

DỰ BÁO CHUỖI THỜI GIAN MỜ DỰA TRÊN NHÓM QUAN HỆ MỜ PHỤ THUỘC THỜI GIAN VÀ TỐI ƯU BẦY ĐÀN

Nguyễn Công Điều¹, Nghiêm Văn Tính²

¹Khoa Toán-Tin, Trường Đại học Thăng Long,

² Trường Đại học Kỹ thuật Công nghiệp, Đại học Thái Nguyên
ncdiu@yahoo.com, nghiemvantinh@tmut.edu.vn

TÓM TẮT— Trong thời gian gần đây, mô hình chuỗi thời gian mờ đang thu hút sự chú ý của các nhà nghiên cứu và phân tích số liệu. Từ mô hình ban đầu của Song và Chissom, đến nay ngày càng nhiều mô hình chuỗi thời gian mờ được đề xuất để nâng cao độ chính xác trong dự báo. Tuy nhiên, vẫn tồn tại một số vấn đề chưa được giải quyết một cách tối ưu trong mô hình chuỗi thời gian mờ. Đó là, làm thế nào để phân chia tập nền thành các khoảng có độ dài thích hợp và xây dựng các quan hệ mờ, nhóm quan hệ mờ một cách có hiệu quả. Sự kết hợp của các phương pháp tối ưu như giải thuật di truyền, kỹ thuật mô phỏng tối luyện, tối ưu đàn kiến hay tối ưu bầy đàn,... nhằm xác định một cách tối nhất các khoảng chia đã được đề cập đến. Trong bài báo này, mô hình chuỗi thời gian mờ dựa trên một khái niệm mới được đề xuất là nhóm quan hệ mờ phụ thuộc thời gian và kỹ thuật tối ưu bầy đàn được phát triển để điều chỉnh độ dài khoảng chia tập nền nhằm tăng độ chính xác dự báo của mô hình. Kết quả thực nghiệm cho thấy mô hình mới này cho độ chính xác dự báo tốt hơn so với một số phương pháp đã đề xuất trước đây.

Từ khóa— Dự báo, chuỗi thời gian mờ, nhóm quan hệ mờ phụ thuộc thời gian, tối ưu bầy đàn.

I. MỞ ĐẦU

Nhiều mô hình đã được đề xuất nhằm giải quyết các bài toán dự báo khác nhau để giúp con người đưa ra các quyết định chẳng hạn như: dự báo tuyển sinh đại học, thị trường chứng khoán, dự báo dân số, dự báo nhiệt độ. Trong một số trường hợp, các phương pháp dự báo truyền thống không giải quyết tốt được đối với chuỗi số liệu biểu diễn bởi giá trị ngữ nghĩa hay với chuỗi số liệu có sự biến thiên mạnh. Vì vậy, Song và Chissom [22-24] đưa ra hai mô hình dự báo chuỗi thời gian mờ phụ thuộc thời gian và mô hình chuỗi thời gian mờ không phụ thuộc thời gian nhằm khắc phục những khó khăn trên và áp dụng để dự báo số lượng sinh viên nhập học của trường Đại học Alabama. Sau công trình này, một loạt các bài báo của nhiều tác giả khác nhau tiếp tục dựa trên ý tưởng này để xây dựng mô hình dự báo và ứng dụng trong nhiều lĩnh vực khác nhau. Chen [3] 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 kết hợp max-min phức tạp trong xử lý mối quan hệ mờ. Công trình này cùng với mô hình chuỗi thời gian mờ bậc cao [4] và mô hình chuỗi thời gian nhiều nhân tố [2], [17] là nền tảng cho sự phát triển mạnh mẽ của mô hình chuỗi thời gian mờ trong những khoảng thời gian tiếp sau. Tuy nhiên có thể thấy, những mô hình chuỗi thời gian mờ trên còn có độ chính xác dự báo chưa cao.

Trong những năm gần đây, nhiều tác giả đã sử dụng các kỹ thuật khác nhau để tìm mô hình hiệu quả cho chuỗi thời gian mờ, chủ yếu là nâng cao độ chính xác của dự báo. Huarng [12] đã phát hiện ra độ chính xác của mô hình dự báo phụ thuộc nhiều vào độ dài khoảng phân chia tập nền và đã đề xuất ra 3 thuật toán xác định độ dài khoảng là phương pháp dựa trên trung bình, độ dài dựa trên phân bố và độ dài dựa trên tỉ lệ. Tiếp sau, một số bài báo liên quan đến vấn đề lựa chọn khoảng tối ưu cũng được thực hiện như trong [9].

Kỹ thuật phân cụm như phân cụm tự động hay phân cụm mờ cũng được một số tác giả sử dụng để phân chia khoảng của tập nền như trong các công trình [1], [7]. Một số tác giả khác lại cố định số lượng khoảng nhưng tìm cách chỉnh lại các điểm chia dựa trên một tiêu chuẩn tối ưu. Các phương pháp tối ưu thường được sử dụng như giải thuật di truyền [5], [18], [19], tối ưu bầy đàn [10], [15-16] hay tìm kiếm Tabu [9]. Trong những công trình mới nhất có thể kết hợp nhiều kỹ thuật khác nhau như tối ưu bầy đàn, phân cụm K – mean và độ đo tương tự để xây dựng mô hình [8].

