
4. Thiết lập các *mối quan hệ logic mờ*, nhóm quan hệ logic mờ như Định nghĩa 3.
5. Dự báo và giải mờ. Trong bước dự báo chuỗi thời gian mờ được thực hiện như sau:

Trường hợp 1: Nếu $A_j \rightarrow A_i$ và giá trị hàm thuộc của A_i đạt giá trị *maximum* tại đoạn u_i và điểm giữa của u_i là m_i thì dự báo của chuỗi thời gian tại thời điểm i là m_i .

Trường hợp 2: Nếu ta có nhóm quan hệ logic mờ sau:

$$A_i \rightarrow A_{j1}, A_{j2}, \dots, A_{jp} \quad \text{thì giá trị dự báo sẽ là } A_{i1}, A_{i2}, A_{i1}, \dots, A_{jp}$$

Nếu $m_{j1}, m_{j2}, \dots, m_{jp}$ điểm giữa của các đoạn u_i , khi đó giải mờ giá trị dự báo sẽ là:

$$\frac{m_{j1} + m_{j2} + \dots + m_{jp}}{p}$$

Trường hợp 3: Nếu về phải của mỗi quan hệ mờ là trống $A_i \rightarrow \emptyset$ thì giá trị dự báo sẽ là A_i và giải mờ giá trị này sẽ là trung điểm m_i của đoạn u_i .

Yu [10] đã xây dựng mô hình chuỗi thời gian mờ có trọng đề xử lý sự lặp lại các tập mờ xuất hiện trong về phải của nhóm quan hệ mờ. Đối với thứ tự xuất hiện của các tập mờ trong nhóm quan hệ logic mờ ta gán chúng với trọng số khác nhau. Phương pháp này trong đa số các trường hợp cho độ chính xác cao hơn. Dưới đây mô tả thuật toán của Yu trong mô hình chuỗi thời gian mờ bậc nhất.

Bước 1: Xác định tập U bao gồm khoảng giá trị của chuỗi thời gian. Khoảng này xác định từ giá trị nhỏ nhất đến giá trị lớn nhất có thể của chuỗi thời gian và chia khoảng này thành các đoạn để xác định tập các biến ngôn ngữ.

Bước 2: Xác định các tập mờ xác định trên các biến ngôn ngữ trên và mờ hoá các giá trị lịch sử.

Bước 3: Thiết lập mỗi quan hệ mờ và nhóm quan hệ mờ. Trong nhóm quan hệ mờ thiết lập toàn bộ lịch sử xuất hiện các tập mờ có trong về phải của mỗi quan hệ logic mờ theo thứ tự xuất hiện. Thí dụ nếu có các quan hệ mờ sau: $A_1 \rightarrow A_2$, $A_1 \rightarrow A_1$, $A_1 \rightarrow A_1$, $A_1 \rightarrow A_3$, $A_1 \rightarrow A_1$ thì nhóm quan hệ logic mờ có dạng $A_1 \rightarrow A_2, A_1, A_1, A_3, A_1$

Bước 4 : Dự báo như thuật toán của Chen theo các luật khác nhau.

Bước 5 : Nếu xảy ra các trường hợp như các Trường hợp 1 và 3 của thuật toán Chen thì phân giải mờ được giữ nguyên. Còn rơi vào Trường hợp 2 có xuất hiện nhóm các quan hệ logic mờ $A_i \rightarrow A_{i1}, A_{i2} \dots A_{ip}$, và $m_{i1}, m_{i2}, \dots, m_{ik}$ là điểm giữa của các đoạn tương ứng với các biến ngôn ngữ $u_{i1}, u_{i2}, \dots, u_{ik}$ ta sẽ gán các trọng 1, 2, ..., k khi giải mờ giá trị dự báo A_i theo công thức sau:

$$\frac{1 \times m_{i1} + 2 \times m_{i2} + \dots + k \times m_{ik}}{1 + 2 + \dots + k}$$

