

MỘT THUẬT TOÁN MỚI CHO MÔ HÌNH CHUỖI THỜI GIAN MỜ HEURICTIC TRONG DỰ BÁO CHỨNG KHOÁN

Nguyễn Công Điều

Phòng Thống kê tính toán, Viện CNTT

Đền Toà soạn ngày: 17/7/2010

1. MỞ ĐẦU

Chuỗi thời gian đang được sử dụng như một công cụ hữu hiệu để phân tích số liệu trong kinh tế, xã hội cũng như trong nghiên cứu khoa học. Chính do tầm quan trọng của lĩnh vực này, rất nhiều tác giả đã đề xuất các công cụ phân tích chuỗi thời gian để trích xuất ra những thông tin quan trọng từ trong dãy số liệu đó.

Trong đây, phương pháp chủ yếu để phân tích chuỗi thời gian là sử dụng các công cụ của thống kê như hồi quy, phân tích Furie và một vài công cụ khác. Nhưng hiệu quả nhất có lẽ là phương pháp sử dụng mô hình ARIMA của Box-Jenkins. Mô hình này đã cho một kết quả khá tốt trong phân tích dữ liệu và đang được sử dụng rất rộng rãi trong thực tế. Tuy nhiên, sự phức tạp của thuật toán đã gây khó khăn khi ứng dụng trong phân tích chuỗi số liệu, nhất là khi chuỗi số liệu có những thay đổi phản ánh sự phi tuyến của mô hình.

Để vượt qua được những khó khăn trên, gần đây nhiều tác giả đã sử dụng mô hình chuỗi thời gian mờ. Khái niệm tập mờ được Zadeh đưa ra từ năm 1965 và ngày càng tìm được ứng dụng trong nhiều lĩnh vực khác nhau nhất là trong điều khiển và trí tuệ nhân tạo. Trong lĩnh vực phân tích chuỗi thời gian, Song và Chissom [10 - 12] đã đưa ra khái niệm chuỗi thời gian mờ không phụ thuộc vào thời gian (chuỗi thời gian dừng) và phụ thuộc vào thời gian (không dừng) để dự báo. Chen [2] đã cải tiến và đưa ra phương pháp mới đơn giản và hữu hiệu hơn so với phương pháp của Song và Chissom. Trong phương pháp của mình, thay vì sử dụng các phép tính tổ hợp Max-Min phức tạp, Chen đã tính toán bằng các phép tính số học đơn giản để thiết lập các mối quan hệ mờ. Phương pháp của Chen cho hiệu quả cao hơn về mặt sai số dự báo và giảm độ phức tạp của thuật toán.

Từ các công trình ban đầu về chuỗi thời gian mờ được xuất hiện năm 1993, hiện nay mô hình này đang được sử dụng để dự báo trong rất nhiều lĩnh vực của kinh tế hay xã hội như giáo dục để dự báo số sinh viên nhập trường [2, 11] hay trong lĩnh vực dự báo thất nghiệp [6], dân số [1], chứng khoán [5, 8] và trong đời sống như dự báo mức tiêu thụ điện, hay dự báo nhiệt độ của thời tiết...

Tuy nhiên xét về độ chính xác của dự báo, các thuật toán trên cho kết quả chưa cao. Để nâng cao độ chính xác của dự báo, một số thuật toán cho mô hình chuỗi thời gian mờ liên tiếp được đưa ra. Chen [2] đã sử dụng mô hình bậc cao của chuỗi thời gian mờ để tính toán. Sah và Degtiarev [9] thay vì dự báo chuỗi thời gian đã sử dụng chuỗi thời gian là hiệu số bậc nhất để nâng cao độ chính xác và làm giảm độ phi tuyến. Đây cũng là một phương pháp hay được sử dụng trong mô hình Box-Jenkins để loại bỏ tính không dừng của chuỗi thời gian. Huarng [5] 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.

Trong báo cáo này, chúng tôi trình bày một cải tiến mô hình heuristic chuỗi thời gian mờ và áp dụng mô hình trong dự báo chỉ số chứng khoán. Tư tưởng chính của phương pháp là sử dụng một số khái niệm của Huarng [5] và Chen, Hsu [4] để phát triển thuật toán mới. Dựa trên thuật toán đề ra, chúng tôi đã tính toán một bài toán thực tế dựa trên dữ liệu lấy từ thị trường chứng khoán Đài Loan để kiểm chứng. Kết quả thu được rất khả quan. Độ chính xác của dự báo được nâng lên khá nhiều so với các thuật toán trước đây đề ra.

2. MỘT SỐ KHÁI NIỆM

Trong phần này, chúng ta sẽ sử dụng khái niệm và phương pháp dự báo của chuỗi thời gian mờ được Song *et. al.* [10, 11] và Chen [2] đưa ra để xây dựng thuật toán dự báo cho chuỗi thời gian.

2.1 Chuỗi thời gian mờ

Giả sử U là không gian nền. Không gian nền này xác định một tập hợp các đối tượng cần nghiên cứu. Nếu A là một tập con rõ của U thì ta có thể xác định chính xác một hàm đặc trưng:

$$\mu_A(x) = \begin{cases} 0 & \text{nếu } x \text{ nằm ngoài } A \\ 1 & \text{nếu } x \text{ nằm trong } A \end{cases}$$

Nhưng với một tập mờ B trong không gian nền U thì phần tử x không xác định chính xác được. Khi đó ta có định nghĩa:

Tập A là mờ trên không gian nền U nếu A được xác định bởi hàm:

$$\mu_A : U \rightarrow [0, 1]$$

μ_A được gọi là hàm thuộc (*Membership function*). Còn với bất kỳ một phần tử u nào của A thì hàm $\mu_A(u)$ được gọi là độ thuộc của u vào tập mờ A .