Đối với các mô hình dự báo hiện tại, có hai yếu tố chính làm ảnh hưởng đến độ chính xác dự báo, đó là độ dài của khoảng chia tập nền và mối quan hệ mờ. Như trên đã nói, việc phân khoảng đã có một số tác giả đề xuất với nhiều kỹ thuật khác nhau. Liên quan đến xây dựng các mối quan hệ mờ chỉ có số ít công trình được thực hiện. Song và Chissom [22-24] đã xây dựng các ma trận quan hệ mờ trên cơ sở các toán tử phức hợp và tính toán dựa trên các phép max - min khá phức tạp. Chen đã cải tiến mô hình khi đưa ra các mối quan hệ mờ và nhóm quan hệ mờ và tính toán các giá trị dự báo chỉ bằng các phép tính đơn giản. Huarng [13] và N.C. Dieu [20] đã xem xét xu hướng của chuỗi thời gian để rút gọn nhóm quan hệ mờ trong mô hình chuỗi thời gian mờ heuristic. Yu [25] đã đưa vào nhóm quan hệ mờ các giá trị ngôn ngữ có tính cả độ lặp của chúng. Huang và các tác giả khác [11] đã xây dựng các mối quan hệ mờ thông qua các ma trận quan hệ mờ có trọng và thực hiện tính toán bằng cả phép tính đơn giản và phép lấy max. Trong báo cáo này, chúng tôi đề xuất khái niệm mới là nhóm quan hệ mờ phụ thuộc thời gian và mô hình chuỗi thời gian mờ mới để dự báo dựa trên sự kết hợp giữa khái niệm nhóm quan hệ mờ phụ thuộc thời gian và thuật toán tối ưu bầy đàn. Các kết quả thực nghiệm cho số liệu sinh viên nhập học của đại học Alabama trên mô hình mới này cho kết quả khá tốt so với những phương pháp khác khi sử dụng mô hình chuỗi thời gian mờ kết hợp với tối ưu bầy đàn hay giải thuật di truyền.

Bài báo này có 4 mục và phần kết luận. Sau phần Mở đầu, mục 2 sẽ trình bày các khái niệm liên quan đến chuỗi thời gian mờ, khái niệm mới đề xuất là nhóm quan hệ mờ phụ thuộc thời gian cùng thuật toán xây dựng, tối ưu bầy đàn

và thuật toán cơ bản. Trong mục 3, mô hình mới dựa trên khái niệm nhóm quan hệ mờ phụ thuộc thời gian và tối ưu bầy đàn được đề xuất. Mục 4 thực hiện các tính toán dự báo số lượng sinh viên nhập học tại Đại học Alabama và so sánh các kết quả dự báo giữa mô hình đề xuất với các mô hình chuỗi thời gian mờ khác kết hợp với giải thuật giải thuật di truyền và tối ưu bầy đàn.

II. CHUỖI THỜI GIAN MỜ VÀ KỸ THUẬT TỐI ƯU BẦY ĐÀN

2.1. Một số khái niệm

Trong phần này sẽ trình bày một số khái niệm và phương pháp dự báo của chuỗi thời gian mờ được Song và Chissom [22]-[24] phát triển và được Chen [3] cải tiến.

Một số định nghĩa sau liên quan đến chuỗi thời gian mờ [3].

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 $R(t-1, t)$ giữa các tập mờ $F(t)$ và $F(t-1)$ sao cho $F(t) = F(t-1) * R(t-1, t)$ trong đó $*$ là ký hiệu của một toán tử xác định trên tập mờ thì có thể nói $F(t)$ được suy ra từ $F(t-1)$. Ta cũng có thể ký hiệu mối quan hệ mờ giữa $F(t)$ và $F(t-1)$ bằng $F(t-1) \rightarrow F(t)$.

Nếu $F(t-1) = A_i$ và $F(t) = A_j$ thì có thể ký hiệu *mối quan hệ logic mờ* giữa chúng như sau: $A_i \rightarrow A_j$ và có thể hiểu là tập mờ A_j được suy ra từ A_i và A_i là về trái còn A_j được gọi là về phải của mối quan hệ logic mờ.

Định nghĩa 3: Giả sử $F(t)$ và $F(t-1)$ là hai tập mờ. Nếu tồn tại một mối quan hệ mờ $R(t-1, t)$ sao cho $F(t) = F(t-1) * R(t-1, t)$ thì ta có thể nói $F(t)$ được suy ra từ $F(t-1)$ và mối quan hệ logic mờ được biểu diễn $F(t-1) \rightarrow F(t)$ trong đó $F(t-1)$ được gọi là trạng thái hiện hành còn $F(t)$ được gọi là trạng thái tiếp. Nếu $R(t-1, t)$ không phụ thuộc vào t thì $F(t)$ được gọi là chuỗi thời gian mờ dừng, còn ngược lại ta có chuỗi thời gian mờ không dừng.

Định nghĩa 4: Giả sử $F(t)$ suy đồng thời từ $F(t-1)$, $F(t-2)$, ..., $F(t-m)$ $m > 0$. Khi đó mối quan hệ mờ có thể viết được $F(t-m), \dots, F(t-2), F(t-1) \rightarrow F(t)$ và gọi đó là mô hình quan hệ logic mờ bậc m .

Định nghĩa 5: Nhóm các mối quan hệ mờ theo Chen.

Khi các mối quan hệ mờ trên có cùng về trái nhưng về phải lại khác nhau thì có thể gộp về phải lại để hình thành nhóm quan hệ logic mờ. 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 6: Nhóm các mối quan hệ logic mờ bậc nhất theo Yu [25] có tính độ lặp và thứ tự về phải

Nếu ta có các mối quan hệ:

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

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

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

Để ý thấy rằng trong nhóm quan hệ mờ này có tính cả các lần lặp của A_k trong thành phần về phải.

2.2. Nhóm quan hệ mờ phụ thuộc thời gian

Tại thời điểm t có thể xác định được mối quan hệ mờ $F(t-1) \rightarrow F(t)$. Đặt $F(t) = A_i$ và $F(t-1) = A_j$ thì mối quan hệ trên được viết thành $A_j \rightarrow A_i$. Để ghi nhớ thời điểm xuất hiện t của mỗi quan hệ thì có thể viết $A_j(t) \rightarrow A_i(t)$ trong đó tham số t trong $A_j(t)$ và $A_i(t)$ chỉ sự xuất hiện của các tập mờ này tại thời điểm t .