4. THUẬT TOÁN CẢI BIẾN

Nhận thấy rằng trong **Định nghĩa 3** nhóm quan hệ logic mờ không xác định thời điểm xuất hiện của mỗi phần tử A_i . Chính vì vậy khi tạo nhóm quan hệ logic mờ dạng $A_i \rightarrow A_{i1}, A_{i2} \dots A_{ip}$, thì các thành phần của về phải xuất hiện tại nhiều thời điểm khác nhau. Nếu ta có mỗi quan hệ $F(t-1) \rightarrow F(t)$ và ứng với $F(t-1)$ và $F(t)$ là các tập mờ $A_i(t-1)$, $A_k(t)$ thì ta sẽ có mỗi quan hệ $A_i(t-1) \rightarrow A_k(t)$. Về tổng thể nhóm quan hệ logic mờ phải được viết $A_i(t-1) \rightarrow A_{i1}(t1), A_{i2}(t2), \dots, A_{ip}(tp)$.

Từ đây có thể định nghĩa nhóm quan hệ logic mờ phụ thuộc thời gian như sau..

Định nghĩa 5: (Nhóm quan hệ logic mờ phụ thuộc thời gian)

Mỗi quan hệ mờ ta đều xác định từ quan hệ $F(t-1) \rightarrow F(t)$. Nếu như trên ta đặt $F(t) = A_i(t)$ và $F(t-1) = A_j(t-1)$ thì ta có mỗi quan hệ $A_j(t-1) \rightarrow A_i(t)$. Nếu tại thời điểm t ta có các mỗi quan hệ mờ : $A_j(t-1) \rightarrow A_i(t), A_j(t1-1) \rightarrow A_{i2}(t1), \dots, A_j(tk-1) \rightarrow A_{ip}(tk)$ thì nhóm quan hệ logic mờ

$$A_j(t-1) \rightarrow A_i(t), A_{i1}(t1), A_{i2}(t2), \dots, A_{ip}(tp)$$

với các giá trị $t1, t2, \dots, tp \leq t$ được gọi là nhóm quan hệ mờ phụ thuộc thời gian.

Có thể thấy rằng trong về phải của nhóm quan hệ logic mờ trên chỉ giữ lại các tập mờ có thời điểm xuất hiện trước thời điểm t .

Từ định nghĩa nhóm quan hệ logic này, chúng tôi đưa ra thuật toán dựa trên thuật toán chuỗi thời gian mờ có trọng số của Yu.

1. Xác định tập nền. Tập nền U được xác định như sau: lấy giá trị lớn nhất $f_{\max}$ và nhỏ nhất $f_{\min}$ của chuỗi thời gian và $U = [f_{\min}, f_{\max}]$ trong đó f_1, f_2 là những giá trị dương nào đó. Chia đoạn U thành m khoảng con bằng nhau $u_1, u_2, \dots, u_m$.
2. Xây dựng các tập mờ A_i tương ứng với các khoảng con như trong bước 2 và sử dụng các hàm thuộc tam giác cho mỗi khoảng con của phép chia và mờ hoá các giá trị chuỗi thời gian.
3. Xây dựng mối quan hệ mờ và xác định nhóm các quan hệ logic mờ theo Định nghĩa 5.
4. Dự báo chuỗi thời gian mờ theo các luật sau:

Luật 1: Nếu nhóm quan hệ mờ $A_i \rightarrow \emptyset$ thì giá trị dự báo mờ tại thời điểm t sẽ là A_i

Luật 2: Nếu nhóm quan hệ logic mờ có dạng $A_i \rightarrow A_k$ giá trị dự báo mờ tại thời điểm t sẽ là A_k

Luật 3: Nếu nhóm mỗi quan hệ mờ phụ thuộc thời gian có dạng $A_i \rightarrow A_{i1}, A_{i2}, \dots, A_{ip}$, thì giá trị dự báo sẽ là: $A_{i1}, A_{i2}, \dots, A_{ip}$

