DỰ BÁO CHUỖI THỜI GIAN MỜ DỰA TRÊN NGỮ NGHĨA

Nguyễn Duy Hiếu[1], Vũ Như Lân[2,3] , Nguyễn Cát Hồ[2,4]
[1]Trường Đại học Tây Bắc
[2]Viện Công nghệ thông tin, Viện Hàn lâm Khoa học và Công nghệ Việt Nam
[3]Trường Đại học Thăng Long
[4]Trường Đại học Duy Tân
hieu3210@gmail.com, vnlan@ioit.ac.vn, ncatho@gmail.com

TÓM TẮT *- Bài toán dự báo chuỗi thời gian mờ đã được nhiều tác giả quan tâm nghiên cứu như: Song, Chissom, S. M. Chen… Các nghiên cứu tập trung giải quyết việc nâng cao độ chính xác của đầu ra dự báo. Có nhiều phương pháp đã được đưa ra nhằm cải tiến mô hình dự báo ban đầu của Song, Chissom, Chen với trung bình sai số bình phương (MSE) ngày càng thấp. Trong vài năm trở lại đây, đại số gia tử đã được ứng dụng có hiệu quả trong nhiều bài toán như điều khiển, phân lớp, tính toán trên từ,… với nhiều kết quả tốt hơn so với tiếp cận mờ. Điểm quan trọng và khác biệt của đại số gia tử là xem xét các biến ngôn ngữ trong quan hệ thứ tự vốn có của chính các giá trị ngữ nghĩa. Bài báo này trình bày về cách tiếp cận mới dựa trên đại số gia tử theo ngữ nghĩa trong bài toán dự báo chuỗi thời gian mờ. Mô hình dự báo chuỗi thời gian mờ bằng đại số gia tử sẽ được kiểm định qua các kết quả tính toán dự báo dựa trên dữ liệu sinh viên nhập học của Trường Đại học Alabama từ năm 1971 đến 1992 mà nhiều tác giả trên thế giới sử dụng. Qua đó có thể thấy được hiệu quả của mô hình dự báo đề xuất mới.*

Từ khóa *- Chuỗi thời gian, mô hình dự báo chuỗi thời gian mờ, chuỗi thời gian mờ, đại số gia tử, ngữ nghĩa.*

I. MỞ ĐẦU

Trong thực tế, chúng ta gặp rất nhiều các dữ liệu dạng chuỗi thời gian như: nhiệt độ trung bình của một vùng theo ngày, chỉ số chứng khoán, giá vàng,… Những dữ liệu ấy thường được biểu diễn dạng chuỗi giá trị biến đổi theo thời gian. Bài toán dự báo cho dữ liệu chuỗi thời gian luôn là vấn đề được quan tâm của các nhà khoa học trên thế giới. Q. Song và B. S. Chissom lần đầu tiên đưa ra khái niệm về chuỗi thời gian mờ, nghĩa là xem xét giá trị định lượng của các giá trị trong chuỗi thời gian từ góc độ định tính. Từ đó, chuyển bài toán dự báo về việc dự báo các giá trị ngôn ngữ của các biến ngôn ngữ. Khi đó có thể sử dụng các luật mờ, các suy luận mờ để có thể đưa ra kết quả dự báo. Đây có thể coi là quan niệm mới, có tính đột phá. Mô hình dự báo chuỗi thời gian của Q. Song và B. S. Chissom [1, 2, 3] đưa ra khả năng dự báo qua quá trình dự báo lại các dữ liệu lịch sử, tuy nhiên độ chính xác chưa cao. S. Chen trong những nghiên cứu của mình [4, 5, 6, 7] đã thay đổi các tính toán của trong [2, 3] thành các phép tính số học đơn giản hơn. Tiếp nối những nghiên cứu đó, nhiều nghiên cứu khác đã thu những kết quả quan trọng [8, 19, 20, 21] trong việc dự báo về chuỗi thời gian mờ. Bài báo số [18] là nghiên cứu đầu tiên về dự báo chuỗi thời gian mờ tại Việt Nam.

Các nghiên cứu về mô hình dự báo chuỗi thời gian tập trung giải quyết việc nâng cao độ chính xác của kết quả dự báo. Trong chuỗi thời gian mờ có thể thấy hai yếu tố ảnh hưởng tới độ chính xác dự báo:

- Mờ hoá dữ liệu.
- Giải mờ.