Khi các mối quan hệ logic mờ có cùng về trái (không tính thời điểm xuất hiện): $A_j(t) \rightarrow A_i(t)$; $A_j(t1) \rightarrow A_i(t1)$; $A_j(t2) \rightarrow A_i(t2)$, ..., $A_j(tp) \rightarrow A_i(tp)$ thì theo Định nghĩa 6 có thể gộp về phải lại thành nhóm quan hệ mờ tại thời điểm hiện hành t :

$$A_j(t) \rightarrow A_i(t), A_i(t1), A_i(t2), \dots, A_i(tp)$$

Các ký hiệu $t, t1, \dots, tp$ trong ngoặc là chỉ thời điểm xuất hiện của các phần tử A_k trong các mối quan hệ mờ. Theo Định nghĩa nhóm quan hệ logic mờ của Chen và Yu có thể tham gia của thành phần $A_{ik}(t_k)$ nào đó mà thời điểm xuất hiện t_k của A_{ik} lại sau thời điểm có nhóm quan hệ mờ này, tức là $t < t_k$ thì sự tham gia của thành phần này để dự báo cho thời điểm t là vô lý. Do đó cần xây dựng nhóm quan hệ mờ mà các thành phần trong về phải chỉ xuất hiện trước hay

đồng thời tại thời điểm t khi xác định nhóm quan hệ mờ mà thôi. Nhóm quan hệ logic mờ này được gọi là nhóm quan hệ logic mờ phụ thuộc thời gian.

Định nghĩa 7: Nhóm quan hệ logic mờ phụ thuộc thời gian

Tại thời điểm t , nếu tồn tại các mối quan hệ mờ có cùng vế trái: $A_j(t) \rightarrow A_i(t); A_j(t_1) \rightarrow A_{i1}(t_1); A_j(t_2) \rightarrow A_{i2}(t_2), \dots, A_j(t_s) \rightarrow A_{is}(t_s)$ thì có thể gộp vế phải lại thành một nhóm:

$$A_j(t) \rightarrow A_i(t), A_{i1}(t_1), A_{i2}(t_2), \dots, A_{is}(t_s)$$

với các giá trị $t_1, t_2, \dots, t_s \leq t$ (tức là chỉ nhóm các tập mờ xuất hiện trước thời điểm t trong vế trái). Nhóm quan hệ mờ này có tính đến thời điểm xuất hiện của các tập mờ được gọi là *nhóm quan hệ logic mờ phụ thuộc thời gian*.

Chú ý rằng các ký hiệu $A_i(t)$ và $A_i(t_1)$ vẫn cùng là tập mờ A_i nhưng chúng xuất hiện trong mỗi quan hệ mờ tại thời điểm t và t_1 tương ứng.

Tương tự ta có thể định nghĩa được *Nhóm quan hệ mờ bậc cao m* phụ thuộc thời gian được định nghĩa như sau:

Nếu có $F(t)$ được suy ra từ $F(t-1), F(t-2), \dots, F(t-m)$ thì ta có nhóm quan hệ mờ bậc cao $F(t-1), F(t-2), \dots, F(t-m) \rightarrow F(t)$ hay $A_{j1}(t), A_{j2}(t), \dots, A_{jm}(t) \rightarrow A_{ki}(t)$. Ký hiệu $A_{i2}(t)$ để chỉ tập A_{i2} xuất hiện tại thời điểm t .

$$A_{j1}(t_1-m), A_{i2}(t_1-m+1), \dots, A_{jm}(t_1-1) \rightarrow A_{ki}(t_1);$$

.....

$$A_{j1}(t_p-m), A_{i2}(t_p-m+1), \dots, A_{jm}(t_p-1) \rightarrow A_{kp}(t_p)$$

Có thể loại bỏ các tham số thời gian tại vế trái của mỗi quan hệ logic mờ để có thể được viết như sau:

$$A_{j1}, A_{i2}, \dots, A_{jm} \rightarrow A_{ki}(t_1);$$

.....

$$A_{j1}, A_{i2}, \dots, A_{jm} \rightarrow A_{kp}(t_p)$$

Với $t_1 < t_2 < \dots < t_p$ thì nhóm quan hệ mờ bậc cao phụ thuộc thời gian tại thời điểm t_p sẽ được viết như sau:

$$A_{j1}, A_{i2}, \dots, A_{jm} \rightarrow A_{ki}(t_1), \dots, A_{kp}(t_p)$$

Chú ý rằng tuy viết như vậy nhưng $A_j(t_1)$ và $A_j(t_2)$ đều cùng một giá trị ngữ nghĩa A_j nhưng thời điểm xuất hiện của tập mờ này là t_1 và t_2 tương ứng.

Xây dựng nhóm quan hệ mờ phụ thuộc thời gian

Giả sử có chuỗi thời gian mờ $F(t), t=1, 2, \dots, q$ hay thể hiện qua các giá trị ngôn ngữ (tập mờ) như sau $\{A_{j1}, A_{j2}, \dots, A_{jq}\}$.