5. Giải mờ dựa vào các luật dự báo:

Luật 1: Nếu nhóm quan hệ mờ của là rỗng khi đó giá trị dự báo của $F(t)$ là giá trị A_i và giải mờ sẽ là điểm giữa của khoảng u_i

$$forecast = m_i$$

Luật 2: Nếu nhóm quan hệ logic mờ có dạng $A_i \rightarrow A_k$ và nếu điểm giữa của khoảng u_k là m_k thì

$$forecast = m_k$$

Luật 3: Nếu mỗi quan hệ mờ bậc cao có dạng $A_{i2} \rightarrow A_{i1}, A_{i2}, \dots, A_{ip}$, thì giá trị dự báo sẽ là:

$$forecast = \frac{1 \times m_{i1} + 2 \times m_{i2} + \dots + k \times m_{ik}}{1 + 2 + \dots + k}$$

với $m_{i1}, m_{i2}, \dots, m_{ip}$ là điểm giữa của các đoạn tương ứng.

5. VÍ DỤ

5.1. Dự báo số lượng sinh viên nhập học

Để xem xét tính hiệu quả của định nghĩa mới về nhóm quan hệ logic mờ, chuỗi dữ liệu về số lượng học sinh nhập học của Trường đại học Alabama được sử dụng. Đây là thí dụ mẫu được đưa ra trong bài báo của Song và Chissom [2 - 3] và được nhiều tác giả sử dụng để so sánh.

Bảng 1. Số lượng sinh viên nhập học.

Năm	Số sinh viên	Năm	Số sinh viên
1971	13055	1982	15433
1972	13563	1983	15497

1973	13867	1984	15145
1974	14696	1985	15163
1975	15460	1986	15984
1976	15311	1987	16859
1977	15603	1988	18150
1978	15861	1989	18970
1979	16807	1990	19328
1980	16919	1991	19337
1981	16388	1992	18876

Thuật toán cải tiến cho chuỗi thời gian mờ bao gồm các bước sau đây và áp dụng cho số liệu tại bảng trên.

Bước 1: Xây dựng tập nền U . Xác định giá trị lớn nhất và nhỏ nhất của chuỗi thời gian trên là 19337 và 13055 sinh viên. Do vậy tập nền U được xác định là giá trị trong khoảng $[13000, 20000]$. Tập U được chia thành 7 khoảng $u_1, u_2, \dots, u_7$ với độ rộng là 1000 như trong [4], như vậy các khoảng sẽ là: $u_1 = [13000, 14000], u_2 = [14000, 15000], \dots, u_7 = [19000, 20000]$.

Bước 2: Xây dựng các tập mờ xác định trên các biến ngôn ngữ là các khoảng đã chia

Trong bước này ta xác định lại các tập mờ A_i tương ứng với từng khoảng và có thể gán lại các giá trị ngôn ngữ cho từng tập mờ này. Các tập mờ $A_i, i = 1, 2, \dots, 7$ được định nghĩa như sau:

$$A_1 = 1/u_1 + 0.5/u_2 + 0/u_3 + \dots + 0/u_6 + 0/u_7$$

$$A_2 = 0.5/u_1 + 1/u_2 + 0.5/u_3 + \dots + 0/u_6 + 0/u_7$$

.....

$$A_6 = 0/u_1 + 0/u_2 + \dots + 0.5/u_5 + 1/u_6 + 0.5/u_7$$

$$A_7 = 0/u_1 + 0/u_2 + \dots + 0/u_5 + 0.5/u_6 + 1/u_7$$

Bước 3: Xác định mối quan hệ mờ và nhóm quan hệ mờ phụ thuộc thời gian

Xác định các mối quan hệ mờ theo thời gian. Các mối quan hệ mờ đầu tiên với $t = 1, 2, \dots, 7$ có thể viết như sau: $A_1(1) \rightarrow A_1(2), A_1(2) \rightarrow A_1(3), A_1(3) \rightarrow A_2(4), \dots, A_7(6) \rightarrow A_6(7)$. Tương tự có lập các mối quan hệ mờ cho các thời điểm tiếp sau cho đến $t=22$.