Việc mờ hoá dữ liệu đòi hỏi phải có kinh nghiệm và trực giác tốt để có thể mô tả định tính các giá trị định lượng một cách phù hợp. Tham số quan trọng trong việc mờ hoá đó 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 chính xác thấp do thiếu 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 có thể tạo ra nhiều khoảng không chứa dữ liệu hoặc chỉ chứa một dữ liệu. Việc tìm ra số lượng khoảng chia phù hợp là một vấn đề khó khăn. Ngoài ra, để tăng độ chính xác người ta cũng có thể tăng bậc của chuỗi thời gian mờ. Từ đó xây dựng được những nhóm quan hệ mờ phù hợp có lợi cho dự báo sau này.

Giải mờ là quá trình dự báo trên cơ sở phép mờ hoá trên đây và cần hướng tới dự báo tối ưu.

Những nghiên cứu tập trung giải quyết hai vấn đề trên để nâng cao độ chính xác dự báo. Vấn đề thứ nhất có thể thấy rõ trong các nghiên cứu [5, 6, 7]. Theo đó, các nghiên cứu chỉ rõ rằng: số lượng khoảng, độ dài khoảng và bậc của chuỗi thời gian mờ ảnh hưởng nhiều tới độ chính xác của dự báo. Vấn đề nghiên cứu tìm ra những giá trị đó phù hợp cũng đã có nhiều kết quả. Ngoài ra, các tác giả cũng đưa ra những cách tiếp cận khác như phân cụm, tham số hoá mức độ thay đổi của chuỗi thời gian. Vấn đề thứ hai là giải mờ để tìm ra giá trị dự báo. Theo S. Chen [4] thì cần dùng 3 luật cơ bản để giải quyết vấn đề này. Có thể coi phép giải mờ này dựa trên cơ sở trung bình hoá các trọng số có giá trị ngôn ngữ trong nhóm quan hệ mờ.

Cách tiếp cận theo lý thuyết mờ cho bài toán dự báo chuỗi thời gian đã tìm ra được nhiều cách làm hay, nhiều phương pháp tốt để có thể ngày một nâng cao kết quả dự báo. Nhưng phương pháp đó cũng ngày càng được cải tiến và cho độ chính xác ngày càng cao.

Năm 1990, N. Cat Ho và W. Wechler đã giới thiệu đại số gia tử (ĐSGT) [10] và từ đó cho đến nay, nhiều công trình nghiên cứu đã cho những kết quả tốt đẹp [12, 13, 14, 15, 16, 17, 18]. Có thể kể tới ứng dụng của ĐSGT trong điều khiển học, trích rút tri thức hay gần đây là tính toán trên từ đều cho những kết quả tốt hơn so với cách tiếp cận mờ. 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. Có thể thấy rằng: tính chất tự nhiên của ngữ nghĩa các giá trị ngôn ngữ của một miền giá trị biến ngôn ngữ là ngữ nghĩa vốn có tính so sánh được, nghĩa là giữa các giá trị ngôn ngữ có tồn tại khách quan một quan hệ thứ tự phản ánh thứ tự vốn có trên tập nền của biến ngôn ngữ. Trong khi ngữ nghĩa ngôn ngữ dựa trên tập mờ bỏ qua quan hệ thứ tự này. Như vậy, ĐSGT mô hình hóa ngữ nghĩa các giá trị ngôn ngữ đúng bản chất hơn, hay nói khác đi, nó cố gắng phát hiện các tính chất tự nhiên của các giá trị ngôn ngữ vốn tồn tại trong cấu trúc thứ tự đó.

Bài báo này là một trong những nghiên cứu để sử dụng lý thuyết của đại số gia tử trong bài toán dự báo chuỗi thời gian. Từ đó, tìm ra những cách tiếp cận mới và tìm cách nâng cao độ chính xác của đầu ra dự báo.

Bài báo được trình bày theo thứ tự: Sau phần MỞ ĐẦU ở mục I sẽ trình bày mục II về MÔ HÌNH DỰ BÁO CHUỖI THỜI GIAN MỜ theo cách tiếp cận của Q. Song, B. S. Chissom [2, 3] và S. Chen [4]. Mục III sẽ nêu TÓM TẮT MÔ HÌNH TÍNH TOÁN CỦA ĐẠI SỐ GIA TỬ trong bài toán dự báo chuỗi thời gian mờ. Phương pháp dự báo theo lý thuyết của ĐSGT, cách tính toán, kết quả dự báo sẽ được đưa ra. Vấn đề tối ưu các tham số cũng sẽ được trình bày. Số liệu phục vụ cho tính toán là số liệu về sinh viên nhập học của Trường Đại học Alabama từ năm 1971 tới 1992 mà nhiều nghiên cứu dùng để so sánh kết quả dự báo thông qua việc đánh giá sai số trung bình bình phương MSE (Mean Square Error) để có thể thấy rõ tính ưu việt của cách tiếp cận ĐSGT so với tiếp cận mờ.