Muốn tạo các nhóm quan hệ mờ phụ thuộc thời gian, cần thiết phải lần lượt tạo các nhóm quan hệ mờ tại từng thời điểm. Thuật toán được mô tả như sau:

Thuật toán tạo nhóm quan hệ mờ phụ thuộc thời gian bậc m

1. Khởi tạo nhóm quan hệ mờ bậc m với thời điểm $t=m: F(1), F(2), \dots, F(m-1) \rightarrow F(m)$ hay $A_{i2}, \dots, A_{jm} \rightarrow A_{ki}(m)$
 2. Vòng lặp for cho $t=m+1, \dots, q$ do
 - 2.1. Lặp for $k = m, \dots, 1$ do

Xây dựng quan hệ bậc m $A_{i2}(t-m), \dots, A_{jm}(t-1) \rightarrow A_{ki}(t)$

endfor
 - 2.2. Vòng lặp for $l=1, \dots, t-1$ do

Kiểm tra xem đã có quan hệ dạng $A_{i2}, \dots, A_{jm} \rightarrow A_{k2}(l)$ có cùng vế trái chưa. Nếu có cùng vế trái thêm A_{k2} vào nhóm thành quan hệ $A_{i2}, \dots, A_{jm} \rightarrow A_{k1}, A_{k2}$

endfor
 3. Kết thúc vòng lặp for
-

2.3. Một số Thuật toán trong mô hình chuỗi thời gian mờ

Thuật toán đầu tiên do Song và Chissom [1-3] đề xuất khá phức tạp vì phải thực hiện các phép max-min trong các toán tử phức hợp của mỗi quan hệ mờ. Chen đã xác định mỗi quan hệ mờ thông qua nhóm quan hệ mờ nên tính mỗi quan hệ mờ chỉ cần sử dụng các phép tính số học đơn giản.

Thuật toán của Chen [3] 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 kỳ.
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 5*.
5. Dự báo và giải mờ. Trong bước dự báo chuỗi thời gian mờ được thực hiện như sau:
 - *Luật 1*: Nếu $A_j \rightarrow A_i$ và giá trị hàm thuộc của A_j 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 .
 - *Luật 2*: Nếu ta có các mối quan hệ logic mờ hình thành nhóm quan hệ logic mờ sau:
 $A_i \rightarrow A_{j_1}, A_{j_2}, \dots, A_{j_p}$ thì giá trị dự báo sẽ là $A_{i1}, A_{i2}, A_{j1}, \dots, A_{jp}$

Khi đó giải mờ giá trị dự báo sẽ là:

$$\frac{m_{j_1} + m_{j_2} + \dots + m_{j_p}}{p}$$

Trong đó $m_{j_1}, m_{j_2}, \dots, m_{j_p}$ điểm giữa của các đoạn u_i .

- *Luật 3*: Nếu vế phải của mối quan hệ mờ là trống như trường hợp sau:

$$A_i \rightarrow \emptyset$$

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 :

Thuật toán cải tiến có trọng của Yu về cơ bản là giống như của Chen nhưng khác tại bước 4 khi thiết lập nhóm quan hệ logic mờ theo *Định nghĩa 6* còn tại bước giải mờ được thực hiện như sau:

2.4. Tối ưu bầy đàn

Tối ưu bầy đàn (PSO- Particle Swarm Optimization) được Kennedy và Eberhart [14] đề xuất năm 1995 là thuật toán tìm kiếm ngẫu nhiên dựa trên mô phỏng hành vi và sự cộng tác của đàn chim tìm kiếm nguồn thức ăn. Mỗi phần tử của đàn chim có những con đường tìm kiếm riêng của mình và trên mỗi bước lặp con đường tối ưu được xác định và các cá thể đều được hiệu chỉnh theo con đường tối ưu đó. Mỗi phần tử được tạo ngẫu nhiên và sau đó bay theo một đường bay trong không gian tìm kiếm. Tại mỗi bước tối ưu, mỗi phần tử đánh giá vị trí của mình và các vị trí của phần tử lân cận. Mỗi phần tử cần nhớ vị trí tốt nhất của mình cũng như các vị trí tốt nhất của các phần tử khác trong đàn. Để nhận được lời giải tối ưu, mỗi cá thể đều cập nhật vị trí và tốc độ theo các phương trình sau:

$$V_{id} = \omega \times V_{id} + C_1 \times \text{Rand}() \times (P_{id} - X_{id}) + C_2 \times \text{Rand}() \times (P_{gbest} - X_{id}) \quad (1)$$

$$X_{id} = X_{id} + V_{id} \quad (2)$$

Trong những phương trình trên, ω là trọng số quán tính, V_{id} là tốc độ của mỗi phần tử id nằm trong $[-V_{Max}, V_{Max}]$ trong đó V_{Max} là hằng số được chọn trước. Thông thường chọn $V_{Max}=100$. C_1 và C_2 là hệ số tự tin cậy và hệ số xã hội và trong giải thuật chuẩn thì đó là những hằng số. $\text{Rand}()$ là hàm tạo số ngẫu nhiên thực trong khoảng $(0,1)$ với phân bố chuẩn. X_{id} và P_{id} là vị trí hiện thời và vị trí tốt nhất của mỗi phần tử id . Ký hiệu P_{gbest} là vị trí tốt nhất của toàn quần thể theo hàm mục tiêu. Trong bài báo này, chúng tôi sử dụng thuật toán tối ưu bầy đàn chuẩn trong đó ω giảm tuyến tính khi các thế hệ kế tiếp nhau được thực hiện. Thuật toán PSO chuẩn được mô tả như sau:

Thuật toán PSO chuẩn

1. Khởi tạo vị trí ban đầu và tốc độ của toàn bộ phần tử
 2. Vòng lặp while cho đến khi điều kiện lặp được thỏa mãn
 - 2.1. Vòng lặp for cho mỗi phần tử id thực hiện
 - a. Đánh giá hàm mục tiêu
 - b. Cập nhật vị trí địa phương và toàn cục tốt nhất
 - c. Tính toán vận tốc theo công thức (1)
 - d. Cập nhật vị trí theo công thức (2)
 - 2.2. Kết thúc vòng lặp for
 3. Kết thúc vòng while
-

III. MÔ HÌNH ĐỀ XUẤT

3.1. Thuật toán cơ bản dựa trên nhóm quan hệ mờ phụ thuộc thời gian

Trước hết, nhắc lại thuật toán cơ bản tương tự như mô hình của Chen tại mục 2.3 nhưng dựa trên mô hình chuỗi thời gian mờ có trọng của Yu [25].