Sau đó lập các nhóm quan hệ mờ theo Định nghĩa 7 ở phần trên. Thí dụ ta có thể nhận được một nhóm quan hệ mờ liên quan đến về trái A_1 nhưng tại thời điểm khác nhau $t = 7, t = 8, t = 9$ ta lại có nhóm quan hệ logic mờ khác nhau: $A_1(7) \rightarrow A_3, A_3, A_1(8) \rightarrow A_3, A_3, A_3$; $A_1(9) \rightarrow A_3, A_3, A_3, A_4$. Toàn thể các nhóm quan hệ mờ sẽ được thể hiện dưới Bảng 2.

Bảng 2. Các nhóm mối quan hệ mờ.

Giá trị	Thời điểm	Giá trị mờ	Nhóm QH mờ Chen	Nhóm QHLG mờ Yu	Nhóm QH logic mờ mới
13055	t = 1	A1			
13563	t = 2	A1	A1,A2	A1,A1,A2	A1
13867	t = 3	A1	A1,A2	A1,A1,A2	A1,A1
14696	t = 4	A2	A1,A2	A1,A1,A2	A1,A1,A2
15460	t = 5	A3	A3	A3	A3
15311	t = 6	A3	A3,A4	A3,A3,A3,A4,A3,A3,A3,A3,A4	A3
15603	t = 7	A3	A3,A4	A3,A3,A3,A4,A3,A3,A3,A3,A4	A3,A3
15861	t = 8	A3	A3,A4	A3,A3,A3,A4,A3,A3,A3,A3,A4	A3,A3,A3
16807	t = 9	A4	A3,A4	A3,A3,A3,A4,A3,A3,A3,A3,A4	A3,A3,A3,A4
16919	t = 10	A4	A3,A4,A6	A4,A4,A3,A6	A4
16388	t = 11	A4	A3,A4,A6	A4,A4,A3,A6	A4,A4
15433	t = 12	A3	A3,A4,A6	A4,A4,A3,A6	A4,A4,A3
15497	t = 13	A3	A3,A4	A3,A3,A3,A4,A3,A3,A3,A3,A4	A3,A3,A3,A4,A3
15145	t = 14	A3	A3,A4	A3,A3,A3,A4,A3,A3,A3,A3,A4	A3,A3,A3,A4,A3,A3
15163	t = 15	A3	A3,A4	A3,A3,A3,A4,A3,A3,A3,A3,A4	A3,A3,A3,A4,A3,A3,A3
15984	t = 16	A3	A3,A4	A3,A3,A3,A4,A3,A3,A3,A3,A4	A3,A3,A3,A4,A3,A3,A3,A3
16859	t = 17	A4	A3,A4	A3,A3,A3,A4,A3,A3,A3,A3,A4	A3,A3,A3,A4,A3,A3,A3,A3,A3
18150	t = 18	A6	A3,A6	A4,A4,A3,A6	A4,A4,A3,A6
18970	t = 19	A6	A6,A7	A6,A7	A6
19328	t = 20	A7	A6,A7	A6,A7	A6,A7
19337	t = 21	A7	A6,A7	A7,A6	A7
18876	t = 22	A6	A6,A7	A7,A6	A7,A6

Từ bảng trên, có thể thấy nhóm các quan hệ mờ phụ thuộc thời gian khác với nhóm quan hệ mờ được xác định theo phương pháp của Chen và Yu.

Bước 4, 5: Dự báo và giải mờ theo các luật đã mô tả ở trên có tính đến trọng số. Kết quả tính toán của phương pháp cải tiến và các phương pháp khác được đưa ra trong Bảng 3 dưới đây:

Bảng 3. Kết quả dự báo của các phương pháp khác nhau.

