Tạp chí Khoa học và Công nghệ 54 (2) (2016) 161-177
DOI: 10.15625/0866-708X/54/2/5242

ỨNG DỤNG ĐẠI SỐ GIA TỬ TRONG DỰ BÁO CHUỖI THỜI GIAN MỜ

Nguyễn Cát Hồ[1], Nguyễn Công Điều[1, 2], Vũ Như Lân[1, 2, *]

[1]*Viện Công nghệ thông tin, Viện HLKHCNVN, 18 Hoàng Quốc Việt, Cầu Giấy, Hà Nội*

[2]*Trường Đại học Thăng Long, Đường Nghiêm Xuân Yêm, Đại Kim, Hoàng Mai, Hà Nội*

[*]Email: *vnlan@ioit.ac vn*

Đến Tòa soạn: 11/11/2014; Chấp nhận đăng: 15/11/2015

TÓM TẮT

Chuỗi thời gian mờ do Song & Chissom đưa ra trên tạp chí "Fuzzy Sets and Systems" năm 1993 đã được nghiên cứu rộng rãi trên thế giới cho mục đích dự báo. Tuy nhiên, độ chính xác của dự báo trên quan điểm xem xét chuỗi thời gian theo tiếp cận mờ của Song & Chissom còn chưa cao do phụ thuộc vào nhiều yếu tố. S. M. Chen (1996) đã đề xuất mô hình dự báo chuỗi thời gian mờ rất hiệu quả chỉ sử dụng các tính toán số học đơn giản. Sau đó mô hình này được nghiên cứu cải tiến trong nhiều ứng dụng dự báo và đã có được nhiều kết quả chính xác hơn.

Đại số gia tử là một tiếp cận mới được các tác giả N. C. Ho và W. Wechler xây dựng vào những năm 1990, 1992 hoàn toàn khác biệt so với tiếp cận mờ. Ở đây, đại số gia tử được sử dụng để mô phỏng biến ngôn ngữ và có được cấu trúc ngữ nghĩa. Phép mờ hóa và phép giải mờ được thay thế bằng phép ngữ nghĩa hóa và phép giải nghĩa tương ứng đơn giản hơn. Đại số gia tử dựa trên hệ mờ là một hướng đi mới, được ứng dụng lần đầu tiên trong điều khiển mờ năm 2008. Những ứng dụng của tiếp cận đại số gia tử cho một số bài toán cụ thể trong lĩnh vực công nghệ thông tin và điều khiển đã mang lại một số kết quả quan trọng khẳng định tính ưu việt của tiếp cận này so với tiếp cận mờ truyền thống. Tiếp tục những ứng dụng của đại số gia tử, bài báo này tập trung nghiên cứu vấn đề dự báo chuỗi thời gian mờ theo tiếp cận đại số gia tử.

Trong bài báo này, chúng tôi đưa ra một tiếp cận mới sử dụng đại số gia tử với khả năng cung cấp một mô hình tính toán hoàn toàn khác biệt so với tiếp cận mờ cho mô hình dự báo chuỗi thời gian mờ. Các kết quả thử nghiệm dự báo số sinh viên nhập học tại Đại học Alabama chứng minh rằng mô hình chuỗi thời gian mờ dựa trên đại số gia tử tốt hơn so với nhiều mô hình hiện có.

Từ khóa: tập mờ, nhóm quan hệ logic mờ, đại số gia tử, mô hình dự báo chuỗi thời gian mờ.

1. MỞ ĐẦU

Dự báo chuỗi thời gian là vấn đề luôn được nhiều nhà khoa học trên thế giới quan tâm nghiên cứu. Q. Song và B. S. Chissom [1] lần đầu tiên đã đưa ra quan niệm mới xem các giá trị thực định lượng trong chuỗi thời gian từ góc độ định tính. Từ đó chuỗi thời gian có thể xem như một biến ngôn ngữ và bài toán dự báo trở thành vấn đề dự báo các giá trị ngôn ngữ của biến

ngôn ngữ. Có thể coi đây là quan niệm mới về chuỗi thời gian có tính đột phá. Tuy nhiên mô hình tính toán nhóm quan hệ mờ [2, 3] quá phức tạp và do đó độ chính xác của dự báo không cao. Chen [4] đã thay đổi cách tính toán nhóm quan hệ mờ trong mô hình dự báo [2, 3] với các phép tính số học đơn giản hơn để thu được kết quả dự báo chính xác hơn. Nhiều nghiên cứu tiếp theo vẫn sử dụng phương pháp luận này và đã thu được nhiều kết quả quan trọng [4 - 6]. Ở Việt Nam, bài báo [7] là kết quả nghiên cứu đầu tiên về dự báo chuỗi thời gian mờ.

Các nghiên cứu trên thế giới chủ yếu tập trung giải quyết vấn đề nâng cao độ chính xác dự báo. Có thể thấy một số vấn đề sau đây ảnh hưởng đến độ chính xác dự báo chuỗi thời gian mờ:

a/ Mờ hóa các dữ liệu: Đây là vấn đề đòi hỏi phải có trực giác tốt để mô tả định tính chuỗi thời gian một cách hợp lí với các tham số đặc thù, qua đó cung cấp thông tin có giá trị cho quá trình dự báo sau này. Đặc tính quan trọng của phép mờ hóa là số lượng khoảng chia, độ dài khoảng chia và bậc của chuỗi thời gian mờ. Nếu số lượng khoảng chia quá ít, dự báo có thể có độ sai lệch lớn do chưa đủ thông tin. Nếu số lượng khoảng chia quá lớn, dự báo có thể mất hết ý nghĩa về tính mờ của giá trị ngôn ngữ khi không còn nhóm quan hệ mờ vì như vậy có thể tạo ra nhiều khoảng không chứa dữ liệu hoặc chỉ chứa 1 dữ liệu. Do đó vấn đề tìm ra khoảng chia tối ưu là một bài toán không dễ.

b/ Giải mờ: Đây là quá trình dự báo trên cơ sở phép mờ hóa trên đây và cần hướng đến dự báo tối ưu.

Vấn đề thứ nhất có thể thấy rõ qua một số nghiên cứu [8, 9]. Trong các nghiên cứu này rõ ràng rằng: số lượng khoảnh, độ dài khoảng và bậc của mô hình chuỗi thời gian mờ có ảnh hưởng đến độ chính xác của mô hình dự báo. Vấn đề nghiên cứu sâu hơn liên quan đến vấn đề tối ưu là xây dựng số lượng khoảng, độ dài khoảng và bậc của mô hình chuỗi thời gian mờ như thế nào để có dự báo tốt nhất cho các dữ liệu trong nhóm quan hệ mờ. Nhiều tác giả cũng đã nghiên cứu thành công vấn đề nêu trên [10, 11]. Vấn đề thứ nhất cũng liên quan đến khả năng tạo ra các tham số hỗ trợ cho vấn đề dự báo trên cơ sở tham số hóa mức độ tăng hay giảm giá trị của dữ liệu theo thời gian [12].

Vấn đề thứ hai có ảnh hưởng đến độ chính xác của dự báo là cách giải mờ tìm ra giá trị dự báo cho các dữ liệu từ nhóm quan hệ mờ trên cơ sở mờ hóa chuỗi thời gian ở trên.Tuy nhiên cách giải mờ phổ biến dựa trên 3 luật cơ bản [4]. Đặc biệt trong [6, 7] tìm ra một số tham số định hướng cho quá trình giải mờ và đã thu được một số kết quả khá tốt.

Tiếp cận đại số gia tử (ĐSGT) [13,14] là tiếp cận khác biệt so với tiếp cận mờ và đã có một số ứng dụng thể hiện rõ hiệu quả ứng dụng trong một số lĩnh vực công nghệ của tiếp cận này so với tiếp cận mờ truyền thống. Có thể kể đến một số lĩnh vực ứng dụng có triển vọng như điều khiển [15 - 19], công nghệ thông tin [20] và đặc biệt gần đây là lĩnh vực tính toán trên từ (Computing with words) [21]. Những kết quả ứng dụng mang tính ưu việt hơn trong một số lĩnh vực công nghệ khác nhau của tiếp cận ĐSGT so với tiếp cận mờ là minh chứng quan trọng cho tính đúng đắn của tiếp cận có xuất phát điểm khoa học dựa trên hệ tiên đề chặt chẽ làm cơ sở cho việc xây dựng ĐSGT- một cấu trúc toán học được nhúng vào tập các giá trị ngôn ngữ để biểu diễn các khái niệm mờ một cách tổng quát dựa trên ngữ nghĩa [22].