II. 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, 2, 3] và được Chen cải tiến [4, 5, 6] để có thể xử lý 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 * ký 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 ký hiệu:

$$F(t-1) \rightarrow F(t) \qquad (2.1)$$

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)=A_i and F (t)=A_j , quan hệ logic giữa F (t) and F(t−1) được ký hiệu bằng $A_i \rightarrow Aj$, trong đó A_i là vế trái và A_j 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:

$$F(t-n),...,F(t-2), F(t-1) \rightarrow F(t) \qquad (2.2)$$

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 ký 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:

$$F(t)=F(t-1)*R(t-1, t) \qquad (2.3)$$

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 t_1 và t_2 khác nhau, R(t_1, t_1−1)=R(t_2, t_2−1), 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:

$A_i \rightarrow A_{j1}$; $A_i \rightarrow A_{j2}$;....; $A_i \rightarrow A_{jn}$

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

$$A_i \rightarrow A_{j1,} A_{j2, , ...}, A_{jn} \quad (2.4)$$

Tập mờ A_{jk} (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, 2, 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 lý 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 $A_i = A_{is-1} * R$, ở đây ký 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 $R_i = A_s^T \times A_q$, với mọi quan hệ mờ k, $A_s \rightarrow A_q$,

$$R = \bigcup_{i=1}^{k} R_i \quad (2.5)$$

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

2.4. Luật dự báo chuỗi thời gian mờ

Luật dự báo cũng chính là phép giải mờ các kết quả đầu ra dự báo như ở bước 6 của mô hình dự báo [4].

Giả sử dữ liệu của chuỗi thời gian *F(t-1)* được mờ hóa bằng A_j, khi đó. Đầu ra dự báo của *F (t)* được xác định theo những luật (nguyên tắc) sau đây:

1. Nếu tồn tại quan hệ một - một trong nhóm quan hệ của A_j, ký hiệu là $A_j \rightarrow A_k$ và mức độ thuộc cao nhất của A_k tại khoảng u_k, thì đầu ra dự báo của *F (t)* là điểm giữa của u_k.

2. Nếu A_k là trống, có nghĩa là $A_j \rightarrow \varnothing$ và A_j có mức độ thuộc cao nhất tại khoảng u_j, thì đầu ra dự báo là điểm giữa của u_j.

3. Nếu tồn tại quan hệ một - nhiều trong nhóm quan hệ mờ của A_j , ký hiệu là $A_j \rightarrow A_1, A_2, \ldots, A_n$, và mức độ thuộc cao nhất của $A_1, A_2, \ldots, A_n$ tại các khoảng $u_1, u_2, \ldots, u_n$ tương ứng, thì đầu ra dự báo được tính bằng trung bình các điểm giữa $m_1, m_2, \ldots, m_n$ của $u_1, u_2, \ldots, u_n$. Đầu ra dự báo khi này có dạng: $(m_1+m_2+\ldots+m_n)/n$.

III. TÓM TẮT MÔ HÌNH TÍNH TOÁN CỦA ĐẠI SỐ GIA TỬ

Đại số gia tử cung cấp một mô hình xử lý các đại lượng không chắc chắn khá hiệu quả cho nhiều bài toán ứng dụng như điều khiển mờ [20, 23], chống động đất [24, 25, 26], phân lớp dựa trên luật mờ [22] và đặc biệt gần đây ĐSGT đã mở ra hướng nghiên cứu mới về tính toán trên từ (computing with words) [21]. 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 [17, 18], từ đó tạo ra môi trường tính toán, suy luận tốt cho nhiều ứng dụng.

Gọi AX = (X, G, C, H, ≤) 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; ≤ 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.