Năm	Số lượng SV	Phương pháp Chen	Phương pháp Yu	Cải tiến
1971	13055			
1972	13563	14000	14000	13500
1973	13867	14000	14000	13500
1974	14696	14000	14000	14000
1975	15460	15500	15500	15500
1976	15311	16000	15789	15500

1977	15603	16000	15789	15500
Năm	Số lượng SV	P/pháp Chen	P/pháp Yu	Cải tiến
1978	15861	16000	15789	15500
1979	16807	16000	15789	15900
1980	16919	16833	17000	16500
1981	16388	16833	17000	16500
1982	15433	16833	17000	16000
1983	15497	16000	15789	15767
1984	15145	16000	15789	15690,5
1985	15163	16000	15789	15643
1986	15984	16000	15789	15611
1987	16859	16000	15789	15789
1988	18150	16833	17000	17000
1989	18970	19000	19167	18500
1990	19328	19000	19167	19167
1991	19337	19000	18833	19500
1992	18876	19000	18833	18833
MSE		407507,3	407321,5	267438,4

Để so sánh các kết quả dự báo theo các phương pháp khác nhau, ta sử dụng sai số trung bình bình phương MSE theo công thức:

$$MSE = \frac{\sum_{i=1}^n (f_i - g_i)^2}{n}$$

trong đó f_i là giá trị thực còn g_i là giá trị dự báo.

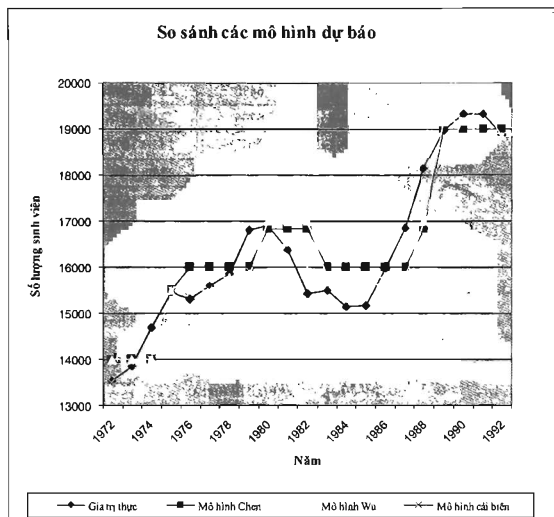
Kết quả sai số theo các phương pháp được đưa ra trong Bảng 4.

Bảng 4 So sánh hiệu quả thuật toán.

Algorithms/MSE	Thuật toán Chen	Thuật toán Yu	Thuật toán cải tiến
MSE	407507,3	407321,5	267438,4

Kết quả tính toán cho thấy trong trường hợp đơn giản của mô hình bậc 1 chúng ta đã thu được sai số MSE chỉ bằng nửa so với thuật toán cơ bản của Chen trong khi thuật toán có trọng của Yu cho sai số tương đương với thuật toán của Chen.

Hình 1 vẽ dưới đây so sánh kết quả tính toán theo phương pháp cải tiến và phương pháp của Chen và Yu. Có thể nhận thấy đồ thị của phương pháp cải tiến phản ánh xu thế tốt hơn so với hai phương pháp Chen và Yu.



Hình 1. Đồ thị kết quả dự báo số sinh viên nhập học.

5.2. Dự báo cho chỉ số chứng khoán Đài Loan

Áp dụng thuật toán cải tiến trên để dự báo cho chuỗi dữ liệu chỉ số thị trường chứng khoán Đài Loan TAIEX mà cũng được nhiều tác giả sử dụng để kiểm định thuật toán Số liệu được đưa ra trong Bảng 5.

Bảng 5. Giá trị chỉ số chứng khoán Đài Loan.