Giả sử $Y(t)$ là chuỗi thời gian ($t = 0, 1, 2, \dots$)

$U = \{u_1, u_2, \dots, u_n\}$ là tập nền. Tập mờ A trên không gian nền U được viết như sau:

$$A = \{(\mu_A(u_1)/u_1, \mu_A(u_2)/u_2, \dots, \mu_A(u_n)/u_n), : u_i \in U ; i=1, 2, \dots, n\}$$

$\mu_A(u_i)$ là độ thuộc của u_i vào tập A hay một cách viết khác:

$$A = \frac{\mu_A(u_1)}{u_1} + \frac{\mu_A(u_2)}{u_2} + \dots + \frac{\mu_A(u_n)}{u_n}$$

Một số định nghĩa sau liên quan đến chuỗi thời gian mờ [5].

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 $Y(t)$.

2.2 Mối quan hệ mờ

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 kí hiệu $F(t-1) \rightarrow F(t)$.

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: $A_i \rightarrow A_j$.

Định nghĩa 3: Nhóm các mối quan hệ mờ.

Các mối quan hệ logic có thể gộp lại thành một nhóm nếu trong kí hiệu trên, cùng một vế trái sẽ có nhiều mối quan hệ tại vế phải. Thí dụ nếu ta có các mối quan hệ:

$$A_i \rightarrow A_k$$

$$A_i \rightarrow A_m$$

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

$$A_i \rightarrow A_k, A_m.$$

Định nghĩa 4:

Giả sử $F(t)$ suy ra từ $F(t-1)$ và $F(t) = F(t-1) * R(t-1, t)$ cho mọi t . 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.

Quá trình dự báo cho chuỗi thời gian mờ cũng dựa trên các bước của phương pháp lập luận xấp xỉ mờ. Như tác giả N. C. Hồ [7] đã tổng kết 4 bước lập luận xấp xỉ mờ như sau:

1. Giải nghĩa các mệnh đề mờ điều kiện
2. Kết nhập các quan hệ mờ
3. Tính kết quả từ phép hợp thành
4. Khử mờ.

Từ những bước lập luận chung như trên, đối với chuỗi thời gian mờ, một số tác giả như Song và Chissom [10, 11], Chen [2] đã đưa ra một số bước trong phương pháp luận xử lí tập mờ cho chuỗi thời gian. Dưới đây chúng tôi mô tả thuật toán của Chen [2] các bước thực hiện trong dự báo mô hình chuỗi thời gian mờ. Thuật toán này 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
3. Xác định các tập mờ trên tập U
4. Mờ hoá các dữ liệu chuỗi thời gian
5. Thiết lập các mối quan hệ mờ và nhóm các quan hệ mờ
6. Dự báo
7. Giải mờ các kết quả dự báo

Các thuật toán để dự báo theo chuỗi thời gian mờ chủ yếu đều dựa vào bước cơ bản trên. Những thay đổi của các tác giả khác nhau chủ yếu tại các bước tính toán mối quan hệ mờ $R(t-1, t)$ và đưa ra các luật để dự báo..

2.3. Mô hình heuristic cho chuỗi thời gian mờ

Huang [5] đã sử dụng mô hình của Chen [2] và đưa vào các thông tin có sẵn của chuỗi thời gian để cải tiến độ chính xác và giảm bớt các tính toán phức tạp của dự báo. Nhờ sử dụng những thông tin có sẵn trong chuỗi thời gian nên mô hình của Huang được gọi là mô hình heuristic.

Các bước thực hiện của mô hình Huang cũng triển khai theo các bước trên. Điều khác biệt là sử dụng một hàm h để xác định mối quan hệ logic mờ. Dưới đây chúng tôi mô tả các bước thực hiện của mô hình heuristic chuỗi thời gian mờ [5].

Bước 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 $U = [f_{\max}, f_{\min}]$. Đôi khi có thể mở rộng khoảng này thêm một giá trị nào đó để dễ tính toán. Chia đoạn U thành m khoảng con bằng nhau u_1, u_2, \dots, u_m .

Bước 2: Xác định các tập mờ A_i và mờ hoá giá trị. Mỗi tập A_i gán cho một biến ngôn ngữ và xác định trên các đoạn đã xác định u_1, u_2, \dots, u_m . Khi đó các tập mờ A có thể biểu diễn như sau:

$$A_i = \frac{\mu_{A_i}(u_1)}{u_1} + \frac{\mu_{A_i}(u_2)}{u_2} + \dots + \frac{\mu_{A_i}(u_m)}{u_m}$$

Bước 3: Thiết lập mối quan hệ mờ và nhóm các mối quan hệ mờ. Như định nghĩa ở trên, đối với chuỗi thời gian mờ ta có thể xác định được mối quan hệ mờ tại mỗi thời điểm t và qua đó xác định được nhóm các mối quan hệ mờ.

Bước 4: Sử dụng hàm h để thiết lập các nhóm mối quan hệ logic mờ heuristic

$$A_i \rightarrow h_j(x, A_{p1}, A_{p2}, \dots) = A_{p1}, A_{p2}, \dots, A_{pk}$$

Bước 5: Dự báo. Từ các nhóm quan hệ logic mờ heuristic. Các giá trị chủ yếu lấy từ điểm giữa hay trung bình các điểm giữa các khoảng trong nhóm quan hệ mờ heuristic.

3. ĐỀ XUẤT MỚI CHO CHUỖI THỜI GIAN MỜ HEURISTIC

3.1. Một số khái niệm

Trước hết ta cần một số khái niệm. Các tập mờ A_1, A_2, \dots, A_k có thể sắp xếp được, có nghĩa là $A_f \geq A_g$ khi $f \geq g$. Nếu $F(t-1) = A_j$ và $F(t) = A_i$ thì khi đó ta có mối quan hệ mờ $A_j \rightarrow A_i$. Ngoài ra ta cũng có thể xác định được nhóm quan hệ mờ:

$$A_j \rightarrow A_{p1}, A_{p2}, \dots, A_{pk}$$

Định nghĩa 5. Hàm h_j phụ thuộc vào một tham số x được xác định như sau:

$$h_j(x, A_{p1}, A_{p2}, \dots) = A_{p1}, A_{p2}, \dots, A_{pk} \quad j \text{ là một chỉ số}$$

khi các chỉ số $p1, p2, \dots, pk \geq j$ với $x > 0$

và $p1, p2, \dots, pk \leq j$ với $x < 0$

Ngoài ra, để đưa ra luật giải mờ heuristic, ta cần thêm thông tin thông qua việc xác định hiệu số bậc nhất và bậc 2 cho chuỗi thời gian. Giả sử các giá trị của chuỗi thời gian tại các thời điểm tương ứng $t, t-1, t-2$ là $f(t), f(t-1), f(t-2)$. Khi đó các hiệu số bậc nhất và bậc 2 được xác định:

$$\Delta_i = f(t) - f(t-1); \quad \Delta_i^2 = (f(t) - f(t-1)) - (f(t-1) - f(t-2))$$