Tiếp tục những nghiên cứu ứng dụng trên đây, tiếp cận ĐSGT cũng cần được nghiên cứu thử nghiệm cho một lĩnh vực ứng dụng mới, đó là bài toán xây dựng mô hình dự báo chuỗi thời gian mờ đã được nhiều tác giả khác trên thế giới quan tâm hiện nay.

Bài báo được trình bày theo thứ tự sau đây: Sau mục MỞ ĐẦU là Mục 2 giới thiệu về mô hình dự báo chuỗi thời gian mờ và ứng dụng cho dự báo số sinh viên nhập học tại trường đại học Alabama của Song & Chissom [2, 3] và Chen [4]. Mục 3 trên cơ sở bài toán dự báo số sinh viên

nhập học của trường đại học Alabama, nêu một số nội dung quan trọng của ĐSGT cần thiết cho bài toán dự báo chuỗi thời gian mờ với bộ tham số hợp lí và so sánh với các phương pháp của Chen và các phương pháp cải tiến khác sử dụng chuỗi thời gian mờ bậc nhất với 7 khoảng chia. Mục 4 tiếp tục trình bày phương pháp dự báo số sinh viên nhập học của trường đại học Alabama trên cơ sở tiếp cận ĐSGT trong điều kiện phép ngữ nghĩa hóa phi tuyến, phép giải nghĩa phi tuyến với các tham số tối ưu và so sánh với một số phương pháp dự báo cải tiến mới sử dụng bậc cao, số khoảng chia lớn hơn 7 và một số mô hình dự báo chuỗi thời gian mờ tối ưu hiện nay. Độ chính xác dự báo của các phương pháp trên được đánh giá qua sai số trung bình bình phương MSE (Mean Square Error), từ đó có thể thấy rõ tính ưu việt của tiếp cận ĐSGT so với tiếp cận mờ.

2. MÔ HÌNH DỰ BÁO CHUỖI THỜI GIAN MỜ

2.1 Một số khái niệm cơ bản của mô hình dự báo chuỗi thời gian mờ

Mô hình chuỗi thời gian mờ lần đầu tiên được Song và Chissom đưa ra [1 - 3] và được Chen cải tiến [4, 23, 24] để có thể xử lí bằng các phép tính số học đơn giản hơn nhưng chính xác hơn phù hợp với các ứng dụng dự báo chuỗi thời gian mờ. Có thể tóm lược qua một số khái niệm cơ bản sau đây:

Định nghĩa 2.1: Chuỗi thời gian mờ

Giả sử Y(t), (t = ..., 0, 1, 2, ...), là tập các số thực và cũng là tập nền trên đó xác định các tập mờ $f_i(t)$, (I = 1,2 , ...). Biến t là thời gian. Nếu F(t) là một chuỗi các tập mờ của $f_i(t)$, (I = 1, 2,...), thì F(t) được gọi là chuỗi thời gian mờ trên Y(t), (t = ..., 0, 1, 2,. ..).

Định nghĩa 2.2: Quan hệ mờ

Nếu tồn tại quan hệ mờ R(t−1, t), sao cho F(t) = F(t−1)*R(t−1, t), trong đó dấu * kí hiệu toán tử nào đó, thì F(t) được suy ra từ F(t−1). Quan hệ giữa F(t) và F(t−1) được xác định bằng kí hiệu:

$$F(t-1) \rightarrow F(t) \tag{2.1}$$

Ví dụ về toán tử * có thể là phép kết hợp MaxMin [2] hoặc MinMax [3] hay phép tính số học [4] . Nếu F (t−1) = Ai và F (t) = Aj, quan hệ logic giữa F (t) và F(t−1) được kí hiệu bằng Ai→Aj, trong đó Ai là vế trái và Aj là vế phải của quan hệ mờ mô tả tập mờ dự báo.

Định nghĩa 2.3: Quan hệ mờ bậc n

Giả sử F(t) là chuỗi thời gian mờ. Nếu F(t) được suy ra từ F(t−1), F(t−2),..., F(t−n), thì quan hệ mờ này được biểu diễn bằng biểu thức:

$$F(t-n),..., F(t-2), F(t-1) \rightarrow F(t) \tag{2.2}$$

và được gọi là chuỗi thời gian mờ bậc n. .

Định nghĩa 2.4: Chuỗi thời gian mờ dừng

Giả sử F(t) được suy ra từ F(t−1) và được kí hiệu bằng F(t−1) → F(t), khi đó quan hệ mờ giữa F(t) và F(t−1) được mô tả bằng phương trình:

$$F(t) = F(t-1)*R(t-1, t). \tag{2.3}$$

Quan hệ mờ R thể hiện mô hình bậc nhất của F(t). Nếu R(t−1, t) không phụ thuộc t, sao cho với mọi t1 và t2 khác nhau, R(t1, t1−1) = R(t2,t2−1), thì F(t) được gọi là chuỗi thời gian mờ dừng, còn lại được gọi là chuỗi thời gian mờ không dừng.

Định nghĩa 2.5: Nhóm quan hệ mờ (NQM)

Các quan hệ mờ với cùng một tập mờ bên vế trái có thể đưa vào một nhóm gọi là nhóm quan hệ mờ hay nhóm quan hệ logic mờ.

Giả sử có các quan hệ mờ sau:

$$Ai \rightarrow Aj1;\ Ai \rightarrow Aj2;....;\ Ai \rightarrow Ajn.$$

Các quan hệ mờ trên có thể đưa vào một nhóm được kí hiệu như sau:

$$Ai \rightarrow Aj1, Aj2, ..., Ajn \qquad (2.4)$$

tập mờ Ajk (k = 1, 2, …, n) chỉ được xuất hiện 1 lần bên vế phải.

2.2 Mô hình dự báo Song và Chissom

Mô hình dự báo chuỗi thời gian mờ lần đầu tiên được Song và Chissom đưa ra vào năm 1993 [1 - 3] và được ứng dụng để dự báo số sinh viên nhập học tại trường Đại học Alabama với dữ liệu lịch sử qua 22 năm kể từ năm 1971 đến 1992 như trong Bảng 2.1 sau đây:

Bảng 2.1. Số sinh viên nhập học tại trường đại học Alabama từ 1971 đến 1992.

Năm	Số sinh viên nhập học	Năm	Số sinh viên nhập học
1971	13055	1982	15433
1972	13563	1983	15497
1973	13867	1084	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

Chuỗi thời gian lần đầu tiên được xem xét dưới góc độ biến ngôn ngữ và bài toán dự báo đã có được một cách nhìn hoàn toàn mới trên quan điểm lí thuyết tập mờ. Mô hình dự báo đầu tiên là mô hình dự báo chuỗi thời gian dừng [2, 3] và được triển khai qua các bước sau đây:

Bước 1. Xác định tập nền

Bước 2. Chia miền xác định của tập nền thành những khoảng bằng nhau.

Bước 3. Xây dựng các tập mờ trên tập nền

Bước 4. Mờ hóa chuỗi dữ liệu

Bước 5. Xác định các quan hệ mờ

Bước 6. Dự báo bằng phương trình $Ai = A_{i-1} * R$, ở đây kí hiệu * là toán tử max-min

Bước 7. Giải mờ các kết quả dự báo.

Trong bước 5, quan hệ mờ R được xác định bằng biểu thức Ri = As ^{T}xAq, với mọi quan hệ mờ k, As $\rightarrow$Aq, R = $U_{i=1,k}$ Ri (2.5)

Ở đây x là toán tử min, T là phép chuyển vị và U là phép hợp.

2.3 Mô hình dự báo Chen

Do mô hình dự báo chuỗi thời gian mờ của Song & Chissom khá phức tạp trong bước 5 và bước 6, vì vậy Chen [4] đã cải tiến cách tính toán sao cho chính xác hơn cho các mô hình dự báo chuỗi thời gian chỉ sử dụng các phép tính số học đơn giản trên cơ sở thông tin từ các quan hệ mờ và nhóm quan hệ mờ theo các bước sau đây:

Bước 1. Chia miền xác định của tập nền thành những khoảng bằng nhau.