NgàyThang	GiáTriThuc	NgàyThang	GiáTriThuc	NgàyThang	GiáTriThuc
3/8/1998	7552	24/08/1998	6955	11/9/1998	6726,5
4/8/1998	7560	25/08/1998	6949	14/09/1998	6774,55
5/8/1998	7487	26/08/1998	6790	15/09/1998	6762
6/8/1998	7462	27/08/1998	6835	16/09/1998	6952,75
7/8/1998	7515	28/08/1998	6695	17/09/1998	6906

10/8/1998	7365	29/08/1998	6728	18/09/1998	6842
11/8/1998	7360	31/08/1998	6566	19/08/1998	7039
12/8/1998	7320	1/9/1998	6409	21/09/1998	6861
13/08/1998	7291	2/9/1998	6430	22/09/1998	6926
14/08/1998	7320	3/9/1998	6200	23/09/1998	6852
15/08/1998	7300	4/9/1998	6403,2	24/09/1998	6890
17/08/1998	7219	5/9/1998	6697,5	25/09/1998	6871
18/08/1998	7220	7/9/1998	6722,3	28/09/1998	6840
19/08/1998	7285	8/9/1998	6859,4	29/09/1998	6806
20/08/1998	7274	9/9/1998	6769,6	30/09/1998	6787
21/08/1998	7225	10/9/1998	6709,75		

Tập nền U được xác định $U = [6200, 7600]$. Chia tập nền thành 14 khoảng mỗi khoảng có độ dài 100. Từ đây xác định các tập mờ:

$$A_1 = 1/u_1 + 0.5/u_2 + 0/u_3 + \dots + 0/u_{13} + 0/u_{14}$$

$$A_2 = 0.5/u_1 + 1/u_2 + 0.5/u_3 + \dots + 0/u_{13} + 0/u_{14}$$

$$A_{13} = 0/u_1 + 0/u_2 + \dots + 0.5/u_{12} + 1/u_{13} + 0.5/u_{14}$$

$$A_{14} = 0/u_1 + 0/u_2 + \dots + 0/u_{12} + 0.5/u_{13} + 1/u_{14}$$

Sử dụng các bước trong thuật toán cải biên có thể thu được kết quả như trong Bảng 6.

Bảng 6. Kết quả dự báo của các phương pháp khác nhau.

Ngày tháng	Thời điểm	Giá trị thực	Giá trị mờ	Chen method	Yu method	Cải tiến
8/3/1998	t = 1	7552	A14			
8/4/1998	t = 2	7560	A14	7450	7416,67	7550
8/5/1998	t = 3	7487	A13	7450	7416,67	7483,33
8/6/1998	t = 4	7462	A13	7500	7516,67	7450
8/7/1998	t = 5	7515	A14	7500	7516,67	7516,67
8/10/1998	t = 6	7365	A12	7450	7416,67	7416,67
8/11/1998	t = 7	7360	A12	7300	7296,67	7350
8/12/1998	t = 8	7320	A12	7300	7296,67	7350
8/13/1998	t = 9	7291	A11	7300	7296,67	7300
8/14/1998	t = 10	7320	A12	7183.33	7169,05	7350