Tương tự như khi khảo sát một hàm số, nếu hiệu số bậc nhất là dương thì đó là hàm tăng, còn hiệu số bậc nhất âm thì đó là hàm giảm. Khái niệm hiệu số bậc 2 cho phép thêm thông tin về tốc độ tăng giảm của hàm: hàm tăng (giảm) nhanh phụ thuộc vào hiệu số bậc 2 âm hay dương.

Ngoài ra còn xét đến điểm lấy giá trị trong khoảng phân chia trong bước giải mờ. Phụ thuộc vào độ tăng giảm của chuỗi thời gian, các điểm được lấy để tính toán trong khoảng không phải là điểm giữa khoảng nữa mà trong thuật toán dưới đây, chúng tôi sẽ lấy các điểm 0.25 (điểm dưới), 0.5 (điểm giữa) và 0.75 (điểm trên) của khoảng.

Thuật toán chúng tôi đề xuất về có những bước tương tự nhưng có những thay đổi tại bước 1 trong chia khoảng giá trị, bước 3 trong việc xác định hàm h và tính các điểm dự báo trong các khoảng trong nhóm các mối quan hệ mờ heuristic. Hàm h_i tính tại thời điểm t và dựa vào tham số hiệu số bậc nhất. Điểm cải tiến cuối cùng là các quy tắc dự báo. Các giá trị để tính dự báo không phải là tại điểm giữa của khoảng nữa mà dựa trên các thông tin có sẵn về hiệu số bậc 1 và hiệu số bậc 2 để tính giá trị tại các điểm dưới, điểm giữa và điểm trên của khoảng đã xác định. Như vậy, thông tin về chuỗi thời gian không chỉ lấy từ hiệu số bậc nhất nữa mà thêm thông tin từ hiệu số bậc 2 của các giá trị chuỗi thời gian.

3.2. Thuật toán đề xuất

Xét bài toán dự báo cho chuỗi dữ liệu chỉ số thị trường chứng khoán Đài Loan TAIFEX [5]. Số liệu được đưa ra trong bảng dưới đây.

Bảng 1. Giá trị chỉ số chứng khoán Đài Loan

NgàyThang	GiaTriThuc	NgàyThang	GiaTriThuc	NgàyThang	GiaTriThuc
03/08/1998	7552	25/08/1998	6949	15/09/1998	6762
04/08/1998	7560	26/08/1998	6790	16/09/1998	6952,75
05/08/1998	7487	27/08/1998	6835	17/09/1998	6906
06/08/1998	7462	28/08/1998	6695	18/09/1998	6842
07/08/1998	7515	29/08/1998	6728	19/08/1998	7039
10/08/1998	7365	31/08/1998	6566	21/09/1998	6861
11/08/1998	7360	01/09/1998	6409	22/09/1998	6926
12/08/1998	7320	02/09/1998	6430	23/09/1998	6852
13/08/1998	7291	03/09/1998	6200	24/09/1998	6890
14/08/1998	7320	04/09/1998	6403,2	25/09/1998	6871
15/08/1998	7300	05/09/1998	6697,5	28/09/1998	6840
17/08/1998	7219	07/09/1998	6722,3	29/09/1998	6806
18/08/1998	7220	08/09/1998	6859,4	30/09/1998	6787
19/08/1998	7285	09/09/1998	6769,6		
0/08/1998	7274	10/09/1998	6709,75		
08/1998	7225	11/09/1998	6726,5		
24/08/1998	6955	14/09/1998	6774,55		

Thuật toán 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 6200 và 7560 điểm. Do vậy tập nền U được xác định là giá trị trong khoảng [6200,7600]. Ta sẽ chia U thành 14 khoảng u_1, u_2, \dots, u_{14} với độ rộng là 100, như vậy các khoảng sẽ là: $u_1 = [6200,6300], u_2 = [6300,6400], \dots, u_{14} = [7500,7600]$.

Bước 2: Xác định các tập A_i ứng với từng khoảng u_i xác định tại bước 1. Ta gán chúng với các biến ngôn ngữ. Thí dụ $A_1 = (\text{Thấp nhất}), A_2 = (\text{rất rất thấp}), A_3 = (\text{rất thấp}), A_4 = (\text{thấp}), A_5 = (\text{hơi thấp}), A_6 = (\text{dưới trung bình}), A_7 = (\text{trung bình}), A_8 = (\text{trên trung bình}), A_9 = (\text{trung bình cao}), A_{10} = (\text{hơi cao}), A_{11} = (\text{rất cao}), A_{13} = (\text{rất rất cao}), A_{14} = (\text{cao nhất})$. Với mỗi tập A_i được xác định bởi một đoạn u_i .

Bước 3: Chia lại khoảng. Tính phân bố của các giá trị chuỗi thời gian rơi vào các khoảng đã chia. Điều này thực hiện để biết các khoảng nào có nhiều giá trị rơi vào để có thể phân khoảng tiếp làm tăng độ chính xác khi dự báo.