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 $U = [f_{min}-f_1, f_{max}+f_2]$ trong đó f_1, f_2 là những giá trị dương nào đó. Chia đoạn U thành m khoảng con u_1, u_2, \dots, u_m .
2. Xây dựng các tập mờ A_i tương ứng với các khoảng đã chia.
3. Mờ hóa các giá trị lịch sử của chuỗi thời gian.
4. Xây dựng mối quan hệ mờ và nhóm các quan hệ logic mờ theo Định nghĩa 6.
5. 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 mối 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: $A_{i1}, A_{i2}, \dots, A_{ip}$

6. 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 mối quan hệ mờ bậc cao có dạng $A_{i2} \rightarrow A_{i1}, A_{i2}, \dots, A_{ip}$, thì giá trị dự báo sẽ là:

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

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

3.2. Thuật toán cải tiến với tối ưu bầy đàn

Như đã biết, độ dài các khoảng chia có ảnh hưởng lớn đến độ chính xác của dự báo. Độ dài của mỗi khoảng chia phụ thuộc vào cách xác định bởi các điểm đầu và cuối mỗi khoảng. Do đó, cần phải tìm các điểm chia sao cho tối ưu hàm sai số dự báo. Tối ưu bầy đàn được sử dụng vào mục đích này và hàm mục tiêu được chọn là giá trị hàm MSE .

Giả sử số khoảng chia được xác định là n . Tập nền U là đoạn $[d_0, d_n]$. Tập nền này được chia làm n khoảng với các nút chia là $d_1, d_2, \dots, d_{n-2}, d_{n-1}$. Như vậy mỗi tập nền U được phân thành n khoảng sau: $u_1 = [d_0, d_1], u_2 = [d_1, d_2], \dots, u_n = [d_{n-1}, d_n]$. Mỗi phần tử trong tối ưu bầy đàn được xác định qua một vectơ $n-1$ thành phần $id = [d_1, d_2, \dots, d_{n-2}, d_{n-1}]$. Cần phải tìm phần tử id nào đây (tức là tìm các điểm chia khoảng d_i) sao cho trong quá trình tính toán dự báo có đại lượng MSE là nhỏ nhất.

Trong quá trình tính toán và xác định vị trí mới của từng phần tử trong mỗi bước lặp, các vectơ vị trí này sẽ được cập nhật và để thỏa mãn tính chất chia khoảng, dãy số $d_i, i=1, 2, \dots, n-1$ cần phải sắp xếp theo thứ tự tăng dần. Biểu diễn mỗi phần tử được trình bày trong dạng dưới đây.

d_1	d_2	\dots	d_j	\dots	d_{n-1}
-------	-------	---------	-------	---------	-----------

Hình 1. Cấu trúc một phần tử

Thuật toán dự báo chuỗi thời gian mờ có sử dụng có sử dụng tối ưu bầy đàn được thực hiện như sau. Giả sử bầy đàn gồm m phần tử, mỗi phần tử có cấu trúc vectơ như hình trên. Đối với mỗi phần tử thể hiện một cách chia khoảng của tập nền. Đối với mỗi phương án chia khoảng như vậy, có thể thực hiện được thuật toán dự báo như thuật toán cơ bản ở trên. Đối với mỗi phần tử có thể tính được hàm MSE như công thức (3). Với lần lặp đầu tiên, các giá trị vận tốc V_{id} và vị trí P_{id} cũng như P_{ld} và P_{gbest} cũng được cho trước. Sau mỗi bước lặp với m phần tử các giá trị P_{ld} và P_{gbest} được mỗi phần tử ghi nhớ lại và thực hiện thay đổi vận tốc và vị trí theo các công thức (1) và (2). Sau khi đã có được các giá trị phân khoảng thực hiện giải thuật trong mục 3.1 để tính dự báo và hàm MSE . Phần tử nào có giá trị MSE nhỏ nhất trong mỗi lần lặp được coi là lời giải tốt nhất của bài toán tại bước lặp đó. Sau khi thực hiện được đủ số lần lặp thì cũng có thể đưa ra lời giải với MSE nhỏ nhất.

Thuật toán PSO cho các dữ liệu dự báo được mô tả như sau:

Thuật toán PSO cho dự báo chuỗi thời gian mờ

1. Xác định tập nền U
 2. Khởi tạo vị trí ban đầu $d_1, d_2, \dots, d_{n-2}, d_{n-1}$ qua đó xác định các khoảng u_1, u_2, \dots, u_m và tốc độ của toàn bộ phần tử
 3. Vòng lặp *while* trong PSO cho mỗi thể hệ đến khi điều kiện lặp được thỏa mãn
 - 3.1. Vòng lặp *for* cho mỗi phần tử id do
 - a. Tạo các tập mờ trên cơ sở phân khoảng theo từng phần tử và mờ hóa dữ liệu trên
 - b. Tạo các quan hệ mờ và nhóm quan hệ mờ phụ thuộc thời gian
 - c. Tạo các luật dự báo mờ như trong bước 5 thuật toán cải tiến của mục 3.1
 - d. Giải mờ theo các luật dự báo mờ tại bước 6 thuật toán cải tiến của mục 3.1
 - e. Tính MSE theo công thức (3) cho mỗi phần tử
 - f. Cập nhật vị trí tốt nhất của từng phần tử và vị trí tốt nhất của toàn bộ hệ thống
 - 3.2. *End for*
 - 3.3. Vòng lặp *for* mọi phần tử id do
 - a. Tính toán vận tốc theo công thức (1)
 - b. Cập nhật vị trí $d_1, d_2, \dots, d_{n-2}, d_{n-1}$ theo công thức (2) cho các phần tử
 - 3.4. *End for*
 - 3.5. Lưu lại dự báo kết quả tốt nhất với MSE nhỏ nhất trong tất cả các phần tử
 4. *End while*
-

Như vậy, mô hình chuỗi thời gian đề xuất được thực hiện thông qua sự kết hợp giữa tối ưu bầy đàn để xác định điểm chia tập nền cho mỗi thể hệ và thuật toán dự báo chuỗi thời gian mờ dựa trên nhóm quan hệ mờ phụ thuộc thời gian bậc cao.