Ký 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 → [0, 1] gọi là độ đo tính mờ nếu thỏa mãn các điều kiện sau:

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

+) Với các phần tử *0*, *W* và *1*, $fm(0) = fm(W) = fm(1) = 0$ (3.2)

+) Và với $\forall x,y \in$ X, $\forall h \in$ H, $\dfrac{fm(hx)}{fm(x)} = \dfrac{fm(hy)}{fm(y)}$ (3.3)

Đẳng thức (3.3) không phụ thuộc vào các phần tử x, y và do đó ta có thể ký 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:

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

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

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

+) $\sum_{i=-1}^{-q} \mu(h_i) = \alpha$ và $\sum_{i=1}^{p} \mu(h_i) = \beta$, với $\alpha, \beta > 0$ và $\alpha+\beta = 1$ (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-) = -1, *Sign(c+)* = +1; (3.8)

Sign(hc) = - *Sign(c),* nếu h là âm đối với c; (3.9)

Sign(hc) = + *Sign(c),* nếu h là dương đối với c; (3.10)

Sign(h'hx) = -*Sign(hx),* nếu h'hx ≠ hx và h' là âm đối với h; (3.11)

Sign(h'hx) = + *Sign(hx)*, nếu h'hx ≠ hx và h' là dương đối với h; (3.12)

Sign(h'hx) = 0 nếu h'hx = hx. (3.13)

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

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

$$v(c^-) = \theta - \alpha fm(c^-) = \beta fm(c^-), \quad (3.15)$$

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

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

với $\omega(h_j x) = \dfrac{1}{2}[1 + Sign(h_j x) sign(h_p h_j x)(\beta - \alpha)] \in \{\alpha, \beta\}$, (3.18)

j ∈ [-q^p], j ≠ 0.

Để thuận tiện cho việc biểu diễn ngữ nghĩa của các giá trị ngôn ngữ [20], 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 [20, 23, 25, 26], 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). Như vậy có thể biểu diễn phép ngữ nghĩa hóa tuyến tính và phép giải nghĩa tuyến tính đơn giản như sau:

- Linear Semantization (x) = $x_s = a_s + (b_s – a_s) (x – a) / (b – a)$ (3.19a)
- Linear Desemantization (x_s) = $x = a + (b – a) (x_s – a_s) / (b_s – a_s)$ (3.20a)
- Normalization (x) = $x_s = (x – a) / (b – a)$ (3.19b)
- Denormalization (x_s) = $x = a + (b – a)x_s$ (3.20b)

Trong đó a, b là các số thực.

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. Tương tự trên, phép ngữ nghĩa hóa phi tuyến và phép giải nghĩa phi tuyến có thể được biểu diễn như sau:

- Nonlinear Semantization (x) = $f(x_s,sp)$ (3.19c)

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

- Nonlinear Desemantization (x_s) = $g(x,dp)$ (3.20c)

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

Các hàm f(.) và g(.) được chọn tùy theo từng ứng dụng và là các hàm liên tục, đồng biến, trong đó sp∈[-1 1] là tham số ngữ nghĩa hóa, dp ∈[-1 1] là tham số giải nghĩa. Ví dụ có thể chọn f(.) phi tuyến theo x_s thể hiện qua $f(x_s,sp)$ và g(.) phi tuyến theo x thể hiện qua Denormalization ($f(x_s,sp)$) như sau:

$f(x_s,sp) = sp*x_s*(1-x_s)+x_s$ (3.19d)

g(x,dp) = dp*((Denormalization ($f(x_s,sp)$)–a)*(b – Denormalization ($f(x_s,sp)$)) / (b-a) +

+ Denormalization ($f(x_s,sp)$) (3.20d)

Trong đó Denormalization ($f(x_s,sp)$) = (sp*x*(1-x)+x)*(b-a) + a. (3.20d1)

Hàm $f(x_s,sp)$ là hàm biểu diễn ngữ nghĩa phi tuyến trong phép giải nghĩa phi tuyến g(x.dp) chưa được sử dụng trong các ứng dụng của ĐSGT. Lưu ý rằng: có thể chọn các hàm $f(x_s,sp)$ và g(x,dp) độc lập với nhau.

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 μ(*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 [20]. 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 lý 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 bao gồm bước 1 đến bước 5. Nếu giai đoạn mờ hóa cung cấp những thông tin định tính hợp lý thì các quan hệ mờ hoặc nhóm quan hệ mờ sẽ tạo ra khả năng dự báo với độ chính xác cao. 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. Khó khăn nhất của giai đoạn này là tìm ra xu hướng dự báo. Trong giai đoạn này phải đánh giá được khả năng tăng hay giảm với mức độ nhiều hay ít một cách chính xác nhất có thể. Chính vì vậy kết quả dự báo trong [6, 19] có độ chính xác rất cao khi đưa ra hướng dự báo cho từng dữ liệu lịch sử dựa trên hiệu ứng tăng hay giảm và tốc độ tăng hay giảm của từng cặp dữ liệu.

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, 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) trên tập nền.

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 và nhiều tác giả trên thế giới quan tâm nghiên cứu cải tiến.