Bảng sau đây sẽ cho thấy sự phân bố các giá trị của chuỗi thời gian rơi vào từng khoảng:

Bảng 2. Phân bố giá trị trong từng khoảng

khoảng	số lượng	khoảng	số lượng
6200 - 6300	1	6900 - 7000	5
6300 - 6400	0	7000 - 7100	1
6400 - 6500	3	7100 - 7200	0
6500 - 6600	1	7200 - 7300	6
6600 - 6700	2	7300 - 7400	5
6700 - 6800	9	7400 - 7500	2
6800 - 6900	9	7500 - 7600	3

Xem xét bảng trên ta thấy sự phân bố các giá trị tại các khoảng khác nhau là không đều nhau. Có 47 giá trị trong 14 khoảng nên số lượng trung bình rơi vào mỗi khoảng là hơn 3. Nhưng có những khoảng rơi vào đến 6 hay 9 giá trị. Vì vậy phải chia những khoảng có nhiều giá trị thành những khoảng con để có thể phân bố đều lại các giá trị này. Vì vậy những khoảng nào có 5, 6 giá trị rơi vào ta chia tiếp làm 2 khoảng con, còn những đoạn nào có 8,9 giá trị rơi vào ta tiếp tục chia thành 3 khoảng để sao cho mỗi khoảng con đó có xấp xỉ 3 giá trị rơi vào. Kết quả sẽ hình thành 21 khoảng sau:

Bảng 3. Phân khoảng

$u_1 = [6200 - 6300]$	$u_8 = [6766 - 6800]$	$u_{15} = [7100 - 7200]$
$u_2 = [6300 - 6400]$	$u_9 = [6800 - 6833]$	$u_{16} = [7200 - 7250]$
$u_3 = [6400 - 6500]$	$u_{10} = [6833 - 6866]$	$u_{17} = [7250 - 7300]$
$u_4 = [6500 - 6600]$	$u_{11} = [6866 - 6900]$	$u_{18} = [7300 - 7350]$
$u_5 = [6600 - 6700]$	$u_{12} = [6900 - 6950]$	$u_{19} = [7350 - 7400]$
$u_6 = [6700 - 6733]$	$u_{13} = [6950 - 7000]$	$u_{20} = [7400 - 7500]$
$u_7 = [6733 - 6766]$	$u_{14} = [7000 - 7100]$	$u_{21} = [7500 - 7600]$

Bước 6: Dự báo

Sử dụng hàm heuristic được mô tả tại Bước 5 để dự báo giá trị cho chuỗi thời gian. Nguyên tắc dự báo như sau:

Giả sử tại thời điểm t , giá trị mờ tại thời điểm này của chuỗi thời gian mờ được suy ra từ giá trị mờ tại thời điểm $t-1$ theo công thức $F(t) = F(t-1) * R(t-1, t)$, hay có thể viết $A_i \rightarrow A_j$.

Như vậy theo các phương pháp truyền thống [11, 12], phải tính được mối quan hệ $R(t-1, t)$. Trong phương pháp heuristic [5], mối quan hệ được sử dụng là nhóm các quan hệ mờ. Trong phương pháp chúng tôi đề xuất, để dự báo giá trị mờ A_j , chúng tôi sử dụng hàm heuristic cho nhóm quan hệ mờ của A_i . Như vậy đối với mỗi thời điểm t ta phải tính hàm h (theo Định nghĩa 5) heuristic tại thời điểm $t-1$ tức là mối quan hệ mờ của A_i . Nhóm mỗi quan hệ mờ và nhóm mỗi quan hệ mờ heuristic tại mỗi thời điểm t được tính toán cụ thể theo bảng 5.

Bảng 5. Nhóm quan hệ mờ và nhóm quan hệ mờ heuristic và điểm tính để dự báo

Actual index	Giá trị mờ	Hiệu số bậc 1	Hiệu số bậc 2	Nhóm quan hệ mờ	Nhóm quan hệ Heuristic	Điểm tính
7552	A21					
7560	A21	8				
7487	A20	-73	-81	A19,A20,A21	A19,A20	0,5, 0,75
7462	A20	-25	48	A20,A21	A20	0,25
7515	A21	53	78	A20,A21	A20,A21	0,5, 0,75
7365	A19	150	97	A19,A20,A21	A19,A20,A21	0,25, 0,5, 0,75
7360	A19	-5	-155	A18,A19	A18,A19	0,5, 0,75
7330	A18	-30	-25	A18,A19	A18	0,75
7291	A17	-29	1	A16,A17,A18	A16,A17	0,5, 0,75
7320	A18	29	58	A16,A17,A18	A18	0,75
7300	A18	-20	-49	A16,A17,A18	A16,A17,A18	0,25, 0,5, 0,75
7219	A16	-81	-61	A16,A17,A18	A16	0,75
7220	A16	1	82	A13,A16,A17	A16,A17	0,5, 0,75
7283	A17	63	62	A13,A16,A17	A17	0,75
7274	A17	-9	-72	A16,A17,A18	A16,A17	0,5, 0,75
7225	A16	-49	-40	A16,A17,A18	A16	0,75
6955	A13	-270	-221	A13,A16,A17	A13	0,75
6949	A12	-6	264	A12	A12	0,25
6790	A8	-159	-153	A8,A10	A8	0,75
6835	A10	45	204	A6,A7,A10	A10	0,75

6695	A5	-140	-185	A5,A8,A9,A11,A14	A5	0,75
6728	A6	33	173	A6	A6	0,75
6566	A4	-162	-195	A4,A6,A8,A10	A4	0,75
6409	A3	-157	5	A3	A3	0,25
6430	A3	21	178	A1,A3,A5	A3,A5	0,5, 0,75
6200	A1	-230	-251	A1,A3,A5	A1	0,75
6403,2	A3	203,2	433,2	A3	A3	0,75
6697,5	A5	294,3	91,1	A1,A3,A5	A5	0,75
6722,3	A6	24,8	-269,5	A6	A6	0,25
6859,4	A10	137,1	112,3	A4,A6,A8,A10	A10	0,75
6769,6	A8	-89,8	-226,9	A5,A8,A9,A11,A14	A5,A8	0,5, 0,75
6709,75	A6	-59,85	29,95	A6,A7,A10	A6	0,25
6726,5	A6	16,75	76,6	A4,A6,A8,A10	A6,A8,A10	0,25, 0,5, 0,75
6774,55	A8	48,05	31,3	A4,A6,A8,A10	A8,A10	0,5, 0,75
6762	A7	-12,55	-60,6	A6,A7,A10	A6,A7	0,5, 0,75
6952,75	A13	190,75	203,3	A13	A13	0,75
6906	A12	-46,75	-237,5	A12	A12	0,75
6842	A10	-64	-17,25	A8,A10	A8,A10	0,5, 0,75
7039	A14	197	261	A5,A8,A9,A11,A14	A14	0,75
6861	A11	-178	-375	A11	A11	0,75
6926	A12	65	243	A10,A11,A12	A12	0,75
6852	A10	-74	-139	A8,A10	A8,A10	0,5, 0,75
6890	A11	38	112	A5,A8,A9,A11,A14	A11,A14	0,5, 0,75
6871	A11	-19	-57	A10,A11,A12	A10,A11	0,5, 0,75
6840	A10	-31	-12	A10,A11,A12	A10	0,75
6806	A9	-34	-3	A5,A8,A9,A11,A14	A5,A8,A9	0,25, 0,5, 0,75
6787	A8	-19	15	A8	A8	0,25