Dưới đây, chuỗi dữ liệu về số lượng học sinh nhập học của Trường đại học Alabama [3] được sử dụng để xây dựng các nhóm quan hệ mờ nêu trên. Các tính toán quan hệ mờ và nhóm quan hệ mờ bậc nhất của Chen của Yu và nhóm quan hệ mờ phụ thuộc thời gian được thực hiện để thấy được sự sai khác giữa các khái niệm này.

IV. KẾT QUẢ THỰC NGHIỆM

Để kiểm tra hiệu quả của thuật toán cải tiến, chuỗi số liệu số sinh viên nhập học của trường Alabama từ năm 1971 đến năm 1992 được lấy để kiểm tra (Bảng 1). Xác định tập nền U bằng cách lấy giá trị lớn nhất và nhỏ nhất cộng trừ thêm một lượng nào đó để thành đoạn để phân hoạch. Như phần trên đã trình bày, tập nền được xác định trong khoảng [13000,20000] và số lượng khoảng được chọn là tùy ý. Sử dụng tối ưu bầy đàn để tìm kiếm các mốc chia khoảng tối ưu theo hàm MSE .

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

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:

$$MSE = \frac{\sum_{i=1}^n (f_i - g_i)^2}{n} \tag{3}$$

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

Các thông số được chọn như sau: số lượng phần tử là 30, số bước dịch chuyển (số lần lặp) của các phần tử là 100, giá trị trọng ω giảm từ 1,4 xuống còn 0,4 trong quá trình dịch chuyển các thể hệ lặp. Tốc độ V_{id} giới hạn trong khoảng $[-100,100]$. Vị trí X_{id} giới hạn trong khoảng $[13000, 20000]$. Các hằng số C_1, C_2 được gán giá trị 2.

Kết quả giá trị dự báo số sinh viên nhập học được đưa ra trong Bảng 2 với số lượng khoảng là 14 và sử dụng mối quan hệ mờ bậc nhất được tính với mô hình đề xuất VGPSO và một số mô hình khác. Các giá trị tại các cột tương ứng với các giá trị dự báo của từng phương pháp.

Bảng 2. Kết quả dự báo của các phương pháp khác nhau

Năm	Số lượng SV	CCO6	HPSO	AFPSO	VGPSO
1971	13055				
1972	13563	13714	13555	13579	13434
1973	13867	13714	13994	13812	13841
1974	14696	14880	14711	14565	14684
1975	15460	15467	15344	15422	15444
1976	15311	15172	15411	15307	15444
1977	15603	15467	15411	15618	15444
1978	15861	15861	15411	15660	15696
1979	16807	15831	16816	16794	16820
1980	16919	17106	17140	17032	16996
1981	16388	16380	16464	16390	16451
1982	15433	15464	15505	15504	15444
1983	15497	15172	15411	15431	15595
1984	15145	15172	15411	15077	15292
1985	15163	15467	15344	15297	15254
1986	15984	15467	16018	15848	15948
1987	16859	16831	16816	16835	16820
1988	18150	18055	18060	18145	18038
1989	18970	18998	19014	18880	18997
1990	19328	19300	19340	19418	19340
1991	19337	19149	19340	19260	19340
1992	18876	19149	19014	19031	18820
MSE		35324	22965	8224	7475

Bảng 3 đưa ra các kết quả của 4 phương pháp khác nhau để so sánh: Phương pháp Chen [3], phương pháp CCO6 của Chen và Chung [5] sử dụng giải thuật di truyền, phương pháp HPSO [15] sử dụng tối ưu bầy đàn, phương pháp AFPSO [11] sử dụng tối ưu bầy đàn kết hợp với hiệu chỉnh dự báo và phương pháp đề xuất VGPSO. Số lượng khoảng chia là 14 và sử dụng quan hệ mờ bậc 1. Có thể thấy thuật toán đề xuất có giá trị MSE nhỏ nhất mặc dù về phương pháp luận đều sử dụng nhóm quan hệ mờ và tối ưu bầy đàn. Đáng chú ý là kết quả thậm chí còn tốt hơn cả phương pháp AFPSO khi phương pháp này ngoài sử dụng nhóm quan hệ mờ theo Chen và phương pháp tối ưu bầy đàn giống như phương pháp HPSO còn hiệu chỉnh cả kết quả dự báo.

Bảng 3. So sánh độ chính xác các mô hình khác nhau theo bậc

Bậc	CCO6[5]	HPSO[15]	AFPSO[11]	VGPSO
2	67834	67123	19594	19868
3	31123	31644	31189	31307
4	32009	23271	20155	23288
5	24948	23534	20366	23552
6	26980	23671	22276	23684
7	26969	20651	18482	20669
8	22387	17106	14778	17116
9	18734	17971	15251	17987

Bảng 4 trình bày các kết quả tính toán trong giai đoạn huấn luyện bằng các phương pháp khác nhau như CCO6, HPSO và AFPSO và phương pháp đề xuất VGPSO với số lượng khoảng cố định là 7 nhưng sử dụng các mối quan hệ mờ bậc cao, từ bậc 2 đến bậc 9. Các giá trị của từng cột là chỉ số MSE được tính cho từng phương pháp theo các bậc khác nhau của mô hình chuỗi thời gian mờ bậc cao.