Bước 2. Xây dựng các tập mờ trên tập nền.

Bước 3. Mờ hóa chuỗi dữ liệu.

Bước 4. Xác định các quan hệ mờ.

Bước 5. Tạo lập nhóm quan hệ mờ.

Bước 6. Giải mờ đầu ra dự báo.

3. MÔ HÌNH DỰ BÁO THEO TIẾP CẬN ĐẠI SỐ GIA TỬ

Đại số gia tử cung cấp một mô hình xử lí các đại lượng không chắc chắn khá hiệu quả cho nhiều bài toán ứng dụng. Có thể thấy rõ rằng các giá trị ngôn ngữ với ngữ nghĩa vốn có thứ tự chặt chẽ trong biến ngôn ngữ đã được mô tả bằng một cấu trúc đại số gia tử [13, 14], từ đó tạo ra môi trường tính toán, suy luận tốt cho nhiều ứng dụng.

Gọi AX = (X, G, C, H, $\leq$) là một cấu trúc đại số, với X là tập nền của AX; G = {*c-*, *c+*} là tập các phần tử sinh; C = {**0**, **W**, **1**}, trong đó **0**, **W** và **1** tương ứng là những phần tử đặc trưng cận trái (tuyệt đối nhỏ), trung hòa và cận phải (tuyệt đối lớn); H là tập các toán tử một ngôi được gọi là các gia tử; $\leq$ là biểu thị quan hệ thứ tự trên các giá trị ngôn ngữ. Gọi H^- là tập hợp các gia tử âm và H^+ là tập hợp các gia tử dương của AX.

Kí hiệu $H^- = \{h_{-1}, h_{-2}, \ldots h_{-q}\}$, trong đó $h_{-1} < h_{-2} < \ldots < h_{-q}$ và $H^+ = \{h_1, h_2, \ldots, h_p\}$, trong đó $h_1 < h_2 < \ldots < h_p$.

Định nghĩa 3.1: Độ đo tính mờ

fm: X $\rightarrow$ [0, 1] gọi là độ đo tính mờ nếu thỏa mãn các điều kiện sau::

$$fm(c-)+fm(c+) = 1 \text{ và } \sum_{h\in H} fm(hx) = fm(x), \text{ với } \forall x \in X. \quad (3.1)$$

Với các phần tử *0*, *W* và *1*,

$$fm(0) = fm(W) = fm(1) = 0. \quad (3.2)$$

Và với $\forall x,y \in X$, $\forall h \in H$,

$$\frac{fm(hx)}{fm(x)} = \frac{fm(hy)}{fm(y)} \quad (3.3)$$

Đẳng thức (3.3) không phụ thuộc vào các phần tử x, y và do đó ta có thể kí hiệu là $\mu(h)$ và đây là độ đo tính mờ của gia tử *h*. Tính chất của *fm*(x) và $\mu(h)$ như sau:

$$fm(hx) = \mu(h)fm(x), \forall x \in X \quad (3.4)$$

$$\sum_{i=-q, i\neq 0}^{p} fm(h_i c) = fm(c), \text{ với c}\in \{c\text{-}, c+\} \tag{3.5}$$

$$\sum_{i=-q, i\neq 0}^{p} fm(h_i x) = fm(x) \tag{3.6}$$

$$\sum_{i=-1}^{-q} \mu(h_i) = \alpha \text{ và } \sum_{i=1}^{p} \mu(h_i) = \beta, \text{ với } \alpha, \beta > 0 \text{ và } \alpha+\beta = 1 \tag{3.7}$$

Định nghĩa 3.2: Hàm dấu

Hàm *Sign*: X→{-1, 0, 1} là một ánh xạ được gọi là hàm dấu với h, h'∈H và c ∈ {c-, c+} trong đó:

$$Sign(c\text{-}) = -1,\ Sign(c+) = +1; \tag{3.8}$$

$$Sign(hc) = -Sign(c), \text{ nếu h là âm đối với c;} \tag{3.9}$$

$$Sign(hc) = +Sign(c), \text{ nếu h là dương đối với c;} \tag{3.10}$$

$$Sign(h'hx) = -Sign(hx), \text{ nếu h'hx} \neq \text{hx và h' là âm đối với h;} \tag{3.11}$$

$$Sign(h'hx) = +Sign(hx), \text{ nếu h'hx} \neq \text{ hx và h' là dương đối với h;} \tag{3.12}$$

$$Sign(h'hx) = 0 \text{ nếu h'hx = hx.} \tag{3.13}$$

Gọi *fm* là một độ đo tính mờ trên X, ánh xạ ngữ nghĩa định lượng v: X → [0, 1], được sinh ra bởi *fm* trên X, được xác định như sau:

$$v(W) = \theta = fm(c^-), \tag{3.14}$$

$$v(c^-) = \theta - \alpha fm(c^-) = \beta fm(c^-), \tag{3.15}$$

$$v(c^+) = \theta + \alpha fm(c^+) = 1 - \beta fm(c^+) \tag{3.16}$$

$$v(h_j x) = v(x) + sign(h_j x)\{\sum_{i=sign(j)}^{J} fm(h_i x) - \omega(h_j x) fm(h_j x)\} \tag{3.17}$$

với

$$\omega(h_j x) = \frac{1}{2}[1 + Sign(h_j x) sign(h_p h_j x)(\beta - \alpha)] \in \{\alpha, \beta\}, \text{ j} \in \text{[-q^p], j} \neq 0. \tag{3.18}$$

Để thuận tiện cho việc biểu diễn ngữ nghĩa của các giá trị ngôn ngữ [15], giả sử rằng miền tham chiếu thông thường của các biến ngôn ngữ X là đoạn [a, b] còn miền tham chiếu ngữ nghĩa Xs là đoạn $[a_s, b_s]$ ($0 \leq a_s < b_s \leq 1$). Việc chuyển đổi tuyến tính từ [a, b] sang $[a_s, b_s]$ được gọi là phép ngữ nghĩa hóa tuyến tính (linear semantization) còn việc chuyển ngược lại từ đoạn $[a_s, b_s]$ sang [a, b] được gọi là phép giải nghĩa tuyến tính(linear desemantization). Trong nhiều ứng dụng của ĐSGT đã sử dụng miền ngữ nghĩa là đoạn $[a_s = 0, b_s = 1]$, khi đó phép ngữ nghĩa hóa tuyến tính được gọi là phép chuẩn hóa (linear Semantization = Normalization) và phép giải nghĩa tuyến tính được gọi là phép giải chuẩn (Linear Desemantization = Denormalization). Nhiều ứng dụng của ĐSGT trong nhiều lĩnh vực khoa học đòi hỏi mở rộng không gian tham số trong các phép ngữ nghĩa hóa và phép giải nghĩa để có nhiều tham số lựa chọn mềm dẻo hơn nữa. Điều này chỉ có thể có được khi mở rộng phép ngữ nghĩa hóa và phép giải nghĩa từ tuyến tính đến phi tuyến. Như vậy có thể biểu diễn các dạng của phép ngữ nghĩa hóa như sau:

$$\text{Linear Semantization (x)} = x_s = a_s + (b_s - a_s)(x - a)/(b - a) \tag{3.19a}$$

$$\text{Normalization (x)} = x_s = (x - a)/(b - a) \tag{3.19b}$$

$$\text{Nonlinear Semantization } (x) = f(x_s, sp) \tag{3.19c}$$

Với điều kiện: $0 \le f(x_s,sp) \le 1$ và $f(x_s=0,sp) = 0$ và $f(x_s=1,sp) = 1$

Có thể thấy rằng, các hàm f(.) được chọn tùy theo từng ứng dụng và là các hàm liên tục, đồng biến. Ví dụ có thể chọn f(.) phi tuyến theo x_s thể hiện qua $f(x_s,sp)$ như sau:

$$\text{Nonlinear Normalization } (x) = f(x_s,sp) = sp.x_s(1-x_s)+x_s \tag{3.19d}$$

Tương tự phép ngữ nghĩa hóa, các dạng của phép giải nghĩa được xác định như sau:

$$\text{Linear Desemantization } (x_s) = x = a + (b - a)\,(x_s - a_s) / (b_s - a_s) \tag{3.20a}$$

$$\text{Denormalization } (x_s) = x = a + (b - a)x_s \tag{3.20b}$$

$$\text{Nonlinear Desemantization } (x_s) = g(x, dp) \tag{3.20c}$$

Với điều kiện: $a \le g(x, dp) \le b$ và $g(x = a, dp) = a$ và $g(x = b, dp) = b$