Các quy tắc dự báo:

Quy tắc 1: Nếu quan hệ mờ heuristic của A_i là rỗng $A_i \rightarrow 0$ thì giá trị dự báo của $F(t)$ là giá trị điểm giữa m_i của u_i .

Quy tắc 2: Nếu quan hệ mờ heuristic của A_i là một một, nghĩa là $A_i \rightarrow A_k$ thì giá trị dự báo của $F(t)$ là điểm giữa, điểm trên hoặc điểm dưới của đoạn u_k tùy thuộc theo tính chất của hiệu số bậc 1 và hiệu số bậc 2 của chuỗi tại thời điểm t (xem bảng 6, lấy giá trị cuối cùng bên phải).

Quy tắc 3: Nếu quan hệ mờ heuristic của A_i là một nhiều thì ta xác định theo các giá trị khác nhau của các khoảng u_i dựa vào các thông tin của chuỗi thời gian sau:

Đối với mỗi thời điểm t , ta cần các giá trị chuỗi thời gian $f(t-2), f(t-1), f(t)$. Tại thời điểm t , ta cũng cần xác định các hiệu số bậc nhất $\Delta = f(t) - f(t-1)$ và hiệu số bậc 2 $\Delta^2 = (f(t)-f(t-1)) - (f(t-1)-f(t-2))$ của giá trị chuỗi thời gian. Dựa vào cách xác định hàm $h(\Delta, A_{p1}, A_{p2}, \dots, A_{pm})$ để xác định mối quan hệ mờ heuristic tại thời điểm t theo giá trị dương hay âm của Δ . Trong bài báo này, chúng tôi sử dụng cả hiệu số bậc 2 để xác định thêm tính chất của chuỗi thời gian. Tuỳ theo tính chất tăng, giảm của chuỗi thời gian tại thời điểm t để xác định các giá trị dự báo tại các khoảng trong mỗi quan hệ mờ. Một khoảng u_i ta xác định các giá trị tại giữa khoảng (0,5), $\frac{3}{4}$ khoảng (0,75) và $\frac{1}{4}$ khoảng (0,25). Các giá trị được xác định tương ứng với các giá trị mờ hoá A_i tương ứng với khoảng u_i . Ta chỉ quan tâm đến 3 giá trị mờ hoá gần với A_j nhất. Các giá trị khác lấy tại điểm gần nhất. Do vậy, ta có quy luật lấy giá trị tại các khoảng tương ứng như bảng 6.

Bảng 6. Các điểm lấy giá trị dự báo trong khoảng

Tính chất chuỗi	Hiệu bậc nhất	Hiệu bậc 2	Các điểm lấy giá trị
Giảm từ từ	$\Delta < 0$	$\Delta^2 > 0$	0,75, ..., 0,75, 0,5, 0,25
Giảm nhanh	$\Delta < 0$	$\Delta^2 < 0$	0,25, ..., 0,25, 0,5, 0,75
tăng nhanh	$\Delta > 0$	$\Delta^2 > 0$	0,25, ..., 0,25, 0,5, 0,75
tăng từ từ	$\Delta > 0$	$\Delta^2 < 0$	0,75, ..., 0,75, 0,5, 0,25

Giá trị dự báo của chuỗi thời gian tại thời điểm t là giá trị trung bình của các giá trị tại khoảng dựa vào bảng trên.

Dựa vào bảng 6, ta có thể tính dự báo cho chuỗi thời gian tại thời điểm t . Chúng tôi đưa ra một trường hợp để làm thí dụ.

Ngày 10/9 và ngày 11/9 có các giá trị tương ứng tại bảng 1 là 6709,7 và 6726,5. Còn tại bảng 5 là hai hàng được bôi đen. Giá trị mờ của chuỗi thời gian tương ứng là A_6 và A_6 (xem cột tương ứng thứ nhất). Tại thời điểm này, hiệu số bậc nhất tương ứng là $-59,85$ và $16,75$ tức là một giá trị âm còn một giá trị dương. Mối quan hệ ngày 10/9 là $A_8 \rightarrow A_6$. Như vậy để dự báo ta cần nhóm quan hệ $A_8 \rightarrow A_6, A_7, A_{10}$. Để tính quan hệ mờ heuristic, ta sử dụng hàm heuristic

$$h_6(\Delta, A_6, A_7, A_{10}) = A_6 \text{ vì } \Delta \text{ âm nên chỉ lấy các chỉ số } \leq 6.$$

Như vậy giá trị dự báo sẽ rơi vào giá trị mờ A_6 tương ứng với khoảng $u_6 = [6700 - 6730]$. Giá trị hiệu số bậc 2 là dương, do vậy để xem lấy điểm nào trong khoảng dự báo ta lại xem Bảng 6: $\Delta < 0, \Delta^2 > 0$ nên theo bảng trên giá trị này lấy ở điểm dưới của khoảng (0,25). Điểm này tương ứng với giá trị xấp xỉ 6708. Như vậy ta đã dự báo xong thời điểm ngày 10/9.