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 dõi và so sánh với các bước dự báo trong mô hình Chen, ở đây sử dụng một số ký hiệu tương tự những ký 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 ϵ 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= h_{A1}c; A2= h_{A2}c;….; Ak= h_{Ak}c, trong đó h_{Ai}, (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 μ(*very*), μ(*little*) và giá trị độ đo tính mờ của phần tử sinh *fm*(c^-) = θ 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 "*smal*" như trên, thì μ(*little*) = α và μ(*very*) = 1- α = β. 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:

$$\nu(\text{very small}) = \theta(1-\alpha)(1-\alpha) \tag{3.21}$$

$$\nu(\text{small}) = \theta(1-\alpha) \tag{3.22}$$

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

$$\nu(\text{midle}) = \theta \tag{3.24}$$

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

$$\nu(\text{large}) = \theta+(1-\theta)\alpha \tag{3.26}$$

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

Nếu chọn trước α = 0.5 và θ = 0.5, thì các phương trình từ (3.21) đến (3.27) trở thành:

$$\nu(\text{very small}) = 0.125 \tag{3.28}$$

$$\nu(\text{small}) = 0.25 \tag{3.29}$$

$$\nu(\text{little small}) = 0.375 \tag{3.30}$$

$$\nu(\text{midle}) = 0.5 \tag{3.31}$$

$$\nu(\text{little large}) = 0.625 \tag{3.32}$$

$$\nu(\text{large}) = 0.75 \tag{3.33}$$

$$\nu(\text{very large}) = 0.875 \tag{3.34}$$

Ký hiệu: SA = Semantization (A) là giá trị ngữ nghĩa định lượng theo nhãn ngữ nghĩa A, khi đó:

SA1 = ν(very small); SA2 = ν(small); SA3 = ν(little small); SA4 = ν(midle); SA5 = ν(little large);

SA6 = ν(large) và SA7 = ν(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:

$$SA1 < SA2 < SA3 < SA4 < SA5 < SA6 < SA7 \tag{3.35}$$

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 lớp 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 ký hiệu là:

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

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)

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 SA_k là trống, SAj → ∅, thì đầu ra dự báo được tính theo (3.20d):

DSAj → Nonlinear Desemantization (∅) 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]. Do cách chọn θ, α, điểm giữa khoảng uk chỉ là một trường hợp riêng của phép giải nghĩa với θ = 0.5, α = 0.5 và khi khoảng giải nghĩa uj được chọn sao cho bao được uj. Nguyên tắc thứ 2 theo tiếp cận ĐSGT mềm dẻo hơn so với nguyên tắc thứ 2 của Song & Chissom [1, 2, 3] và Chen [4].

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), hay 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: 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 tỷ số giữa số dữ liệu thuộc khoảng ui và tổng số dữ liệu thuộc các khoảng ui, uk,…, ur của NQHNN. Như vậy tính chuẩn hóa của các trọng số được đảm bảo: WSAiAj + WSAkAj +…+ WSArAj = 1.

Trong bài toán dự báo số sinh viên nhập học tại Trường Đại học Alabama, có thể chọn 21 khoảng giải nghĩa 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 21 khoảng cho 21 giá trị dự báo

Khoảng giải nghĩa cho các điểm dự báo	Giá trị đầu khoảng	Giá trị cuối khoảng	Khoảng giải nghĩa cho các điểm dự báo	Giá trị đầu khoảng	Giá trị cuối khoảng
1 (1972)	13000	17000	12 (1983)	14000	17000
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	17000
6 (1977)	14000	17000	17 (1988)	15000	20000
7 (1978)	15000	18000	18 (1989)	17000	20000
8 (1979)	15000	19000	19 (1990)	17000	20000
9 (1980)	15000	19000	20 (1991)	17000	20000
10 (1981)	14000	19000	21 (1992)	17000	20000
11 (1982)	14000	19000			

Mô hình dự báo chuỗi thời gian mờ theo tiếp cận ĐSGT gồm 6 bước cơ bản, trong đó các bước 1,2.., 5 có thể coi là những bước chuẩn bị cho tính toán dự báo cho bước 6 theo 3 nguyên tắc trên đây. Như vậy, có thể xây dựng được nguồn số liệu cung cấp thông tin tổng hợp phục vụ bài toán dự báo số sinh viên nhập học (SSVNH) tại Trường Đại học Alabama dưới dạng bảng 3.2 sau đây:

Bảng 3.2. Tổng hợp thông tin làm cơ sở cho mô hình dự báo theo tiếp cận ĐSGT

Khoảng ui, nhãn ngữ nghĩa Ai và ngữ nghĩa định lượng SAi	Số sinh viên nhập học	Năm nhập học	Ngữ nghĩa định lượng SAi với NQHNN, trọng số ngữ nghĩa của từng thành phần trong nhóm theo nhãn ngữ nghĩa Ai và tổng số dữ liệu	Số dữ liệu thuộc ui
u1 = [13000 – 14000] A1 = very small SA1 = ν(very small)	13055 13563 13867	1971 1972 1973	SA1 → (SA1,SA1, SA2) WSA1A1 = 3/(3x2+1) = 3/7 WSA2A1 = 1/(3x2+1) = 1/7 Tổng số dữ liệu: 7	3
u2 = [14000 – 15000] A2 = small SA2 = ν(small)	14696	1974	SA2 → SA3 WSA3A2 = 1/1 = 1 Tổng số dữ liệu: 1	1
u3 = [15000 – 16000] A3= little small SA3 = ν(little small)	15460 15311 15603 15861 15433 15497 15145 15163 15984	1975 1976 1977 1978 1982 1983 1984 1985 1986	SA3 → SA3, SA3, SA3, SA4, SA3, SA3, SA3, SA3, SA4 WSA3A3 = 9/(9x7+4x2) = 9/71 WSA4A3 = 4/(9x7+4x2) = 4/71 Tổng số lượng dữ liệu: 71	9
u4 = [16000 – 17000] A4 = midle SA4 = ν(middle)	16807 16919 16388 16859	1979 1980 1981 1987	SA4 → SA4, SA4, SA3, SA6 WSA4A4 = 4/(4x2+9+3) = 4/20 WSA3A4 = 9/(4x2+9+3) = 9/20 Tổng số dữ liệu: 20	4
u5 = [17000 – 18000] A5 = little large SA5 = ν(little large)			SA5 Không dự báo	0
u6 = [18000 – 19000] A6 = large SA6 = ν(large)	18150 18970 18876	1988 1989 1992	SA6 → SA6, SA7 WSA6A6 = 3/(3+2) = 3/5 WSA7A6 = 2/(3+2) = 2/5 Tổng số dữ liệu: 5	3
u7 = [19000 – 20000] A7 = very large SA7 = ν(very large)	19328 19377	1990 1991	SA7 → SA7, SA6 WSA7A7 = 2/(3+2) = 2/5 WSA6A7 = 3/(3+2) = 3/5 Tổng số dữ liệu: 5	2

Ví dụ tính toán dự báo cho điểm dự báo thứ nhất ứng với năm 1972:

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 (3x2 = 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 (3x2+1) = 7 và trọng số ngữ nghĩa của SA1 theo nhãn ngữ nghĩa A1 là WSA1A1 = 3 / (3x2+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 được chọn cho điểm dự báo thứ nhất (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 = 0.5: Denormalization (x_s) = f(0.143,0.5) = (0.5* 0.143*(1-0.143)+0.143) *(17000-13000) + 13000 = 13817. Tiếp tục tính giá trị giải nghĩa phi tuyến cho phép ngữ nghĩa hóa phi tuyến theo (3.20d) với dp = -0.5: Nonlinear Desemantization (x) = g(13817,-0.5) = (-0.5)*(13817-13000)*(17000-13817) / (17000-13000) + 13817 = 13492.

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

DSA1 → NonlinearDesemantization (x) = g(13817,-0.5) = 13492

Bằng cách tương tự có thể tính toán cho các điểm dự báo thứ 2, 3,… , 21 để nhận được các giá trị dự báo cụ thể cho năm 1973, 1974, …, 1992. Như vậy với số sinh viên nhập học của 22 năm từ 1971 đến 1992, trên cơ sở 6 bước theo tiếp cận ĐSGT, xây dựng được mô hình dự báo cho 21 năm 1971 → 1972 , 1972 → 1973, 1973 → 1974,….. , 1991 → 1992. Chương trình tính toán được thể hiện trong PHỤ LỤC 1 trên cơ sở sử dụng MATLAB R2013a. Kết quả của mô hình dự báo sử dụng ĐSGT được mô tả trong bảng 3.3. để so sánh với các kết quả của nhiều mô hình dự báo khác hiện có.