Các hàm g(.) cũng được chọn tùy theo từng ứng dụng và là các hàm liên tục, đồng biến. Ví dụ có thể chọn g(.) phi tuyến theo x thể hiện qua Denormalization $(f(x_s,sp))$ như sau:

$$\text{Nonlinear Denormalization } (x_s) = dp((\text{Denormalization } (f(x_s,sp))-a).$$
$$(b - \text{Denormalization } (f(x_s,sp))) / (b-a) + \text{Denormalization } (f(x_s,sp)) \tag{3.20d}$$

trong đó

$$\text{Denormalization } (f(x_s,sp)) = (sp.x.(1-x)+x).(b-a) + a \tag{3.20d1}$$

Hàm $f(x_s,sp)$ là hàm biểu diễn phép ngữ nghĩa hóa phi tuyến và hàm g(x.dp) là hàm biểu diễn phép giải nghĩa phi tuyến chưa được sử dụng trong các ứng dụng của ĐSGT, trong đó $sp \in [-1\ 1]$ là tham số ngữ nghĩa hóa, $dp \in [-1\ 1]$ là tham số giải nghĩa. Khi sp = dp = 0; tính phi tuyến bị loại bỏ và biểu thức (3.19d) trở thành (3.19b) và (3.20d) trở thành (3.20b).

Cho trước độ đo tính mờ của các gia tử $\mu(h)$ và các giá trị độ đo tính mờ của các phần tử sinh $fm(c^-)$, $fm(c^+)$ và θ là phần tử trung hoà (neutral). Khi đó mô hình tính toán của ĐSGT được xây dựng trên cơ sở các biểu thức từ (3.1) đến (3.20) được kích hoạt và thực tế đã được sử dụng hiệu quả trong rất nhiều ứng dụng. Phép mờ hóa và phép giải mờ trong tiếp cận mờ được thay thế tương ứng bằng phép ngữ nghĩa hóa và phép giải nghĩa trong tiếp cận ĐSGT. Hệ luật được thể hiện bằng siêu mặt làm cơ sở cho quá trình suy luận xấp xỉ. Một lưu ý quan trọng của quá trình tính toán trong tiếp cận ĐSGT là cần xác định các tham số ban đầu như độ đo tính mờ của các phần tử sinh và độ đo tính mờ của các gia tử trong biến ngôn ngữ một cách thích hợp dựa trên cơ sở phân tích ngữ nghĩa của miền ngôn ngữ trong từng bài toán ứng dụng cụ thể. Khi đó mô hình tính toán của tiếp cận ĐSGT sẽ cho các kết quả hợp lí trong các ứng dụng.

Đối với mô hình dự báo chuỗi thời gian mờ của Song & Chissom và Chen, có thể thấy rõ hai giai đoạn quan trọng được các tác giả sử dụng dựa trên tiếp cận mờ. Đầu tiên là giai đoạn có nội dung của phép mờ hóa và những vấn đề liên quan bao gồm bước 1 đến bước 5. Nếu giai đoạn mờ hóa cung cấp những thông tin định tính hợp lí thì các quan hệ mờ hoặc nhóm quan hệ mờ sẽ tạo ra khả năng dự báo với độ chính xác cao cho các dữ liệu. Giai đoạn tiếp theo là giai đoạn có nội dung của phép giải mờ (bước 6 và bước 7 của mô hình Song & Chissom hoặc bước 6 của mô hình Chen). Đây là giai đoạn tìm ra kết quả dự báo dựa trên cơ sở các bước của giai đoạn mờ hóa.

Dựa trên các phân tích trên đây, rõ ràng rằng: hoàn toàn có thể thay thế tiếp cận mờ với hai giai đoạn có nội dung của phép mờ hóa và phép giải mờ trong các mô hình của Song & Chissom hoặc Chen bằng tiếp cận ĐSGT cũng với hai giai đoạn có nội dung của phép ngữ nghĩa hóa và phép giải nghĩa tương ứng. Như vậy có thể xây dựng được mô hình dự báo chuỗi thời gian mờ

tương tự như mô hình Chen nhưng không sử dụng tập mờ mà dựa trên tiếp cận ĐSGT với mô hình tính toán qua các biểu thức (3.1), (3.2), ...(3.20) như sau:

Bước 1. Xác định tập nền và chia miền xác định của tập nền thành những khoảng bằng nhau.

Bước 2. Xây dựng các nhãn ngữ nghĩa (giá trị ngôn ngữ theo tiếp cận ĐSGT).

Bước 3. Ngữ nghĩa hóa phi tuyến chuỗi dữ liệu.

Bước 4. Xác định các quan hệ ngữ nghĩa theo nhãn ngữ nghĩa .

Bước 5. Tạo lập nhóm quan hệ ngữ nghĩa theo nhãn ngữ nghĩa.

Bước 6. Giải nghĩa phi tuyến đầu ra dự báo.

Các bước trên đây tương tự với các bước dự báo trong mô hình Chen nhưng trong tiếp cận ĐSGT không sử dụng tập mờ mà dùng ngữ nghĩa định lượng mô tả định lượng giá trị ngôn ngữ. Ở đây, phép mờ hóa được thay bằng phép ngữ nghĩa hóa, quan hệ mờ được thay bằng quan hệ ngữ nghĩa và nhóm quan hệ mờ được thay bằng nhóm quan hệ ngữ nghĩa. Cuối cùng phép giải mờ được thay bằng phép giải nghĩa.

Bài toán được chọn để so sánh và làm rõ hiệu quả dự báo của mô hình trên là bài toán dự báo số sinh viên nhập học tại trường Alabama do Song & Chissom [2, 3] và Chen [4] đặt ra đầu tiên để nghiên cứu mô hình chuỗi thời gian mờ trên quan điểm biến ngôn ngữ. Từ đó có thể mô tả định tính số lượng sinh viên nhập học tại trường Đại học Alabama từ các số liệu lịch sử có từ năm 1971 đến năm 1992 và đưa số liệu này vào mô hình dự báo chuỗi thời gian mờ. Đây cũng là bài toán cho đến nay vẫn được Chen [8, 9] và nhiều tác giả trên thế giới quan tâm nghiên cứu cải tiến [25 - 27].

Các bước tính toán dựa trên ĐSGT cụ thể như sau:

Bước 1: Xác định tập nền, chia miền xác định của tập nền thành những khoảng bằng nhau.

Tập nền U được chọn tương tự mô hình Chen có khoảng xác định: [Dmin−D1, Dmax + D2] với Dmin và Dmax là số sinh viên nhập học thấp nhất và cao nhất theo dữ liệu lịch sử nhập học của trường. Cụ thể Dmin = 13055 và Dmax = 19337. Các biến D1 và D2 là các số dương được chọn sao cho khoảng [Dmin−D1, Dmax+D2] bao được hoàn toàn số sinh viên nhập học thấp nhất và cao nhất trong tương lai. Sử dụng cách chọn của Chen [4], D1 = 55 và D2 = 663, như vậy U= [13000, 20000]. Khoảng xác định tập nền U được Chen [4] và nhiều tác giả khác [15, 29, 32, 38] chia thành 7 khoảng bằng nhau u1, u2, u3, u4, u5, u6 và u7. Trong đó u1 = [13000, 14000], u2 = [14000, 15000], u3 = [15000, 16000], u4 = [16000, 17000], u5 = [17000, 18000], u6 = [18000, 19000] và u7 = [19000, 20000].

Bước 2: Xây dựng các nhãn ngữ nghĩa (giá trị ngôn ngữ không biểu diễn dưới dạng tập mờ) của tiếp cận ĐSGT) trên tập nền. Để có thể dễ theo rõi và so sánh với các bước dự báo trong mô hình Chen, ở đây sử dụng một số kí hiệu tương tự những kí hiệu Chen đã sử dụng nhưng với ý nghĩa của tiếp cận ĐSGT. Giả sử A1, A2,..., Ak là các nhãn ngữ nghĩa được gán cho các khoảng u1, u2,...uk, k là số khoảng trên tập nền. Khác với tập mờ trong nghiên cứu của Chen, các nhãn ngữ nghĩa ở đây được xây dựng từ các phần tử sinh c-, c+ với các gia tử h □ H tạo thành các giá trị ngôn ngữ của biến ngôn ngữ "số sinh viên nhập học ". Khi đó các nhãn ngữ nghĩa A1, A2,..., Ak có dạng sau đây: A1= hA1c; A2 = hA2c;...; Ak = hAkc, trong đó hAi, (i = 1,2,...k) là chuỗi gia tử tác động lên c với c ∈ {c-, c+}.