Tính tiếp dự báo cho ngày 11/9. Dự báo theo quan hệ $F(10/9) \rightarrow F(11/9)$ hay $A_6 \rightarrow A_6$. Nhóm quan hệ mờ của nó là $A_6 \rightarrow A_4, A_6, A_8, A_{10}$. Xác định nhóm quan hệ mờ heuristic sử dụng hàm heuristic với hiệu số bậc nhất tại thời điểm này có giá trị 16.75 tức là giá trị dương, ta thu được như sau:

$$h_6(\Delta, A_4, A_6, A_8, A_{10}) = A_6, A_8, A_{10} \text{ vì } \Delta \text{ dương nên chỉ lấy các chỉ số } \geq 6.$$

Như vậy giá trị dự báo sẽ chỉ lấy trung bình trong các khoảng u_6, u_8, u_{10} . Điểm lấy giá trị tương ứng trong khoảng lại xét dấu của hiệu số bậc nhất và hiệu bậc 2 tại thời điểm này. Tính

toán cho thấy cả hai giá trị đều dương nên tính chất của chuỗi số liệu là tăng nhanh nên các điểm tính tương ứng sẽ là 0,25, 0,5, 0,75 của ba khoảng trên và dự báo sẽ là giá trị trung bình của 3 giá trị trên. Điểm 0,25 của khoảng u_6 là 6708. Điểm 0,5 của u_8 có giá trị là 6785, còn điểm 0,75 của khoảng u_{10} có giá trị là 6852. Như vậy giá trị dự báo $f(11/9)$ sẽ là:

$$f(11/9) = (6708 + 6786 + 6852) / 3 = 6781,7 \approx 6782.$$

Lập được bảng 6 ta dễ dàng tính được các giá trị dự báo.

4. KẾT QUẢ TÍNH TOÁN

Chúng tôi đã sử dụng thuật toán trên để tính toán các chỉ số của thị trường chứng khoán Đài Loan TAIFEX theo số liệu của đưa ra trong [5]. Kết quả tính toán này được so sánh với các kết quả của thuật toán Chen [2] và thuật toán heuristic hai tham số và ba tham số của Huarng [5]. Kết quả cho trong bảng 7.

Bảng 7. Kết quả tính toán

Ngày tháng	Actual index	Chen	Huarng1	Huarng2	Dự báo
3/8/1998	7552	7450	7450	7450	7550
4/8/1998	7560	7450	7450	7450	7550
5/8/1998	7487	7450	7450	7450	7425
6/8/1998	7462	7500	7450	7500	7425
7/8/1998	7515	7500	7500	7500	7512.5
10/8/1998	7365	7450	7450	7450	7464
11/8/1998	7360	7300	7350	7300	7355
12/8/1998	7330	7300	7300	7300	7334
13/8/1998	7291	7300	7350	7300	7255
14/8/1998	7320	7183.33	7100	7188.33	7334
15/8/1998	7300	7300	7350	7300	7275
17/8/1998	7219	7300	7300	7300	7234
18/8/1998	7220	7183.33	7100	7100	7255
19/8/1998	7283	7183.33	7300	7300	7284
20/8/1998	7274	7183.33	7100	7188.33	7255
21/8/1998	7225	7183.33	7100	7100	7234
24/8/1998	6955	7183.33	7100	7100	6984
25/8/1998	6949	6850	6850	6850	6916
26/8/1998	6790	6850	6850	6850	6790
27/8/1998	6835	6775	6650	6775	6850
28/8/1998	6695	6850	6750	6750	6675
29/8/1998	6728	6750	6750	6750	6720

31/08/1998	6566	6775	6650	6650	6575
1/9/1998	6409	6450	6450	6450	6425
2/9/1998	6430	6450	6550	6550	6562.5
3/9/1998	6193	6450	6350	6350	6275
4/9/1998	6403.2	6450	6450	6450	6475
5/9/1998	6697.5	6450	6550	6550	6675
7/9/1998	6722.3	6750	6750	6750	6710
8/9/1998	6859.4	6775	6850	6850	6850
9/9/1998	6769.6	6850	6750	6750	6720
10/9/1998	6709.75	6775	6650	6650	6708
11/9/1998	6726.5	6775	6850	6775	6782
14/09/1998	6774.55	6775	6850	6775	6818
15/09/1998	6762	6775	6650	6775	6734
16/09/1998	6952.75	6775	6850	6850	6984
17/09/1998	6906	6850	6950	6850	6934
18/09/1998	6842	6850	6850	6850	6816
19/08/1998	7039	6850	6950	6950	7075
21/09/1998	6861	6850	6850	6850	6886
22/09/1998	6926	6850	6950	6850	6934
23/09/1998	6852	6850	6850	6850	6816
24/09/1998	6890	6850	6950	6850	6978
25/09/1998	6871	6850	6850	6850	6866
28/09/1998	6840	6850	6750	6750	6850
29/09/1998	6806	6850	6750	6850	6743
30/09/1998	6787	6850	6750	6750	6780
MSE		9737	7905	5437	1700