Lưu ý rằng về nguyên tắc, độ chính xác của các phương pháp dự báo chuỗi thời gian mờ theo tiếp cận của Song & Chisson, Chen và nhiều tác giả khác phụ thuộc rất nhiều vào quá trình mờ hóa chuỗi thời gian và giải mờ đầu ra dự báo và đặc biệt rất khó tối ưu hóa đồng thời hai quá trình này. Trong khi đó, mô hình tính toán theo tiếp cận ĐSGT đưa ra cách chọn bộ tham số θ, α, μ(.) , sp, dp để xây dựng dự báo tối ưu dựa trên phép ngữ nghĩa hóa và phép giải nghĩa tuyến tính hoặc phi tuyến. Đây là tính chất rất quan trọng của tiếp cận ĐSGT và là cơ sở khoa học cho tính hiệu quả cao trong nhiều bài toán ứng dụng nói chung và bài toán dự báo chuỗi thời gian mờ nói riêng.

Trong bảng 3.3 so sánh kết quả dự báo theo tiếp cận ĐSGT với các mô hình dự báo khác cùng sử dụng chuỗi thời gian mờ với 7 khoảng chia.

Bảng 3.3. So sánh các phương pháp dự báo với 7 khoảng chia

Năm	**Số sinh viên nhập học**	**Phương pháp Chen [4]**	**Phương pháp Lee [9]**	**Phương pháp Hwang [23]**	**Phương pháp Qiu [24]**	**Phương pháp Huarng [20]**	**Phương pháp ĐSGT**
1971	13055						
1972	13563	14000	13833		14195	14000	13571
1973	13867	14000	13833		14424	14000	13714
1974	14696	14000	13833		14593	14000	14000
1975	15460	15500	15500		15589	15500	15375
1976	15311	16000	15722	16260	15645	15500	15167
1977	15603	16000	15722	15511	15634	16000	15167
1978	15861	16000	15722	16003	16100	16000	16167
1979	16807	16000	15722	16261	16188	16000	16556
1980	16919	16833	16750	17407	17077	17500	16925
1981	16388	16833	16750	17119	17105	16000	16406
1982	15433	16833	16750	16188	16369	16000	16406
1983	15497	16000	15722	14833	15643	16000	15167
1984	15145	16000	15722	15497	15648	15500	15167
1985	15163	16000	15722	14745	15622	16000	15167
1986	15984	16000	15722	15163	15623	16000	16167
1987	16859	16000	15722	16384	16231	16000	15778
1988	18150	16833	16750	17659	17090	17500	17406
1989	18970	19000	19000	19150	18325	19000	19400
1990	19328	19000	19000	19770	19000	19000	19400
1991	19337	19000	19000	19928	19000	19500	19400
1992	18876	19000	19000	19537	19000	19000	19400
MSE		**407507**	**397537**	**321418**	**261473**	**226611**	**198219**

IV. KẾT LUẬN

Qua kết quả dự báo trên có thể thấy cách tiếp cận theo đại số gia tử cho bài toán dự báo chuỗi thời gian có độ chính xác cao hơn các phương pháp hiện có cho bài toán cổ điển là dự báo số sinh viên nhập học của Trường Đại học Alabama. Mô hình tính toán theo đại số gia tử do có cấu trúc chặt chẽ đã làm giảm các tham số cần tối ưu, nâng cao khả năng chọn được những giá trị tối ưu tốt nhất. Việc áp dụng mô hình dự báo chuỗi thời gian dựa trên ngữ nghĩa này cũng cần được thử nghiệm với những bộ đầu vào khác để xem xét tính thích nghi. Ngoài ra, việc áp dụng các lý thuyết tối ưu, phân cụm trong mô hình này cũng cần được thử nghiệm và đối chiếu với những phương pháp hiện có.

LỜI CẢM ƠN

Bài báo này đã được Quỹ Phát triển Khoa học và Công nghệ Quốc gia (NAFOSTED) hỗ trợ theo hợp đồng số 102.05-2013.34.

V. TÀI LIỆU THAM KHẢO

[1] Song Q, Chissom B. S. Fuzzy time series and its models. Fuzzy Sets and Syst. 54 269–277, 1993.

[2] Song Q, Chissom B. S, Forecasting enrollments with fuzzy time series – part 1. Fuzzy Sets and Syst. 54, 1–9, 1993.

[3] Song Q, Chissom, B S, Forecasting enrollments with fuzzy time series – part 2. Fuzzy Sets and Syst. 62, 1–8, 1994.

[4] Chen, S.M, Forecasting Enrollments Based on Fuzzy Time Series. Fuzzy Sets and Syst. 81, 311–319, 1996.

[5] Chen S M, Forecasting Enrollments based on High Order Fuzzy Time Series. Cybernetics and Systems: An International Journal. 33,1-16, 2002.