Từ bảng này cho thấy trong trường hợp sử dụng quan hệ mờ bậc cao, 3 phương pháp HPSO, AFPSO và VGPSO có sử dụng tối ưu bầy đàn đều nhìn hơn phương pháp CCO6 sử dụng giải thuật di truyền. Phương pháp mới đề xuất với mô hình bậc 2 và 3 tốt hơn phương pháp HPSO nhưng với các bậc cao hơn thì tương đương. Còn với phương pháp AFPSO thì mô hình đề xuất kém hiệu quả hơn đôi chút. Điều này cũng dễ hiểu vì phương pháp AFPSO ngoài sử dụng tối ưu bầy đàn để hiệu chỉnh giá trị dự báo khi lấy thêm thông tin để xác định giá trị dự báo. Cũng có thể quan sát thấy sự biến thiên đồng đều của các mô hình có sử dụng tối ưu bầy đàn để phân khoảng. Mô hình bậc 2 lại tốt hơn bậc 3 rồi giá trị MSE đều giảm dần nhưng mô hình bậc 6 lại tốt hơn mô hình bậc 5. Tương tự mô hình bậc 9 lại tốt hơn chút ít với mô hình bậc 8. Điều này chứng tỏ trong mô hình chuỗi thời gian mờ đối với từng trường hợp không phải cứ mô hình bậc cao nào cũng tốt hơn mô hình bậc thấp.

Bảng 4. So sánh hiệu quả các mô hình với số lượng khoảng khác nhau

Phương pháp	Khoảng						
	8	9	10	11	12	13	14
CCO6	132963	96244	85486	55742	54248	42497	35324
HPSO	119962	90527	60722	49257	34709	24687	22965
AFPSO	27435	24860	19698	19040	16995	11589	8224
VGPSO	34457	25855	20533	15625	14630	10004	7475

Đối với trường hợp chỉ sử dụng mô hình quan hệ mờ bậc nhất nhưng với các khoảng khác nhau thì có sự khác biệt rõ rệt về độ chính xác. Bảng 5 đưa ra các giá trị MSE để so sánh các mô hình đã kể trên cho số lượng khoảng tăng từ 8 đến 12. Rõ ràng mô hình mới đề xuất có độ chính xác tốt hơn rất nhiều so với hai mô hình CCO6 và HPSO. Còn đối với mô hình có thêm điều chỉnh giá trị dự báo AFPSO thì mô hình mới đề xuất với số khoảng là 8, 9 có độ chính xác tốt hơn. Nhưng khi số lượng khoảng tăng lên thì mô hình VGPSO tỏ rõ ưu thế so với mô hình AFPSO. Về mặt tổng thể cả bốn phương pháp đều cho thấy khi tăng số lượng khoảng thì giá trị MSE đều giảm dần.

V. KẾT LUẬN

Trong báo cáo này chúng tôi đưa ra một mô hình chuỗi thời gian mờ mới dựa trên khái niệm nhóm quan hệ mờ phụ thuộc thời gian bậc cao và tối ưu bầy đàn để nâng cao độ chính xác của dự báo. Khái niệm nhóm quan hệ mờ phụ thuộc thời gian đưa ra là một khái niệm mới, bổ sung cho khái niệm nhóm quan hệ mờ của Chen và của Yu. Đây là một công cụ cơ bản mà hầu hết các mô hình chuỗi thời gian mờ sau này đều dựa vào để xây dựng mô hình dự báo. Qua những thực nghiệm tính toán, nhìn chung mô hình chuỗi thời gian mờ mới này thực sự có hiệu quả. So sánh với các mô hình có cùng cấu trúc khi chỉ sử dụng các khái niệm nhóm quan hệ mờ bậc nhất để dự báo và phân khoảng theo phương pháp tối ưu bầy đàn như HPSO thì mô hình mới đề xuất hơn hẳn khi tăng số lượng khoảng. Còn với mô hình bậc cao với cùng số lượng khoảng chia thì hai phương pháp này tương đương và còn kém mô hình AFPSO. Điều này có thể giải thích được khi mô hình AFPSO ngoài các kỹ thuật nhóm quan hệ mờ và tối ưu bầy đàn, khi tính dự báo còn có phần hiệu chỉnh thêm. Hơn nữa, có thể thấy rằng với số lượng khoảng càng tăng thì sự khác biệt trong các nhóm quan hệ mờ theo Chen và nhóm quan hệ mờ phụ thuộc thời gian càng tăng lên và như vậy cũng có sự khác biệt về độ chính xác dự báo. Nhưng đối với các nhóm quan hệ mờ bậc cao thì về mặt cấu trúc, về phải của hai nhóm quan hệ mờ này hầu như không có sự thay đổi nên độ chính xác dự báo tương đương nhau. Vì thế đối với các mô hình sử dụng nhóm quan hệ mờ bậc cao thì độ chính xác sẽ không có nhiều khác biệt. Hy vọng trong mô hình chuỗi thời gian mờ mới đề xuất có đưa thêm kỹ thuật hiệu chỉnh giá trị dự báo bằng phân thêm khoảng phụ như mô hình trong [16] hay điều chỉnh bằng tổ hợp giữa thông tin toàn cục và thông tin địa phương như mô hình AFPSO trong [10] thì mô hình chuỗi thời gian mờ với nhóm quan hệ mờ phụ thuộc thời gian sẽ hiệu quả hơn.