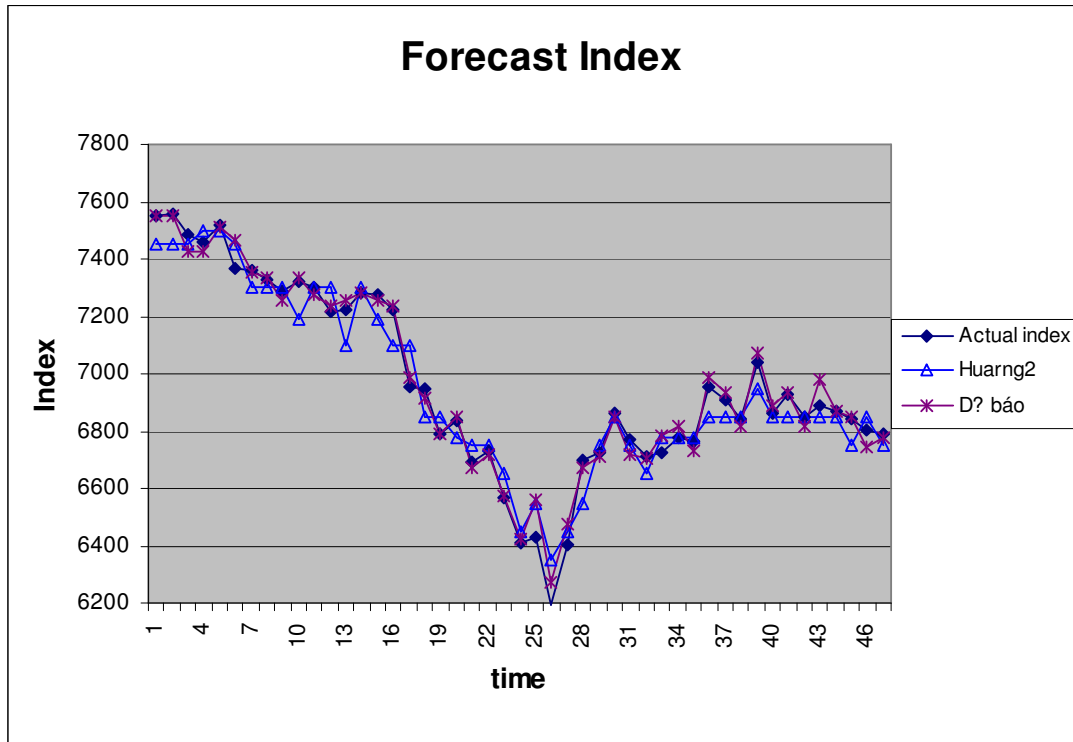
Cột cuối cùng là để tính sai số trung bình bình phương MSE theo công thức:

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

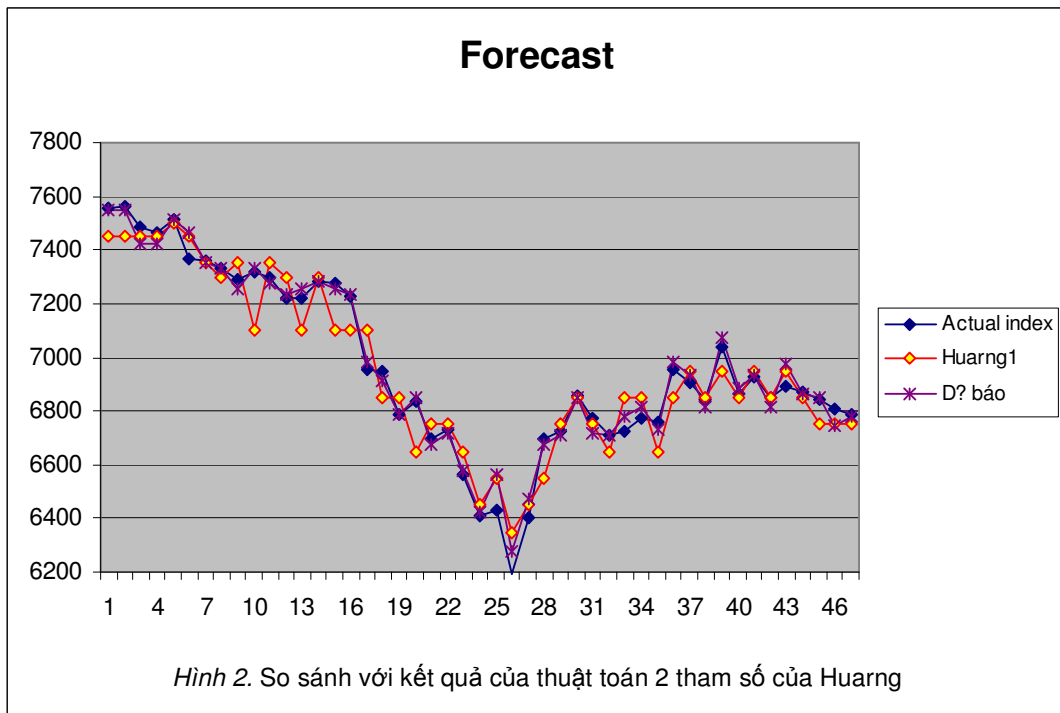
trong đó f_i là giá trị thực còn g_i là giá trị dự báo.

Ta có thể thấy rõ độ chính xác của phương pháp mới này chỉ bằng 1/3 phương pháp tốt nhất của Huarng.

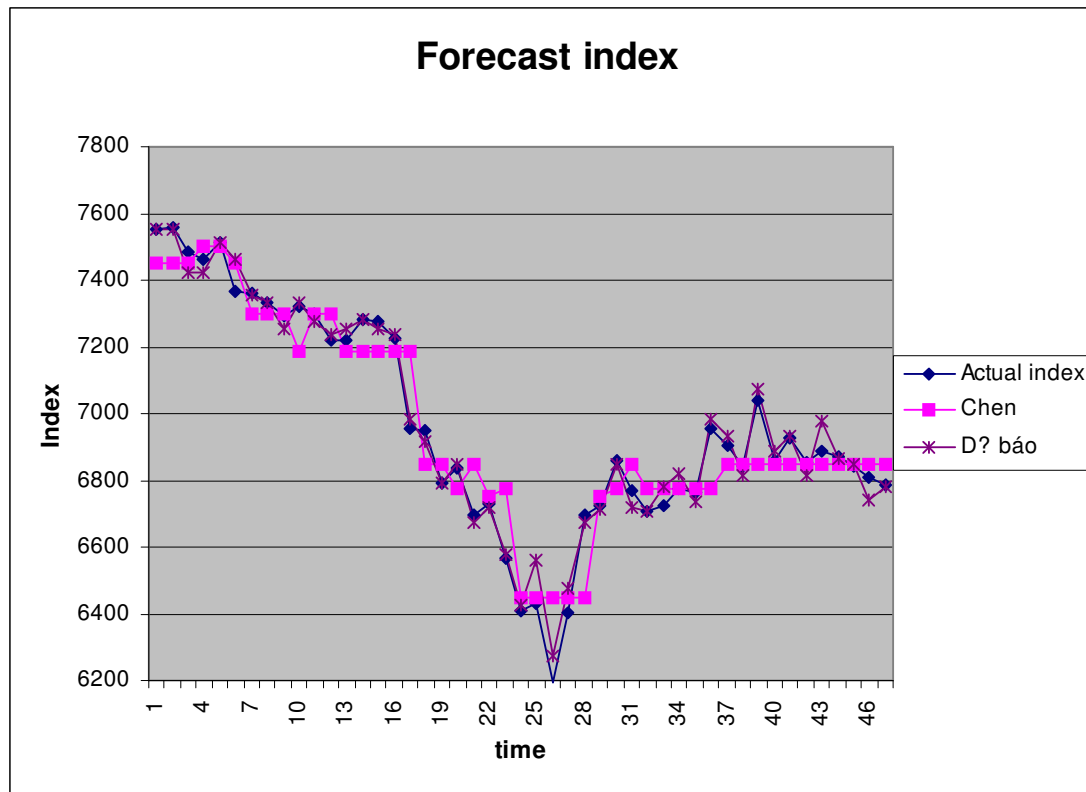
Sau đây là một số đồ thị so sánh các kết quả với nhau.