Trong bài toán dự báo số sinh viên nhập học tại trường Đại học Alabama, Chen sử dụng các giá trị ngôn ngữ A1 = (not many), A2 = (not too many), A3 = (many), A4 = (many many), A5 = (very many), A6 = (too many) và A7 = (too many many). Trong bài toán dự báo này theo

tiếp cận ĐSGT, sử dụng 2 gia tử "very"và "little" tác động lên 2 phần tử sinh "small"và "large"để tạo ra 7 nhãn ngữ nghĩa tương ứng với 7 giá trị ngôn ngữ của Chen như sau: A1 = (very small), A2 = (small), A3 = (little small), A4 = (midle), A5 = (little large), A6 = (large) và A7 = (very large).

Bước 3: Ngữ nghĩa hóa chuỗi dữ liệu.

Để xác định ngữ nghĩa định lượng cho các nhãn ngữ nghĩa A1, A2,...,A7 ở bước 2, cần chọn trước độ đo tính mờ của các gia tử $\mu(very)$, $\mu(little)$ và giá trị độ đo tính mờ của phần tử sinh $fm(c) = \theta$ với θ là phần tử trung hoà được cho trước. Nếu các nhãn ngữ nghĩa được tạo thành chỉ từ 1 gia tử dương và 1 gia tử âm ví dụ gia tử dương "*very*" và gia tử âm "*little* " tác động lên các phần tử sinh "*large*" hoặc "*small*" như trên, thì $\mu(little) = \alpha$ và $\mu(very) = 1 - \alpha = \beta$. Như vậy ngữ nghĩa định lượng của các nhãn ngữ nghĩa sẽ chỉ phụ thuộc vào các tham số của ĐSGT α, θ và hoàn toàn được xác định sau khi thay các giá trị α, θ vào các phương trình tính toán ngữ nghĩa định lượng từ (3.14) đến (3.18). Cụ thể là 7 giá trị ngữ nghĩa định lượng của 7 nhãn ngữ nghĩa A1,A2, ...A7 được gán tương ứng cho 7 khoảng u1, u2,..., u7 có dạng tham số hóa sau đây:

$$v(\text{very small}) = \theta(1-\alpha)(1-\alpha) \quad (3.21)$$

$$v(\text{small}) = \theta(1-\alpha) \quad (3.22)$$

$$v(\text{little small}) = \theta(1-\alpha+\alpha^2) \quad (3.23)$$

$$v(\text{midle}) = \theta \quad (3.24)$$

$$v(\text{little large}) = \theta+\alpha(1-\theta)(1-\alpha) \quad (3.25)$$

$$v(\text{large}) = \theta+(1-\theta)\alpha \quad (3.26)$$

$$v(\text{very large}) = \theta+\alpha(1-\theta)(2-\alpha) \quad (3.27)$$

Nếu chọn trước $\alpha = 0.5$ và $\theta = 0.5$, thì các phương trình từ (3.21) đến (3.27) trở thành:

$$v(\text{very small}) = 0.125 \quad (3.28)$$

$$v(\text{small}) = 0.25 \quad (3.29)$$

$$v(\text{little small}) = 0.375 \quad (3.30)$$

$$v(\text{midle}) = 0.5 \quad (3.31)$$

$$v(\text{little large}) = 0.625 \quad (3.32)$$

$$v(\text{large}) = 0.75 \quad (3.33)$$

$$v(\text{very large}) = 0.875 \quad (3.34)$$

Kí hiệu: SA = Semantization (A) là giá trị ngữ nghĩa định lượng theo nhãn ngữ nghĩa A, khi đó: SA1 = v(very small); SA2 = v(small); SA3 = v(little small); SA4 = v(midle); SA5 = v(little large); SA6 = v(large) và SA7 = v(very large) là các giá trị ngữ nghĩa định lượng theo các tham số được chọn trước α, θ. Khi đó dễ dàng thấy rằng:

$$SA1 < SA2 < SA3 < SA4 < SA5 < SA6 < SA7 \quad (3.35)$$

Tương tự như trên, có thể xây dựng các công thức tính toán các giá trị ngữ nghĩa định lượng theo các nhãn ngữ nghĩa khi có nhiều gia tử tác động lên phần tử sinh.

Biểu thức (3.35) thể hiện rõ những tính chất quan trọng sau đây:

1. Thứ tự ngữ nghĩa luôn được đảm bảo.
2. Các nhãn ngữ nghĩa Ai có giá trị ngữ nghĩa định lượng SAi và luôn có quan hệ ngữ nghĩa với nhau thông qua bộ tham số của ĐSGT α, θ, $\mu(h_{Ai})$, I = 1, 2,...

Như vậy, trong các ứng dụng cụ thể của tiếp cận ĐSGT, ảnh hưởng của bộ tham số mang tính hệ thống. Có nghĩa là tất cả các giá trị ngôn ngữ trong biến ngôn ngữ đều chịu ảnh hưởng bởi bộ tham số của ĐSGT. Những tính chất trên đây tạo ra sự khác biệt giữa tiếp cận ĐSGT và tiếp cận mờ. Có thể thấy rằng: trong tiếp cận mờ, các giá trị ngôn ngữ sử dụng tập mờ của biến ngôn ngữ hoàn toàn không có ràng buộc với nhau. Sự khác biệt này đã đưa đến hiệu quả cao trong nhiều ứng dụng của tiếp cận ĐSGT.

Bước 4: Xác định các quan hệ ngữ nghĩa theo nhãn ngữ nghĩa.

Các quan hệ ngữ nghĩa được xác định trên cơ sở các dữ liệu lịch sử. Nếu đặt chuỗi thời gian mờ F(t-1) là Ak có ngữ nghĩa định lượng SAk và F(t) là Am có ngữ nghĩa định lượng SAm, thì Ak có quan hệ với Am và dẫn đến SAk có quan hệ với SAm. Quan hệ này được gọi là quan hệ ngữ nghĩa theo nhãn ngữ nghĩa và được kí hiệu là:

$$SAk \rightarrow SAm \text{ hoặc Semantization (Aj)} \rightarrow \text{Semantization (Ak)} \quad (3.36)$$

Trong bài toán dự báo số sinh nhập học tại trường Alabama, ở đây Ak là nhãn ngữ nghĩa mô tả số sinh viên nhập học của năm hiện tại với ngữ nghĩa định lượng SAk, Am là nhãn ngữ nghĩa mô tả số sinh viên nhập học của năm tiếp theo với ngữ nghĩa định lượng SAm.

Như vậy, trên cơ sở số liệu của Chen [4], có thể xác định được các quan hệ ngữ nghĩa theo nhãn ngữ nghĩa (kể cả số lần trùng nhau) sau đây:

SA1 → SA1 (trùng nhau 2 lần); SA1 → SA2;

SA2 → SA3; SA3 → SA3 (trùng nhau 7 lần);

SA3 → SA4 (trùng nhau 2 lần); SA4 → SA4 (trùng nhau 2 lần);

SA4 → SA3; SA4 → SA6; SA6 → SA6; SA6 → SA7;

SA7 → SA7 và SA7 → SA6. (3.37)

Bước 5: Tạo lập nhóm quan hệ ngữ nghĩa theo nhãn ngữ nghĩa.

Nếu một ngữ nghĩa định lượng (vế trái (3.37)) có quan hệ với nhiều ngữ nghĩa định lượng (vế phải (3.37)), thì vế phải được chập lại thành một nhóm. Quan hệ được lập theo nhóm như vậy được gọi là nhóm quan hệ ngữ nghĩa (NQHNN). Như vậy từ (3.37) nhận được các NQHNN sau đây:

Nhóm 1: SA1 → (SA1, SA1, SA2)

Nhóm 2: SA2 → (SA3)

Nhóm 3: SA3 → (SA3, SA3, SA3, SA3, SA3, SA3, SA3, SA4, SA4)

Nhóm 4: SA4 → (SA4, SA4, SA3, SA6)

Nhóm 5: SA6 → (SA6, SA7)

Nhóm 6: SA7 → (SA7, SA6)

Bước 6: Giải nghĩa đầu ra dự báo.