[6] Chen S M, Hsu, C C. : A New Method to Forecast Enrollments using Fuzzy Time Series. Int. Journal Applied Science and Engineering 2, 234-244, 2004.

[7] Chen S.M and Chung N.Y, Forecasting enrollments using high-order fuzzy time series and genetic algorithms, Int. Journal of Intelligent Systems 21, 485-501. 2006.

[8] Chen S M, Tanuwijaya K, Multivariate fuzzy forecasting based on fuzzy time series and automatic clustering techniques. Expert Systems with Applications 38, 10594–10605, 2011.

[9] Lee M H, Efendi R, Ismad Z, Modified Weighted for Enrollments Forecasting Based on Fuzzy Time Series. MATEMATIKA, 25(1), 67-78, 2009.

[10] N. Cat Ho and W. Wechler, Hedge algebras: An algebraic approach to structures of sets of linguistic domains of linguistic truth variable, Fuzzy Sets and Systems, Vol. 35,3, pp.281-293, 1990.

[11] N. Cat Ho and W. Wechler, Extended hedge algebras and their application to Fuzzy logic, Fuzzy Sets and Systems 52, 259-281, 1992.

[12] Cat Ho, N. and H. Van Nam: An algebraic approach to linguistic hedges in Zadeh's fuzzy logic, Fuzzy Set and System, 129, 229-254, 2002.

[13] Nguyen Cat Ho, Vu Nhu Lan, Le Xuan Viet, Optimal hedge-algebras-based controller: Design and Application, Fuzzy Sets and Systems 159, 968– 989, 2008.

[14] Dinko Vukadinović, Mateo Bašić, Cat Ho Nguyen, Nhu Lan Vu, Tien Duy Nguyen Hedge-Algebra-Based Voltage Controller for a Self-Excited Induction Generator, Control Engineering Practice, 30, 78–90, 2014.

[15] 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, 483–495, 2013.

[16] Hai Le Bui, Duc Trung Tran, Lan Nhu Vu, Optimal fuzzy control of inverted pendulum. Journal of Vibration and Control, 18 (14), 2097-2110, 2012.

[17] 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), 2186–2200, 2012.

[18] Cong Nguyen Huu, Duy Nguyen Tien, Trung Ngo Kien, Ha Le Thi Thu, A Research on Parabolic Trough Solar Collector System Control based on Hedge Algebra, 11^{th} International Conference on Control, Automation, Robotics and Vision, December, 715-720, 2010, Singapore.

[19] Nguyễn Công Điều: Một thuật toán mới cho mô hình chuỗi thời gian mờ. Tạp chí Khoa học và Công nghệ, Tâp 49, Số 4, 11-25, 2011.

[20] Huarng, K. Heuristic Models of Fuzzy Time Series for Forecasting. Fuzzy Sets and Syst. 123, 369–386, 2001.

[21] Huarng, K. Effective Lengths of Intervals to Improve Forecasting in Fuzzy Time Series. Fuzzy Sets and Syst. 123, 387–394, 2004.

[22] Hwang, J.-R., Chen, S.-M., Lee, C.-H.: Handling Forecasting problems using fuzzy time series. Fuzzy Sets and Systems 100, 217-228, 1998.

[23] Hwang, J.-R., Chen, S.-M., Lee, C.-H.: Handling Forecasting problems using fuzzy time series. Fuzzy Sets and Systems 100, 217-228, 1998.

[24] Qiu W, Liu X, Li H, Generalized Method for Forecasting Based on Fuzzy Time Series. Expert Systems with Applications 38, 10446-10453, 2011.

FUZZY TIME SERIES FORCASTING BASE ON SEMANTICS

Nguyen Duy Hieu, Vu Nhu Lan, Nguyen Cat Ho

ABSTRACT - *Forcasting on fuzzy times series has research by many reseachers such as Q. Song, B. S. Chissom, S. M. Chen… There are many methods have researched and presented to handle and improve the result of forcasting method that were introduce by Q. Song, B. S. Chissom and S. M. Chen. Those method has MSE (Mean Square Error) increasely lower. During recent years, hedge algebras has been applied to many problems such as control, classification or computing with words. Hedge algebras approaches have many better result than fuzzy approaches. Important and distinctive point of hedge algebras is recognize the values of linguistic variable in themselves order. This paper present hedge algebras approach for fuzzy times series forcasting problem. The computational model will be inspected and compare with other method base on data of general example about enrollments of Alabama University from 1971 to 1992. Thereby we can see and discuss the effects of new forcasting method.*

Keywords - *Times Series, Fuzzy Time Series, Forecasting, Hedge Algebra, semantic.*