Nhóm quan hệ logic mờ phụ thuộc thời gian trong mô hình chuỗi thời gian mờ

8/15/1998	t = 11	7300	A12	7300	7296,67	7320
8/17/1998	t = 12	7219	A11	7300	7296,67	7296,67
8/18/1998	t = 13	7220	A11	7183.33	7169,05	7283,33
8/19/1998	t = 14	7285	A11	7183.33	7169,05	7266,67
8/20/1998	t = 15	7274	A11	7183.33	7169,05	7260
8/21/1998	t = 16	7225	A11	7183.33	7169,05	7256,67
8/24/1998	t = 17	6955	A8	7183.33	7169,05	7169,05
8/25/1998	t = 18	6949	A8	6850	6863,33	6950
8/26/1998	t = 19	6790	A6	6850	6863,33	6816,67
8/27/1998	t=20	6835	A7	6775	6794,44	6850
8/28/1998	t = 21	6695	A5	6850	6843,33	6650
8/29/1998	t = 22	6728	A6	6750	6750	6750
8/31/1998	t = 23	6566	A4	6775	6794,44	6650
9/1/1998	t=24	6409	A3	6450	6450	6450
9/2/1998	t = 25	6430	A3	6450	6483,33	6450
9/3/1998	t = 26	6200	A1	6450	6483,33	6316,67
9/4/1998	t = 27	6403.2	A3	6450	6450	6450
9/5/1998	t = 28	6697.5	A5	6450	6483,33	6483,33
9/7/1998	t = 29	6722.3	A6	6750	6750	6750
9/8/1998	t = 30	6859.4	A7	6775	6794,44	6750
9/9/1998	t = 31	6769.6	A6	6850	6843,33	6716,67
9/10/1998	t = 32	6709.75	A6	6775	6794,44	6750
9/11/1998	t = 33	6726.5	A6	6775	6794,44	6750
9/14/1998	t = 34	6774.55	A6	6775	6794,44	6750
9/15/1998	t = 35	6762	A6	6775	6794,44	6750
9/16/1998	t = 36	6952.75	A8	6775	6794,44	6794,44
9/17/1998	t = 37	6906	A8	6850	6863,33	6883,33
9/18/1998	t = 38	6842	A7	6850	6863,33	6870
8/19/1998	t = 39	7039	A9	6850	6843,33	6883,33
9/21/1998	t = 40	6861	A7	6850	6850	6850
9/22/1998	t = 41	6926	A8	6850	6843,33	6910
9/23/1998	t = 42	6852	A7	6850	6863,33	6863,33
9/24/1998	t = 43	6890	A7	6850	6843,33	6863,33

9/25/1998	t = 44	6871	A7	6850	6843,33	6878,57
9/28/1998	t = 45	6840	A7	6850	6843,33	6871,43
9/29/1998	t = 46	6806	A7	6850	6843,33	6866,67
9/30/1998	t=47	6787	A6	6850	6843,33	6843,33
MSE				9658,10	10138,09	4676,51

Để so sánh các kết quả dự báo theo các phương pháp khác nhau, sai số trung bình bình phương MSE theo công thức được sử dụng:

$$MSE = \frac{\sum_{i=1}^n (f_i - g_i)^2}{n}$$

trong đó f_i là giá trị thực còn g_i là giá trị dự báo.

Kết quả sai số theo các phương pháp được đưa ra trong Bảng 7.

Bảng 7. So sánh hiệu quả thuật toán.

Algorithms/MSE	Thuật toán Chen	Thuật toán Yu	Thuật toán cải biên
MSE	9658.10	10138.09	4676,51

Một lần nữa kết quả này cho thấy hiệu quả của thuật toán đề xuất chỉ gần bằng nửa của thuật toán Chen trong khi thuật toán có trọng của Yu lại cho kết quả tồi hơn thuật toán của Chen.

6. KẾT LUẬN

Bài báo này đưa ra một cải biên mới để sử dụng được trong mô hình chuỗi thời gian mờ. Tương tự như cải biên của Yu khi xây dựng nhóm quan hệ logic mờ đã tính đến sự lặp lại của các giá trị trùng nhau bên về phải và gán trọng khác nhau cho từng vị trí của giá trị đó, chúng tôi xét thời điểm xuất hiện của từng giá trị về phải mỗi quan hệ logic mờ. Như vậy tại từng thời điểm, nhóm quan hệ logic mờ đối với về trái giống nhau nhưng lại khác nhau ở về phải. Với định nghĩa mới này về nhóm quan hệ logic mờ phụ thuộc thời gian, chưa cần sử dụng các phương pháp nâng cao độ chính xác khác như phân đoạn lại, sử dụng chuỗi thời gian mờ bậc cao hay mô hình hai nhân tố [5 - 9], kết quả đã tốt hơn rất nhiều so với thuật toán cơ bản của Chen.