Giả sử số sinh viên nhập học tại năm (t-1) của chuỗi thời gian mờ F(t-1) được ngữ nghĩa hóa theo (3.19) là SAj, khi đó đầu ra dự báo của F(t) hay số sinh viên nhập học dự báo tại năm t được xác định theo các nguyên tắc (luật) sau đây:

1. Nếu tồn tại quan hệ 1-1 trong nhóm quan hệ ngữ nghĩa theo nhãn ngôn ngữ Aj như sau:

SAj → SAk, theo (3.19d): Nonlinear Semantization (Aj) → Nonlinear Semantization (Ak)

Thì đầu ra dự báo được tính theo (3.20d): DSAj → Nonlinear Desemantization (SAk) trên khoảng giải nghĩa uk được chọn sao cho bao được uk và thuộc khoảng xác định của tập nền chuỗi thời gian mờ [Dmin−D1, Dmax+D2].

2. Nếu SAk là trống, SAj → Ø,

Thì đầu ra dự báo được tính theo (3.20d): DSAj → Nonlinear Desemantization (Ø) trên khoảng giải nghĩa được chọn sao cho bao được uj và thuộc khoảng xác định của tập nền chuỗi thời gian mờ [Dmin−D1, Dmax+D2].

3. Nếu tồn tại quan hệ 1-nhiều trong nhóm quan hệ ngữ nghĩa (kể cả quan hệ trùng) theo nhãn ngôn ngữ Aj: SAj → (SAi,SAk,..., SAr), theo (3.19d): NonlinearSemantization (Aj) → (NonlinearSemantization (Ai), NonlinearSemantization (Ak), ..., NonlinearSemantization (Ar)),

Thì đầu ra dự báo được xác định theo (3.20d) cho từng dữ liệu lịch sử của nhóm quan hệ ngữ nghĩa: DSAj → NonlinearDesemantization (WSAiAj * SAi+ WSAkAj * SAk+...+ WSArAj * SAr) trên một khoảng giải nghĩa được chọn sao cho bao được ui, uk... ur và thuộc khoảng xác định của tập nền chuỗi thời gian mờ [Dmin−D1, Dmax+D2]. Trong đó WSAiAj, WSAkAj..., WSArAj là trọng số ngữ nghĩa của từng thành phần trong NQHNN theo nhãn ngữ nghĩa Aj và được tính bằng tỉ số giữa số dữ liệu thuộc khoảng ui và tổng số dữ liệu thuộc các khoảng ui, uk,..., ur của NQHNN.

Lưu ý rằng cách chọn khoảng giải nghĩa như trên luôn đảm bảo không phá vỡ nhóm quan hệ mờ nhưng đồng thời có thể cho phép tính toán dự báo cho từng điểm dự báo trong cùng nhóm quan hệ mờ.

Trong bài toán dự báo số sinh viên nhập học tại trường đại học Alabama, có thể chọn các khoảng giải nghĩa hợp lí theo phép *thử – sai* với các giá trị đầu, giá trị cuối như trong Bảng 3.1 sau đây:

Bảng 3 1 Giá trị đầu và giá trị cuối của các khoảng giải nghĩa được chọn.

Các điểm dự báo	Giá trị đầu khoảng	Giá trị cuối khoảng	Các điểm dự báo	Giá trị đầu khoảng	Giá trị cuối khoảng
1 (1972)	13000	17000	12 (1983)	14000	18000
2 (1973)	13000	18000	13 (1984)	14000	17000
3 (1974)	13000	20000	14 (1985)	14000	17000
4 (1975)	15000	16000	15 (1986)	15000	18000
5 (1976)	14000	17000	16 (1987)	15000	19000
6 (1977)	14000	18000	17 (1988)	15000	20000
7 (1978)	15000	18000	18 (1989)	16000	20000
8 (1979)	15000	19000	19 (1990)	17000	20000
9 (1980)	15000	19000	20 (1991)	17000	20000
10 (1981)	14000	19000	21 (1992)	15000	20000
11 (1982)	13000	18000			

Ví dụ tính toán dự báo cho năm 1972 với θ = 0.5, α = 0.5, sp = 0.3 và dp = - 0.2:

Thực hiện các bước 1, 2, 3 và 4 bước như ở trên, sau đó tính toán ngữ nghĩa cho nhóm 1 tại bước 5 với NQHNN SA1 → (SA1, SA1, SA2) như sau:

Theo Bảng 3.2: Nhóm 1 có NQHNN thuộc các khoảng u1 và u2. Số dữ liệu thuộc khoảng u1 gồm 3 giá trị: 13055, 13563 và 13867 nhưng trùng nhau 2 lần. Do đó số dữ liệu thuộc khoảng u1 là (3*2 = 6). Số dữ liệu thuộc khoảng u2 gồm 1 giá trị: 14696. Như vậy tổng số dữ liệu thuộc các khoảng u1, u2 của nhóm 1 là (3*2+1) = 7 và trọng số ngữ nghĩa của SA1 theo nhãn ngữ nghĩa A1 là WSA1A1 = 3 / (3*2+1) = 3/7. Tương tự tính được trọng số ngữ nghĩa của SA2 theo nhãn ngữ nghĩa A1 là WSA2A1 = 1/7. Với SA1 = 0,125; SA2 = 0,25, ngữ nghĩa của nhóm 1 là:

(SA1, SA1, SA2) = WSA1A1*SA1 + WSA1A1*SA1 + WSA2A1*SA2

= (3/7)*0.125 + (3/7)*0.125 + (1/7)*0.25 = 0.143.

Khoảng giải nghĩa dự báo được chọn cho năm 1972 theo Bảng 3.1 là [13000 – 17000].

Trước hết tính toán giá trị giải nghĩa tuyến tính cho phép ngữ nghĩa hóa phi tuyến theo (3.20d1) với sp = 0.3:

Denormalization ($f(x_s,sp)$) = f(0.143,0.3) = (0.3*0.143*(1-0.143)+0.143)*(17000-13000) + 13000 = 13719.

Tiếp tục tính giá trị giải nghĩa phi tuyến cho phép ngữ nghĩa hóa phi tuyến theo (3.20d) với dp = -0.2:

Nonlinear Denormalization ($f(x_s,sp_s)$ g(13719,-0.2) = (-0.2)*(13719-13000)*(17000-13719) / (17000-13000) + 13719 = 13600

Như vậy, giá trị dự báo cho năm 1972 theo (3.20d) là:

DSA1 → NonlinearDeNormalization ($f(x_s,sp)$) = g(13719,-0.2) = 13600

Lưu ý rằng: trong các công thức trên, đã sử dụng dấu '.' thay cho dấu ',' của các số thập phân.

Bằng cách tương tự có thể tính toán dự báo cho các năm 1973, 1974… để nhận được các giá trị dự báo cụ thể cho năm 1973, 1974, …, 1992. Như vậy với số sinh viên nhập học từ 1971 đến 1992, trên cơ sở 6 bước theo tiếp cận ĐSGT, xây dựng được mô hình dự báo cho năm 1971 → 1972 , 1972 → 1973, 1973 → 1974,….. , 1991 → 1992.

Chương trình tính toán trên cơ sở sử dụng MATLAB R2013a. Kết quả của mô hình dự báo sử dụng ĐSGT được mô tả trong Bảng 3.2 để so sánh với các kết quả của một số mô hình dự báo khác hiện có với cùng 7 khoảng chia.

Bảng 3 2. So sánh các phương pháp dự báo với 7 khoảng chia.