Hình 1. Đồ thị kết quả dự báo so sánh với thuật toán 3 tham số của Huarng



Hình 2. So sánh với kết quả của thuật toán 2 tham số của Huarng



Hình 3. So sánh kết quả với thuật toán Chen

Chú ý: Số liệu tính toán được làm tròn cho dễ theo dõi.

5. KẾT LUẬN

Bản báo cáo này đưa ra một thuật toán mới cho mô hình chuỗi thời gian mờ heuristic. Thuật toán này phát triển trên cơ sở thuật toán của Huarng [5] nhưng đã cải tiến khá nhiều, chủ yếu là chia lại khoảng từ tập nền và trong dự báo đã lấy thêm các thông tin từ tốc độ tăng giảm của giá trị chuỗi thời gian để xác định điểm dự báo trong khoảng. Các kết quả tính toán cho thấy độ chính xác của dự báo được tăng lên đáng kể so với các thuật toán đã xây dựng trước đây như thuật toán heuristic hai biến và ba biến của Huarng. Điều này thể hiện thông qua chỉ số trung bình bình phương (MSE) của phương pháp chỉ bằng 1/3 chỉ số trên của phương pháp ba biến của Huarng. Cũng có thể thấy được điều này thông qua các đồ thị kết quả đưa ra ở trên. Đường dự báo của chúng tôi đưa ra bám khá sát với giá trị thực tế. Điều này cho thấy dự báo theo phương pháp của chúng tôi khá tốt. Đạt được kết quả này nhờ phân chia khoảng hợp lí hơn và sử dụng được các thông tin thêm có sẵn của chuỗi thời gian để dự báo.

TÀI LIỆU THAM KHẢO

1. A. M. Abbasov, M. H. Mamedova - Application of Fuzzy Time Series to population forecasting, Transactions of Vienna University of Technology (2003) 545-552.

2. S. M. Chen - Forecasting Enrollments based on Fuzzy Time Series, Fuzzy set and system **81** (1996) 311-319.
3. S. M. Chen - Forecasting Enrollments based on high-order Fuzzy Time Series, Inter. Journal: Cybernetic and Systems **33** (2002) 1-16.
4. S. M. Chen, C. C. Hsu - A New Methods to Forecast Enrollments Using Fuzzy Time Series, Inter. Journal of Applied Science and Engineering **2** (3) (2004) 234-244.
5. K. Huarng - Heuristic models of fuzzy time series forecasting, Fuzzy sets and Systems **123** (2001) 369-386.
6. T. S. Lee, C.C. Chiu, F. C. Lin - Prediction of the Unemployment Rate Using Fuzzy Time Series with Box-Jenkins Methodology, International Journal of Fuzzy Systems **3** (4) (2001) 577-585.
7. Nguyễn Cát Hồ - Lí thuyết mờ và công cụ tính toán mềm, Hệ mờ, Mạng Noron và ứng dụng (Tuyển tập các bài giảng), NXB Khoa học và Kỹ thuật, 2001, pp. 37-77.
8. Nguyễn Công Điều, Nguyễn Văn Hùng - Ứng dụng chuỗi thời gian mờ trong dự báo chỉ số chứng khoán, Báo cáo khoa học Viện CNTT năm 2005.
9. M. Sah, K. Y. Degtiarev - Forecasting Enrollment Model Based on First Order Fuzzy Time Series, Transactions on Engineering, Computing and technology, Enfomatika **IV** (2004) 375-378.
10. Q. Song, B. S. Chissom - Fuzzy time series and its model, Fuzzy set and system **54** (1993) 269-277.
11. Q. Song, B.S. Chissom - Forecasting Enrollments with Fuzzy Time Series – Part I, Fuzzy set and system **54** (1993) 1-9.
12. Q. Song, B. S. Chissom - Forecasting Enrollments with Fuzzy Time Series – Part II, Fuzzy set and system **62** (1994) 1-8.

SUMMARY

A NEW ALGORITHM FOR FORECASTING STOCK INDEX BASED ON HEURISTIC FUZZY TIME SERIES MODEL

In this report a new improved method is proposed for time series forecasting based on results of fuzzy time series. Since Song and Chissom (1993) proposed the concept of fuzzy time series and first order fuzzy time series model, a number of researchers are developed this idea for time series forecasting. However the accuracy of these algorithms is necessary to improve. Therefore finding the effective algorithms for time series forecasting is priority.

In this article, we develop an algorithm for heuristic time series model in order to improve the forecast accuracy of time series. Our modification in this algorithm is exposed in two aspects: Partition the Universe of discourse into intervals so that distribution of value observations in these intervals is regular and using the information of first and second orders differences of time series for definition the law in defuzzification process. An example for forecasting Tapei index stock is used for demonstration of effectiveness of the algorithm.

Liên hệ với tác giả:

Nguyễn Công Điều,

Email: ncdieu@ioit.ac.vn