Nhóm các quan hệ logic mờ là khái niệm cơ bản để cải tiến các thuật toán trong mô hình chuỗi thời gian mờ. Chúng được sử dụng trong hầu hết các công trình sau này của các tác giả khác nhau. Chính vì vậy, sử dụng nhóm quan hệ mờ mới này trong các phương pháp cải tiến khác nhau hi vọng sẽ làm tăng hiệu quả của các thuật toán này.

TÀI LIỆU THAM KHẢO

1. Song Q., Chissom B. S. - Fuzzy Time Series and its Mode, Fuzzy set and systems 54 (1993) 269-277.
2. Song Q., Chissom B.S. - Forecasting Enrollments with Fuzzy Time Series – Part I, Fuzzy set and systems 54 (1993) 1-9.
3. Song Q., Chissom B. S. - Forecasting Enrollments with Fuzzy Time Series – Part II, Fuzzy set and systems 62 (1994) 1-8.
4. Chen S. M. - Forecasting Enrollments based on Fuzzy Time Series, Fuzzy set and systems 81 (1996) 311-319.
5. Chen S. M. - Forecasting Enrollments based on high-order Fuzzy Time Series, Int. Journal: Cybernetic and Systems 33 (2002) 1-16.
6. Huang K. - Heuristic models of fuzzy time series forecasting, Fuzzy set and systems 123 (2001) 369-386.
7. 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 (2009) 6108–6117.
8. Lee L. W., Wang L. H., S. Chen S. M., Leu H. C. - Handling forecasting problem based on two-factors high-order fuzzy time series, IEEE Transactions on Fuzzy Systems 14 (3) (2006) 468–477.
9. Nguyễn Công Diêu - Một thuật toán mới cho mô hình chuỗi thời gian mờ heuristic trong dự báo chứng khoán, Tạp chí Khoa học và Công nghệ 49 (4) (2011) 11-25.
10. Yu H. K. - Weighted fuzzy time series models for TAIEX forecasting, Physica A 349 (2005) 609–624.
11. Singh P., Borah B. - Forecasting stock index price based on M-factors fuzzy time series and particle swarm optimization, Int. J. of Approximation Reasoning 55 (2014) 812–833.
12. Chen M. U. - A high order fuzzy time series forecasting model for Internet stock trading, Int. J. Future Generation Computer Generation. Doi/10.10.16/2013. 09.025 (2013).

ABSTRACT

FUZZY TIME-DEPENDING LOGICAL RELATIONSHIP GROUPS IN FUZZY TIME SERIES MODELS

Nguyen Cong Dieu^{1,2}

¹*Institute of Information Technology, VAST, 18 Hoang Quoc Viet, Cau Giay, Hanoi*

²*Thang Long University, Dai Kim, Hoang Mai, Hanoi*

Email: ncdieu@joit.ac.vn

Fuzzy time series models have many applications in forecasting, especially in the economic forecast. In recent years many works have been completed towards improving accuracy and reducing the amount calculated in fuzzy time series models such as the article by

Chen and Hsu, Huang, Kuo, Wu. A different approach to improve efficiency for time series prediction is to use fuzzy techniques in data mining such as clustering, neural networks, ... to build the model. The most of methods are based on the technique of Chen's fuzzy logic relationship groups to reduce the amount of computation to just perform arithmetic calculations instead of min-max as in the model of Song-Chissom. Yu (2005) developed the recurrent fuzzy relationships and constructed the weighted fuzzy time series model.

In this paper, we propose a modified way to define fuzzy time-depending logical relationship groups. Thanks to the new concept of fuzzy relationship groups, the new fuzzy time series model is proposed. Using this model for forecasting fuzzy time series model, we obtained better results for enrollments and Taiwan stock index forecasting than Chen and Yu results.

Keywords: fuzzy time series, weighted fuzzy time series model, fuzzy logical relationship groups.