Năm	Số sinh viên nhập học	Phương pháp Chen [4]	Phương pháp Lee [10]	Phương pháp Hwang [23]	Phương pháp Qiu [25]	Phương pháp Huarng [22]	Phương pháp ĐSGT
1971	13055						
1972	13563	14000	13833		14195	14000	13600
1973	13867	14000	13833		14424	14000	13750
1974	14696	14000	13833		14593	14000	14050
1975	15460	15500	15500		15589	15500	15396
1976	15311	16000	15722	16260	15645	15500	15232
1977	15603	16000	15722	15511	15634	16000	15642

1978	15861	16000	15722	16003	16100	16000	16232
1979	16807	16000	15722	16261	16188	16000	16643
1980	16919	16833	16750	17407	17077	17500	17027
1981	16388	16833	16750	17119	17105	16000	16533
1982	15433	16833	16750	16188	16369	16000	15533
1983	15497	16000	15722	14833	15643	16000	15642
1984	15145	16000	15722	15497	15648	15500	15232
1985	15163	16000	15722	14745	15622	16000	15232
1986	15984	16000	15722	15163	15623	16000	16232
1987	16859	16000	15722	16384	16231	16000	16643
1988	18150	16833	16750	17659	17090	17500	17534
1989	18970	19000	19000	19150	18325	19000	19288
1990	19328	19000	19000	19770	19000	19000	19466
1991	19337	19000	19000	19928	19000	19500	19466
1992	18876	19000	19000	19537	19000	19000	19111
MSE		**407507**	**397537**	**321418**	**261473**	**226611**	**65020**

Trong trường hợp phép ngữ nghĩa hóa tuyến tính và phép giải nghĩa tuyến tính với sp = 0 và dp = 0, kết quả tính toán nhận được **MSE = 69304.**

Trong trường hợp phép ngữ nghĩa hóa phi tuyến và phép giải nghĩa phi tuyến với sp = 0,3 và dp = - 0,2, kết quả tính toán nhận được **MSE = 65020.**

4. MÔ HÌNH DỰ BÁO TỐI ƯU THEO TIẾP CẬN ĐSGT

Vấn đề dự báo tối ưu chuỗi thời gian mờ theo nghĩa cực tiểu sai số trung bình bình phương MSE có thể được thực hiện trên cơ sở phép ngữ nghĩa hóa tối ưu (3.19d) và phép giải nghĩa tối ưu (3.20d) với các khoảng giải nghĩa được chọn dựa trên luật 1, luật 3 và bộ tham số cần tối ưu θ, α, sp, dp của tiếp cận ĐSGT.

Chương trình tính toán trên cơ sở sử dụng phần mềm tối ưu hóa GA của MATLAB R2013a. Kết quả của mô hình dự báo dựa trên ĐSGT với các tham số tối ưu θ*, α*, sp*, dp* theo nghĩa cực tiểu hàm MSE được mô tả trong Bảng 4.1, trong đó MSE có dạng:

$$(\sum_{i=1}^{21}(SSVNHTTi - SSVNHDBi))/21$$

ở đây: MSE (Mean Square Error) là sai số trung bình bình phương;

SSVNHTTi là số sinh viên nhập học thực tế năm i;

SSVNHDBi là số sinh viên nhập học dự báo năm i, i = 1 (1972), 2 (1973), ..., 21 (1992).

Bảng 4.1. Kết quả tính toán dự báo tối ưu số sinh viên nhập học tại trường đại học Alabama từ 1971 đến 1992 theo tiếp cận ĐSGT.

Năm	Số sinh viên nhập học thực tế	Số sinh viên nhập học dự báo	Năm	Số sinh viên nhập học thực tế	Số sinh viên nhập học dự báo
1971	13055		1982	15433	16439
1972	13563	13865	1983	15497	15219
1973	13867	14082	1084	15145	15219
1974	14696	14514	1985	15163	15219
1975	15460	15391	1986	15984	16219
1976	15311	15219	1987	16859	15812
1977	15603	15219	1988	18150	17439
1978	15861	16219	1989	18970	19165
1979	16807	16625	1990	19328	19165
1980	16919	16951	1991	19337	19165
1981	16388	16439	1992	18876	19165

Bộ tham số tối ưu nhận được: $\theta^* = 0.815$; $\alpha^* = 0.286$; $sp^* = -0.552$ và $dp^* = -0.529$ với MSE= 44507.

Trong trường hợp phép ngữ nghĩa hóa tuyến tính và phép giải nghĩa tuyến tính với sp = 0 và dp = 0, kết quả tính toán nhận được các giá trị tối ưu $\theta^* = 0{,}531$; $\alpha^* = 0{,}411$ với MSE = 45222 kém hơn so với trường hợp phi tuyến với MSE = 44507. Mô hình dự báo chuỗi thời gian mờ tối ưu theo tiếp cận ĐSGT ứng dụng cho bài toán dự báo số sinh viên nhập học tại trường đại học Alabama được so sánh với các mô hình dự báo khác theo tiếp cận mờ sử dụng bậc cao, số khoảng lớn hơn 7, được tổng hợp trong Bảng 4.2.

Bảng 4.2. So sánh các kết quả mô hình dự báo tối ưu theo tiếp cận ĐSGT và các kết quả mô hình dự báo cải tiến khác.

Phương Pháp	MSE	Phương Pháp	MSE
Qiu [30]	**199809**	Chen [5]	**86694**
Bai [29]	**140676**	Ozdemir [31]	**78073**
Singh [27]	**133700**	Egrioglu [32]	**60714**
Uslu [24]	**106276**	Tiếp cận ĐSGT với sp = 0 và dp = 0	**45222**
Singh [26]	**87025**	Tiếp cận ĐSGT với $sp^* = -0.552$ và $dp^* = -0.529$	**44507**

5. KẾT LUẬN

Vấn đề dự báo chuỗi thời gian mờ trong những năm gần đây được rất nhiều chuyên gia trên thế giới quan tâm nghiên cứu. Nhiều nghiên cứu sử dụng mô hình chuỗi thời gian mờ bậc cao

với độ dài khoảng và số lượng khoảng hợp lí đã cho kết quả dự báo số sinh viên nhập học tại trường Đại học Alabama khá chính xác [8, 10, 11, 27]. Mô hình dự báo dựa trên ĐSGT là một mô hình mới, hoàn toàn khác biệt, có khả năng dự báo chuỗi thời gian mờ với độ chính xác cao hơn so với một số mô hình dự báo hiện có. Sự khác biệt thể hiện ở phương pháp luận khi lần đầu tiên sử dụng phép ngữ nghĩa hóa phi tuyến thay cho phép mờ hóa, nhóm quan hệ ngữ nghĩa thay cho nhóm quan hệ mờ và phép giải nghĩa phi tuyến thay cho phép giải mờ. Mặc dù chỉ sử dụng mô hình chuỗi thời gian mờ bậc nhất với 7 khoảng chia dữ liệu lịch sử như mô hình dự báo đầu tiên của Chen [4], nhưng kết quả ứng dụng mô hình dự báo dựa trên ĐSGT với sự tham số hóa các nhãn ngữ nghĩa từ (3.21) đến (3.27) của biến ngôn ngữ thể hiện định tính chuỗi số liệu nêu trên có trong Bảng 3.3 đã cho thấy rõ hiệu quả dự báo tốt hơn so với một số phương pháp dự báo cùng sử dụng 7 khoảng hiện có [4, 5, 6, 28, 29] trong trường hợp phép ngữ nghĩa hóa và phép giải nghĩa tuyến tính với **MSE = 69304,** hoặc phi tuyến với **MSE = 65020**. Hơn nữa, mô hình dự báo với bộ tham số tối ưu của ĐSGT (theo Bảng 4.2 trong trường hợp phép ngữ nghĩa hóa và phép giải nghĩa tuyến tính với **MSE = 45222,** hoặc phi tuyến với **MSE = 44507)** cũng tốt hơn so với một số mô hình cải tiến hiện nay, khi các mô hình này đã phải sử dụng chuỗi thời gian mờ bậc cao, số khoảng chia lớn hơn 7 [5, 24], hoặc một số mô hình dự báo tối ưu khác trong [10, 11, 25]. Có thể thấy rằng: các phương pháp dự báo chuỗi thời gian mờ hiện nay đều có khả năng xây dựng phép mờ hóa tối ưu để tận dụng tốt nhất nguồn thông tin từ nhóm quan hệ mờ cho bài toán dự báo. Nhưng các phương pháp này chưa chú ý đến phép giải mờ tối ưu. Từ kết quả thể hiện trong Bảng 4.2 rõ ràng rằng: tính chính xác hơn của mô hình dự báo tối ưu sử dụng ĐSGT so với một số mô hình dự báo tối ưu khác được đảm bảo ở khả năng tính toán tối ưu các tham số của ĐSGT θ^*, α^* trong sự kết hợp với các tham số mở rộng của phép ngữ nghĩa hóa phi tuyến tối ưu sp^* và phép giải nghĩa phi tuyến tối ưu dp^* trên cơ sở sở khai thác có tính hệ thống những thông tin ẩn chứa trong nhóm quan hệ ngữ nghĩa. Những kết quả của bài báo này đã mở ra hướng nghiên cứu mới cho lĩnh vực dự báo chuỗi thời gian mờ.