TÀI LIỆU THAM KHẢO

- [1] S. Askari, N. Montezzin, A high-order multivariable fuzzy time series forecasting algorithm based on fuzzy cluster, *Expert Systems with Applications*, (2015) 42 2121 – 2135.
- [2] M. Avazbeigi, S.H. Hashemi Doulabi, B. Karimi, Choosing the appropriate order in fuzzy time series: A new N-factor fuzzy time series for prediction of the auto industry production. *Expert Systems with Applications*. 37, (2010), 5630–5639.
- [3] S. M. Chen, Forecasting Enrollments based on Fuzzy Time Series., *Fuzzy set and systems*, vol. 81, (1996), pp. 311-319.
- [4] S. M. Chen, Forecasting Enrollments based on high-order Fuzzy Time Series. *Int. Journal: Cybernetic and Systems*, N.33, (2002), pp. 1-16.
- [5] S. M. Chen, N. Y. Chung, Forecasting enrollments of students by using fuzzy time series and genetic algorithms. *International Journal of Information and Management Sciences*, 17, (2006) 1–17.
- [6] S. M. Chen, N. I. Wang, J. S. Pan, Forecasting enrollments using automatic clustering techniques and fuzzy logical relationships, *Expert Systems with Applications*. 36, (2009) 11070–11076.

- [7] S. M. Chen, K. Tanuwijaya, Fuzzy forecasting based on high- order fuzzy logical relationships and automatic clustering techniques, *Expert Systems with Applications*. (2011), 38 15425–15437.
- [8] S. H. Cheng, S. M. Chen, W. S. Jian, Fuzzy time series forecasting based on fuzzy logical relationship similarity measures. *Information Sciences*, 327, (2016), 272–287.
- [9] E. Egriglu, C. H. Aladag, et al. Finding an optimal interval length in high order fuzzy time series. *Expert Systems with Applications*. 37 (2010) 5052–5055.
- [10] Y. L. Huang, S. J. Horng, T. W. Kao, R. S. Run, J. L. Lai, R. J. Chen, I. H. Kuo, M. K. Khan, An improved forecasting model based on the weighted fuzzy relationship matrix combined with PSO adaptation for enrollment. *Intern. J. of Innovation Computing, Information and Control*, 7, (2011), 4027–4045.
- [11] Y. L. Huang, S. J. Horng, M. He, P. Fan, T. W. Kao, M. K. Khan, A hybrid forecasting model for enrollments based on aggregated fuzzy time series and particle swarm optimization, *Expert Systems with Applications*. (2011), 38, 8014 – 8023.
- [12] K. Huarng, Effective lengths of intervals to improve forecasting in fuzzy time series. *Fuzzy Sets and Systems*, (2001b), 123, 387–394.
- [13] K. Huarng, Heuristic models of fuzzy time series for forecasting. *Fuzzy Sets and Systems*, 123, (2001a), 369–386.
- [14] J. Kennedy, R. Eberhart, Particle swarm optimization, In *Proceedings of IEEE international conference on neural network*, (1995) 1942–1948
- [15] I. H. Kuo, 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.
- [16] I. H. Kuo, S. J. Horng, Y. H. Chen, R. S. Run, T. W. Kao, R. J. Chen, Forecasting TAIEX based on fuzzy time series and particle swarm optimization, *Expert Systems with Applications*. 37, (2010), 1494–1502.
- [17] L. W Lee, L. H. Wang, S. H. Chen, Y. H. Leu A new method for handling forecasting problems based on two-factors high-order fuzzy time series. In *Proceedings ninth conference on artificial intelligence and applications*, Taipei, Taiwan, Republic of China , (2004).
- [18] L. W. Lee, L. H. Wang, S. M. Chen, Temperature prediction and TAIEX forecasting based on fuzzy logical relationships and genetic algorithms. *Expert Systems with Applications*, 33, (2007), 539–550.
- [19] L. W. Lee, L. H. Wang, S. M. Chen, Temperature prediction and TAIEX forecasting based on hight order fuzzy logical relationship and genetic simulated annealing techniques. *Expert Systems with Applications*, 34 (2008) 328–336.
- [20] Nguyễn Công Điều. Một thuật toán mới cho mô hình chuỗi thời gian mờ heuristic trong dự báo chứng khoán, *Khoa học và Công nghệ*, Viện KH&CN VN, 52(6), (2011), 659-672.
- [21] Nguyễn Công Điều, Nhóm quan hệ mờ phụ thuộc thời gian trong mô hình chuỗi thời gian mờ. *Tạp chí KHCN*, Viện Hàn lâm KH và CN Việt Nam, 52(6), (2014) 659-672.
- [22] Q. Song, B.S. Chissom, Fuzzy Time Series and its Model. *Fuzzy set and systems*, vol. 54, (1993), pp. 269-277
- [23] Q. Song, B.S. Chissom, Forecasting Enrollments with Fuzzy Time Series – Part I, *Fuzzy set and systems*, 54 (1993) 1-9.
- [24] Q. Song, B.S. Chissom, Forecasting Enrollments with Fuzzy Time Series – Part II. *Fuzzy set and systems*, vol. 62, (1994) 1-8.
- [25] H.K. Yu, Weighted fuzzy time series models for TAIEX forecasting , *Physica A*, 349 (2005) 609–624.

FUZZY TIME SERIES FORECASTING BASED ON TIME-DEPENDING FUZZY RELATIONSHIP GROUPS AND PARTICLE SWARM OPTIMIZATION

Nguyen Cong Dieu, Nghiem Van Tinh

ABSTRACT— In recent times, fuzzy time series model is attracting the attention of researchers and data analysts. From the original model proposed by Song and Chilssom, more fuzzy time series models had been proposed to improve the accuracy of forecasts. However, there still exist some unsolved problems to better product forecasting in the fuzzy time series model. That is, how to split the universe of discourse into the appropriate lengths and how to build fuzzy logical relationships and fuzzy relationship groups effectively. Some approachs have been proposed to solve the problems. One of which is the combination of fuzzy time series with optimization methods such as genetic algorithms, simulated annealing techniques, particle swarm optimization and ant colony optimization. In this paper, fuzzy time series model based on a new concept of time-dependent fuzzy logical relationship groups and the particle swarm optimization techniques is developed. Experimental results show that the proposed model has better forecasting accuracy than previous methods.

Keywords— Fuzzy time series, forecasting, time-depending fuzzy logical relationship groups, particle swarm optimization.