Lời cảm ơn. Bài báo nghiên cứu được Quỹ phát triển Khoa học và Công nghệ Quốc gia (NAFOSTED) tài trợ theo Hợp đồng số 102.05-2013.34.

TÀI LIỆU THAM KHẢO

1. Song Q., Chissom B.S. - Fuzzy time series and its models, Fuzzy Sets and Syst. **54** (1993) 269–277.
2. Song Q., Chissom B.S. - Forecasting enrollments with fuzzy time series – part 1, Fuzzy Sets and Syst. **54** (1993) 1–9.
3. Song Q., Chissom B. S. - Forecasting enrollments with fuzzy time series – part 2, Fuzzy Sets and Syst. **62** (1994) 1–8.
4. Chen S. M. - Forecasting Enrollments Based on Fuzzy Time Series, Fuzzy Sets and Syst. **81** (1996) 311–319.
5. Lee M. H., Efendi R., Ismad Z. - Modified Weighted for Enrollments Forecasting Based on Fuzzy Time Series, MATEMATIKA **25** (1) (2009) 67-78.
6. Huarng K. - Heuristic Models of Fuzzy Time Series for Forecasting, Fuzzy Sets and Syst. **123** (2001) 369–386.
7. Nguyễn Công Điều - Một thuật toán mới cho mô hình chuỗi thời gian mờ, Tạp chí Khoa học và Công nghệ **49** (4) (2011) 11-25.
8. Chen S. M. - Forecasting Enrollments based on High Order Fuzzy Time Series, Cybernetics and Systems: An International Journal **33** (2002) 1-16.

9. Chen S. M and Chung N. Y. - Forecasting enrollments using high-order fuzzy time series and genetic algorithms, Int. Journal of Intelligent Systems **21** (2006) 485-501.
10. Ozdemir O., Memmedli M. - Optimization of Interval Length for Neural Network Based Fuzzy Time Series. IV International Conference "Problems of Cybernetics and Informatics", September **12-14** (2012) 104-105.
11. Egrioglu E., Aladag C.H., Yolcu U., Uslu V. R., Basaran M. A. - Finding an optimal interval length in high order fuzzy time series. Expert Systems with Applications **37** (2010) 5052–5055.
12. Bai E., Wong W. K., Chu W. C., Xia M. and Pan F. - A heuristic time invariant model for fuzzy time series forecasting, Expert Systems with Applications **38** (2011) 2701-2707.
13. Ho N. C. and Wechler W. - Hedge algebras: An algebraic approach to structures of sets of linguistic domains of linguistic truth variable, Fuzzy Sets and Systems **35**(3) (1990) 281-293.
14. Ho N. C. and Wechler W. - Extended hedge algebras and their application to Fuzzy logic, Fuzzy Sets and Systems **52** (1992) 259-281.
15. Nguyen Cat Ho, Vu Nhu Lan, and Le Xuan Viet - Optimal hedge-algebras-based controller: Design and Application, Fuzzy Sets and Systems **159** (2008) 968– 989.
16. Dinko Vukadinović, Mateo Bašić, Cat Ho Nguyen, Nhu Lan Vu, and Tien Duy Nguyen - Hedge-Algebra-Based Voltage Controller for a Self-Excited Induction Generator, Control Engineering Practice **30** (2014) 78–90.
17. Nguyen Dong Anh, Bui Hai Le, Vu Nhu Lan, and Tran Duc Trung - Application of hedgealgebras-based fuzzy controller to active control of a structure against earthquake Struct. Control Health Monit **20** (2013) 483–495.
18. Hai Le Bui, Duc Trung Tran, and Lan Nhu Vu - Optimal fuzzy control of inverted pendulum, Journal of Vibration and Control **18** (14) (2012) 2097-2110.
19. Nguyen Dinh Duc, Vu Nhu Lan, Tran Duc Trung and Bui Hai Le - A study on the application of hedge algebras to active fuzzy control of a seism-excited structure, Journal of Vibration and Control **18** (14) (2012) 2186–2200.
20. Duong T. L., Nguyen C. H., Pedrycz W., Tran T. S. - A Genetic Design of Linguistic Terms for Fuzzy Rule Based Classifiers, Int. J. Approx. Reason **54** (2013) 01-21.
21. Nguyen C. H., Huynh V. N., Pedrycz W. - A Construction of Sound Semantic Linguistic Scales Using 4-Tuple Representation of Term Semantics, Int. J. Approx. Reason 55 763–786, 2014
22. Ho N. C and Nam H. V. - An algebraic approach to linguistic hedges in Zadeh's fuzzy logic, Fuzzy Set and System **129** (2002) 229-254.
23. Chen S. M. and Wang N. Y. - Fuzzy Forecasting Based on Fuzzy-Trend Logical Relationship Groups, IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics **40** (5) (2010) 1343-1358.
24. Chen M., Chen C. D. - Handling forecasting problems based on high-order fuzzy logical relationships, Expert Systems with Applications **38** (2011) 3857–3864.
25. Uslu V. R., Bas E., Yolcu U., Egrioglu E. - A New Fuzzy Time Series Analysis Approach by using Differential Evolution Algorithm and Chronologically-Determined Weights, Journal of Social and Economic Statistics **2**(1) (2013) 18-30.
26. Qiu W., Liu X., and Li H. - High-Order Fuzzy Time SeriesModel Based on Generalized Fuzzy Logical Relationship. Hindawi Publishing Corporation. The Scientific World Journal, Mathematical Problems in Engineering **2013** (2013) Article ID 927394, 11 pages.

27. Singh S. R. - A robust method of forecasting based on fuzzy time series.Applied Mathematics and Computation **188** (2007) 427-484.
28. Hwang J. R., Chen S. M., and Lee C. H. - Handling Forecasting problems using fuzzy time series, Fuzzy Sets and Systems **100** (1998) 217-228.
29. Qiu W., Liu X., and Li H. - Generalized Method for Forecasting Based on Fuzzy Time Series. Expert Systems with Applications **38** (2011) 10446-10453.

ABSTRACT

THE APPLICATION OF HEDGE ALGEBRAS IN FUZZY TIME SERIES FORECASTING

Nguyen Cat Ho[1], Nguyen Cong Dieu[1, 2], Vu Nhu Lan[1, 2, *]

[1]*Institute of Information Technology, VAST, 18 Hoang Quoc Viet str., Cau Giay, Hanoi*
[2]*Thang Long University, Nghiem Xuan Yem Road, Dai Kim, Hoang Mai, Hanoi*

[*]Email: *vnlan@ioit.ac.vn*

Fuzzy time series given by Song & Chissom (1993) in magazine "Fuzzy Sets and Systems" has been widely studied for forecasting purposes. However, the accuracy of forecasts based on the concept of fuzzy approach of Song & Chissom is not high because such depends on many factors. Chen (1996) proposed an efficient fuzzy time series model which consists of simple arithmetic calculations only. After that, this has been widely studied for improving accuracy of forecasting in many applications to get better results.

The hedge algebras developed by Nguyen and Wechler (1990) was completely different from the fuzzy approach. Here the hedge algebras was used to model linguistic domains and variables and their semantic structure is obtained. Instead of performing fuzzification and defuzzification, more simple methods are adopted, termed as semantization and desemantization, respectively. The hedge algebras based fuzzy system is a new topic, which was first applied to fuzzy control 2008 [15]. Hedge algebras applications for some specific problems in the field of information technology and control has a number of important results and confirm advantages of this approach in comparing with fuzzy approach. In continuilty of hedge algebras applications, this paper is mainly focused on the field of fuzzy time series forecasting under hedge algebras approach.

In this paper, we present a new approach using hedge algebras to provide a computational model, which is completely different from the fuzzy approach for fuzzy time series forecasting. The experimental results of forecasting enrollments of students of the University of Alabama show that the model of fuzzy time series based on hedge algebras is better than many existing models.

Keywords: fuzzy sets, fuzzy logical relationship groups, Hedge algebras, fuzzy time series forecasting.